ARDHI UNIVERSITY

HOTSPOT ANALYSIS OF WILDLIFE ROAD CARNAGE ALONG HIGHWAY CROSSING THROUGH MIKUMI NATIONAL PARK

CLINTON JOHN LYIMO

BSc. (Geoinformatics) Dissertation

Ardhi University, Dar es Salaam

July,2023

HOTSPOT ANALYSIS OF WILDLIFE ROAD CARNAGE ALONG HIGHWAY CROSSING THROUGH MIKUMI NATIONAL PARK

LYIMO, CLINTON JOHN

A Dissertation submitted to the Department of Geospatial Sciences and Technology in Partial fulfilment of the Requirements for the Award of Bachelor of Science in Geoinformatics (BSc. GI) of Ardhi University

CERTIFICATION

The undersigned certify that they have read and hereby recommend for acceptance by the Ardhi University a dissertation titled **"Hotspot analysis of wildlife road carnage along highway crossing through Mikumi National Park"** in partial fulfillment of the requirements for the award of degree of Bachelor of Science in Geoinformatics at Ardhi University.

………………………………… ……………………………..

Dr. Zakaria Robert Ngereja **Dr. Atupelye Komba**

(Main Supervisor) (Second Supervisor)

Date ……………………… Date………………………

DECLARATION AND COPYRIGHT

I, LYIMO CLINTON JOHN declare that, the contents of this dissertation are the results of my own original findings through my study and investigation, and to the best of my knowledge they have not been presented anywhere else as a dissertation for diploma, degree or any similar academic award in any institution of higher learning.

…………………………….

LYIMO, CLINTON JOHN

22702/T.2019

ACKNOWLEDGMENT

First and foremost, I would like to express my heartfelt gratitude to the Almighty God for His divine guidance, unwavering presence, and abundant blessings throughout the course of this research journey. I am eternally grateful for the strength, wisdom, and inspiration that I have received from Him.

I extend my sincere appreciation to my esteemed supervisors, Dr. Zakaria Robert Ngereja and Dr. Atupelye Komba, for their invaluable guidance, expertise, and unwavering support. Their profound knowledge and mentorship have been instrumental in shaping the direction and quality of this study. My gratitude extends also to members of Geospatial Science and Technology Department (DGST) for their treasured support which was really influential in shaping my methods and critiquing my results.

I would like to express my deep gratitude to the Tanzania National Parks Authority (TANAPA) and the management of Mikumi National Park for providing the necessary data and resources for this research. Their cooperation and support have been invaluable in conducting this study.

Furthermore, I am also thankful to my classmates for their camaraderie, intellectual discussions, and shared experiences. Their support and collaboration have enriched my academic journey and fostered a conducive learning environment. I am grateful to my family for their love, understanding, and support throughout this research endeavor. Their prayers and encouragement have been a source of motivation and strength.

Lastly, I would like to acknowledge and express gratitude to all the individuals and organizations who have directly or indirectly contributed to this study. Their assistance, cooperation, and provision of resources have been instrumental in the successful completion of this research.

DEDICATION

I dedicate this work to my dear mother, Doris Issay Macha, whose unwavering love and support have guided my path. Her strength has been my inspiration to overcome challenges and pursue excellence. In loving tribute to my late grandparents, Faraj Abdallah and Marry Mgunya, whose wisdom and values live on in me.

This dedication is a testament to the magnificence and intrinsic value of the wildlife species that call Mikumi National Park their home. It is a tribute to their resilience and the urgent need to protect and mitigate the threats they face, particularly in relation to road carnage.

I dedicate this work to the Tanzania National Parks Authority (TANAPA) and the management of Mikumi National Park, whose unwavering commitment to conservation and environmental stewardship has made this research possible. Your dedication to preserving the biodiversity and ecological integrity of this remarkable park is an inspiration.

To the researchers, scientists, and conservationists who tirelessly strive to understand and mitigate the impacts of human activities on wildlife, I dedicate this work. May our collective efforts pave the way for innovative solutions and sustainable practices that ensure the coexistence of humans and wildlife.

This work is also dedicated to future generations, whose stewardship of our planet will determine the fate of our natural heritage. May they find inspiration in this research to continue the vital work of conservation, protecting the delicate balance between human development and the preservation of wildlife and their habitats.

May this dedication serve as a reminder of our collective responsibility to cherish and protect the rich biodiversity that surrounds us, so that future generations can experience the wonders of nature as we have.

ABSTRACT

This study presents a comprehensive analysis of wildlife road carnage along the highway that crosses through Mikumi National Park. The study focuses on hotspot analysis and non-random distribution patterns of collision occurrences. By employing Ripley's K analysis and spatial autocorrelation (Moran's I) analysis, the study confirms the presence of hotspot areas where wildlife vehicle collisions are concentrated, indicating a non-random spatial distribution influenced by specific factors related to roads and functional groups. The study utilizes Generalized Linear Mixed-Effects Models (GLMMs) to analyze collision occurrences involving different species including, amphibians and reptiles, small-medium vertebrates (SMV) and birds, as well as Ungulates, primates, and carnivores. The GLMMs incorporate variables such as topographic visibility risk, traffic volume, and land classification categories to explore their impact on collision predictions. The findings of this study emphasize the importance of considering specific factors, such as traffic volume and land classifications related to built-up structures, water bodies, and dense vegetation, in understanding and mitigating wildlife road carnage. The developed GLMM models offer valuable predictive capabilities, aiding the identification of hotspot areas and informing targeted strategies to reduce collision occurrences in the study area. These research findings contribute to the field of wildlife road carnage analysis and provide valuable insights for policymakers, conservationists, and transportation authorities involved in managing road safety and wildlife conservation within Mikumi National Park. By implementing recommended measures such as targeted road infrastructure improvements, driver awareness campaigns, habitat connectivity planning, and ongoing monitoring, it is possible to mitigate wildlife-vehicle collisions and ensure the coexistence of wildlife and road infrastructure in this ecologically significant area.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF PLATES

LIST OF TABLES

ABREVIATIONS AND ACRONYMS

AADT	Annual Average Daily Traffic
AOI	Area of Interest
ARCGIS	Aeronautical Reconnaissance Coverage Geographic Information System
CSV	Comma Separated Values
DEM	Digital Elevation Model
GIS	Geographic Information System
GLMM	Generalized Linear Mixed Model
RS	Remote Sensing
SMV	Small-Medium Vertebrate
SRTM	Shuttle Radar Topography Mission
TANAPA	Tanzania National Park Authority
TANROAD	Tanzania National Roads Agency
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WVC	Wildlife Vehicle Collision

CHAPTER ONE

INTRODUCTION

1.1 Background

The intricate interplay between modern transportation infrastructure and the delicate ecosystems of natural wildlife habitats has led to a paradigm shift in the ecological landscape (Cramer et al., 2015). The far-reaching consequences of this transformation cannot be overstated, as the expansive network of roads has engendered a profound impact on wildlife, catalyzing the rise of an emergent field of inquiry known as road ecology. This multidisciplinary arena merges the realms of biology, ecology, geography, and urban planning to dissect the multifaceted ramifications of road networks on wildlife (Alexander and Waters, 2014).

One prominent and visibly discernible consequence of this evolution is the heart-wrenching phenomenon of wildlife-road carnage, a tragic occurrence where animals fall victim to the swift, unforgiving movement of vehicles, often succumbing to injury or fatality. However, the repercussions extend far beyond the realm of immediate mortality. A nuanced cascade of indirect effects reverberates across ecosystems, fragmenting habitats, permeating the air with noise and light pollution, obstructing natural movement patterns, and instigating a cascade of factors that prompt animals to either shy away from these roadways or see their habitats degraded (Alexander and Waters, 2014).

The compounding consequences are far-reaching and multifarious. The threat of extinction looms larger as species struggle to adapt to these rapidly changing circumstances, a delicate balance between survival and obliteration. Simultaneously, these developments amplify human safety concerns, as vehicle collisions with wildlife pose dangers not only to animals but also to human travelers (Aoife Moroney, 2018).

The Tanzania-Zambia highway that traverses the Mikumi National Park over a span of 50 kilometers. This thoroughfare underwent rehabilitation between May 1990 and November 1991, culminating in elevated average driving speeds. During this transformative period, meticulous data were collected, documenting the distressing toll of road kills on the park's wildlife. The register, comprising 183 species encompassing mammals, birds, and reptiles,

underscored the staggering scope of the issue, with a minimum of three road kills per day (equivalent to 21-8 road kills/km/year) recorded by the study's conclusion (Drew, 1995).
Amidst the intricate tapestry of challenges posed by the dynamic interaction between transportation infrastructure and wildlife, Geographic Information Systems (GIS) and Remote Sensing emerge as indispensable tools in deciphering this complex conundrum. These cutting-edge technologies serve as beacons of hope, illuminating pathways towards understanding and mitigating the perilous dance between roads and wildlife within the framework of the Mikumi National Park.

GIS, a sophisticated amalgamation of spatial data and analytical techniques, unveils a new dimension of comprehension in the face of wildlife-road carnage. By seamlessly integrating geospatial data encompassing road networks, ecological zones, animal habitats, and human activities, GIS crafts an intricate map of interactions. This cartographic revelation paves the way for an all-encompassing perspective, enabling researchers and stakeholders to delineate high-risk zones and potential collision hotspots (Valerio et al., 2021). Armed with this spatial knowledge, tailored mitigation strategies can be devised to safeguard both animal populations and human travelers, as the interplay between roads and wildlife is unraveled with unprecedented clarity.

In tandem with GIS, Remote Sensing takes to the skies, probing the depths of landscapes from afar. By harnessing satellite imagery and other remote data collection methods, this technology transcends physical boundaries to encapsulate the broader dynamics of Mikumi National Park. Remote Sensing not only aids in identifying critical habitat fragmentation zones but also empowers scientists to observe long-term patterns of animal movement, helping discern where and when road-crossing risks peak (Cibot et al., 2015). The marriage of these technologies furnishes a robust platform for comprehensive risk assessment, serving as a crucial underpinning for strategic interventions.

In the context of the Tanzania-Zambia highway's passage through the park, GIS and Remote Sensing could jointly orchestrate a solution. High-resolution imagery could be wielded to analyze road-kill data, extracting patterns of mortality, while GIS layers could superimpose traffic flow, animal habitats, and landscape characteristics. The resulting synergy would facilitate the identification of precise locations where species vulnerability intersects with vehicular movement (McLennan and Asiimwe, 2016). Armed with this knowledge, targeted measures such as the establishment of wildlife crossings, speed reduction zones, and warning signage could be strategically situated to mitigate the occurrence of wildlife-road collisions (Needham et al., 2020).

In essence, GIS and Remote Sensing unfurl an expansive canvas upon which the intricate brushstrokes of ecological interactions can be meticulously painted. As their synergy unravels the enigmatic tapestry of wildlife-road dynamics, a beacon of hope emerges, guiding the way toward harmony between human progress and the preservation of our natural world.

1.2 Statement of The Problem

Wildlife-vehicle collisions (WVCs) represent a pressing concern within Tanzania's national parks, with Mikumi National Park emerging as a significant corridor for both animal and human movement. Although a recognizable pattern of WVCs exists, the specific factors that underpin these collisions remain elusive. Addressing this, the research aims to unearth the nuanced contributors to WVCs, thereby enhancing our comprehension of their occurrence and fostering targeted mitigation strategies.

The geographical attributes of the road and its surroundings wield substantial influence over WVCs, as they interact with natural habitats, migration routes, and food sources. Animals might be drawn to roadsides in search of sustenance, while traversing roads as they transition between habitats. Simultaneously, the interplay of road design, environmental factors, and vehicle speed forms a complex matrix that significantly shapes the likelihood of WVCs.

Embracing the capabilities of Geographic Information Systems (GIS) technology is pivotal to this endeavor. By transcending the limitations of field observations, GIS broadens the study's scope, enabling the assessment of factors at a scale that is otherwise unattainable due to constraints of distance, time, and resources.

Central to this research is the imperative to evaluate the impact of highways on the incidence of Wildlife-road carnage. Through meticulous analysis, critical areas exhibiting elevated WVC rates will be pinpointed. This strategic insight will galvanize the development of laser-focused mitigation measures, aimed at curtailing these collisions and fostering coexistence between roadways and wildlife.

Ultimately, the research seeks to unravel the enigma of WVCs, shedding light on the hidden facets that contribute to this complex challenge. By merging an exhaustive exploration of contributing factors with the power of GIS technology, the study aspires to lay the groundwork for a safer, harmonious environment for both animals and travelers within Mikumi National Park.

1.3 Objectives

1.3.1 Main Objective

To assess wildlife road carnage hotspots along the highway passing through Mikumi national park using hotspot analysis techniques and Geographic Information Systems.

1.3.2 Specific objectives

The following are the specific objectives of the study,

- **i.** To determine the Topographic visibility of the road.
- ii. To determine the different Landcover factors for wildlife-road carnage.
- **iii.** To perform statistical analysis to determine which variable has the greatest impact on the spatial distribution of wildlife-road carnage.

1.4 Research Questions

- i. Are areas of low driver visibility (topographical) correlated with Wildlife-road carnage?
- ii. Does the roadside land classification category affect the location of wildlife-road carnage?
- iii. Which variables are the greatest predictor of collision risk?

1.5 Significance of the research

This study aims to look into the geographic distribution of wildlife-road carnage, In situations of animal deaths and injuries on the road, hotspot analysis techniques and Geographic Information Systems (GIS) are used to examine the relationship between driver visibility (topography), land classification, and traffic volume.

1.6 Beneficiaries

The results of this study will provide a cartographic depiction of the areas of low driver visibility compared to these Wildlife-road carnage hotspots to provide knowledge that can be implemented in management plans for reducing the risk of collisions.

1.7 Scope and limitations

The accuracy of the research findings may be limited by the availability and quality of data. There may be difficulties in obtaining accurate data on wildlife road fatalities, particularly in remote areas where there is limited access or monitoring capability.

1.8 Study Area

The area to be considered is the Tanzania-Zambia highway that crosses Mikumi National Park over a stretch of 50 km, located near Morogoro, Tanzania with an area of 3,230km sq that was established in 1964. It is thus the most accessible part,with an estimation of 41,666 visitors per 2012 statistics leading to an intense traffic flow, thus resulting to higher average driving speed that facilitates the deaths and injuries of wildlife crossing through the highway.

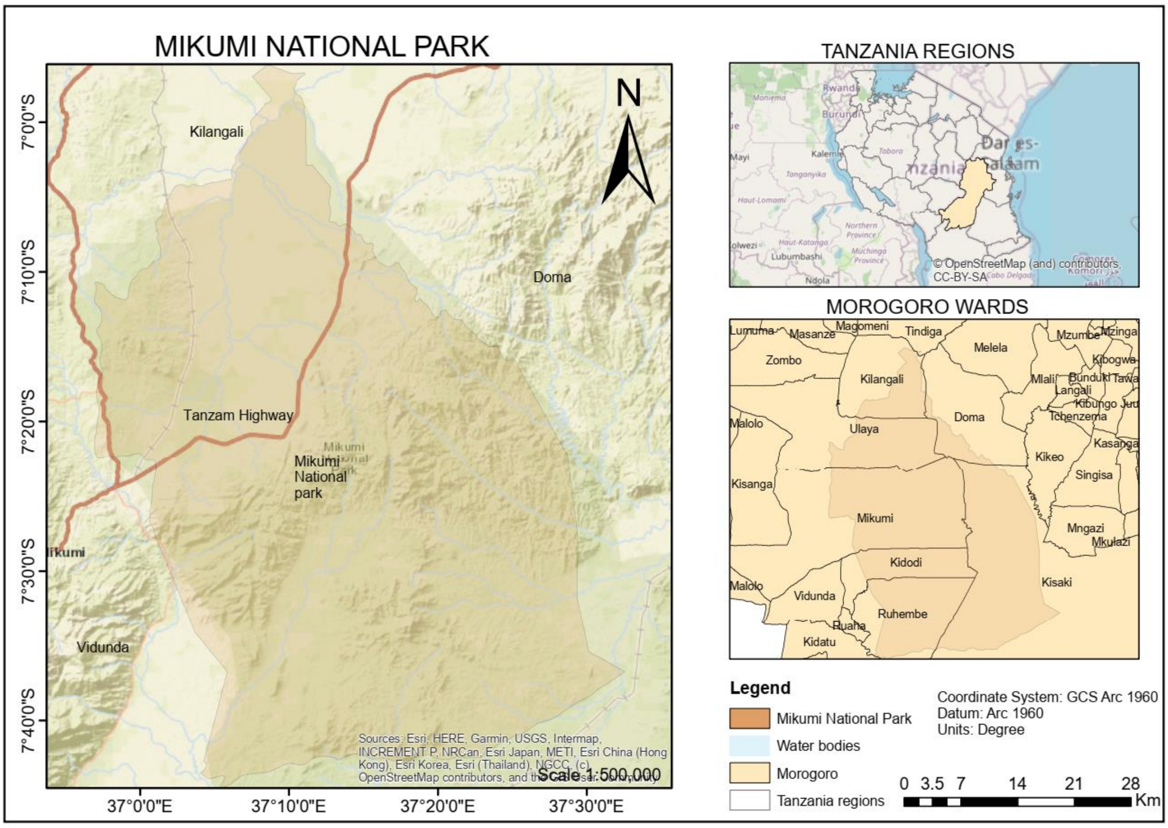

Figure 1-1: Location Map of Mikumi National Park

CHAPTER TWO

LITERATURE REVIEW

2.1 Overview

The sources cited throughout the report are listed in this section. It can simply be a summary of the source of specific information, and it typically follows a set organizational structure. It will provide information on the variety of other studies that are cited and referenced. The literature review also discusses the information that has been published on a subject matter recently

2.2 Hotspost Analysis

Hotspot analysis is a spatial analysis and mapping technique interested in the identification of clustering of spatial phenomena. These spatial phenomena are depicted as points in a map and refer to locations of events or objects.

A hotspot refers to an area with a higher concentration of events than what would be expected if events were randomly distributed. The detection of hotspots originated from the study of point distributions or spatial arrangements of points in a given space (Chakravorty, 1995).

Point patterns are examined by comparing the density of points within a specified area against a complete spatial randomness model, which outlines a process where point events occur in a completely random fashion (also known as a homogeneous spatial Poisson process). In addition to analyzing the density of points in a given area, hotspot techniques also evaluate the degree of interaction between point events to gain insight into spatial patterns (Baddeley, 2010).

Hotspot analysis involves the following type of data:

- Points – Locations of objects/events occurring in a study (e.g., violent trauma, crime, earthquake epicenters, cases of avian influenza, etc.)

Depending on the research question, the hotspot analysis may involve the following type of data:

i. Attributes – Categorical or continuous variable that further describes the objects/events
ii. Period – Date or time of events
iii. Other covariates – Explanatory variables of any kind

Hotspot analysis is a powerful tool for identifying areas with high frequency of wildlife-vehicle collisions (WVCs) and prioritize them for implementing effective measures to reduce the risk of WVCs. The process involves collecting data on WVC locations, creating a spatial point dataset, and using a hotspot analysis tool such as the Getis-Ord Gi* statistic to identify areas with statistically significant clustering of WVCs. The results of the analysis can be visualized using a map, and measures such as wildlife crossings, fences, or reduced speed limits can be implemented in high-risk areas. However, it is important to consider additional factors such as road design and wildlife behavior in developing effective measures to reduce the risk of WVCs (Huijser et al., 2009).

Some basic steps involved in performing hotspot analysis include:

1. Data preparation: This involves collecting and preparing the necessary data for analysis, including geospatial data on the location of WVCs, road networks, and other relevant factors such as land use or topography.
2. Define study area: The study area needs to be defined where the analysis will be conducted. The study area may be defined by administrative boundaries or geographical features.
3. Spatial interpolation: The geospatial data is then interpolated to create a continuous surface, allowing for the identification of clusters of high WVC activity.
4. Hotspot analysis: Once the data has been prepared and the study area defined, hotspot analysis can be conducted to identify areas of statistically significant clustering of WVCs.
5. Interpretation and visualization: The results of the analysis can then be interpreted and visualized to identify areas of high risk for WVCs and inform the development of mitigation strategies.

(Johansson et al., 2019)

2.3 Viewshed Analysis

Viewshed analysis is a GIS spatial analysis technique used to determine the visible and non-visible areas from a given observer point in a terrain or landscape (Wood, 2018).. It is used to analyze the visibility of the terrain from the viewpoint of an observer and to determine the areas that are visible or obscured from the observer's perspective. The analysis takes into account the topography of the landscape, the height of the observer, and any obstructions such as buildings, trees, or other features that block the view. Viewshed analysis is commonly used in various fields, such as urban planning, military, environmental management, and natural resource management, to assess visibility, plan viewpoints, and locate resources. It can also be used to evaluate the impact of new developments or infrastructure on the surrounding landscape.

Viewshed analysis can provide valuable information on the line-of-sight visibility of drivers along roadways. This information can be used to identify areas where drivers may have difficulty seeing wildlife, such as curves or hills. By analyzing the line-of-sight visibility of drivers, viewshed analysis can help identify areas where wildlife crossings or fencing could be installed to guide animals safely across the road. Viewshed analysis can also be used to assess the impact of roadway modifications, such as road widening or tree removal, on the visibility of wildlife along the roadway (Barr et al., 2016).

The process of line-of-sight analysis involves creating a map of visible areas. If two points, A and B, do not have their line of sight intersect with any terrain, then they are said to be mutually visible. To determine this, the slope of the line between A and B is calculated using the formula$(Z_b - Z_a)/d_{AB}$ where z represents the height and d represents the horizontal distance. If the line does not cross any point C with a slope larger than the slope of AB, then points A and B are considered visible to each other (as shown in Figure 2.1). When using a raster DEM, the viewshed of point A is the set of all cells of the DEM that are visible from A. The height of observer A above the terrain has a significant impact on the extent of the viewshed, and can be adjusted to model different viewpoints such as a standing person, the view from a multi-story building, or the view from a cell tower. The resulting viewshed can be represented as either a binary raster containing only 0s and 1s or as a raster with values of vertical angles in relation to the viewpoint.

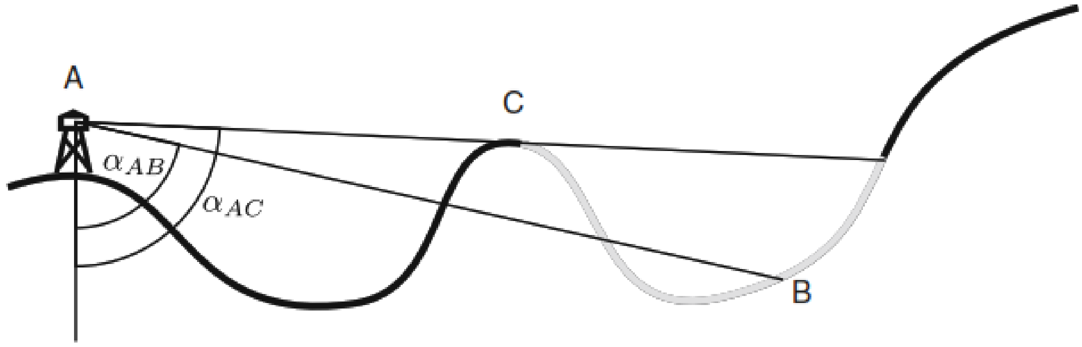

Figure 2-1 Line of sight analysis: point B is not visible from point A because the line of sight between A and B has a smaller slope tan AB than the line of sight between A and C.

One way to evaluate the visibility of an area is to calculate the viewsheds from multiple viewpoints and combine them into a cumulative viewshed, which provides a comprehensive view of the overall visibility of the area (Wheatley, 1995). The use of visibility analysis has been extended to a wide range of geospatial applications, such as landform identification using geomorphons that utilizes visibility analysis with computer vision (Jasiewicz & Stepinski 2013). Additionally, numerous visualization techniques for terrain relief employ visibility analysis to emphasize subtle terrain features. For instance, the sky-view factor, which reveals the visible part of the sky that is not obscured by relief, can be used to render topographic relief as though it were uniformly and diffusely illuminated, thereby clearly illustrating the relative height of features.

2.4 Image Classification

Image classification is the process of assigning pixels to nominal, which results to the thematic classes (Mather & Koch, 2011). The principle of image classification is that a pixel is assigned to a class based on its feature vector by comparing it to the predefined clusters in the feature space where by doing so all image pixels results in a classified image. This is also a process in which the (human) operator instructs the computer to perform an interpretation according to certain conditions. Image classification is based on the different spectral characteristics of different materials on the earth's surface.

Spectral pattern is a set of radiance measurements from various wavelength bands for each pixel. Classification procedures can be based on Spectral pattern (spectral pattern recognition), Spatial patterns (spatial pattern recognition), Temporal patterns (temporal pattern recognition). Spectral pattern recognition uses pixel-by-pixel spectral information as a basis for automated classification. (Rehna & Natya, 2016).

1. PRINCIPLE OF IMAGE CLASSIFICATION

   By comparing a pixel's feature vector to predetermined clusters in the feature space, a class is assigned to it based on the feature vector. A categorized image is produced by doing this for each and every image pixel. Comparing an image to specified clusters is the essence of image classification, which calls for the definition of clusters and comparison techniques. During the training phase, cluster definition takes place in an interactive manner. Using classifier techniques, the clusters and individual pixels are compared (Rehna & Natya, 2016).

2. CLASSIFICATION SCHEME

   This shows how the classes will be chosen during the process of image classification. There are several classification schemes and one of them is Anderson's classification scheme. This was developed for the use with remote sensing data both aircraft and satellite based. The advantages of this are can be used for many applications by selecting the level of the detail desired and many of the classes are not separable over large areas using remote sensing observations. (Rehna & Natya, 2016)

Levels of Anderson classification scheme are;

- i. Level one Urban built up areas, Agriculture, Rangeland, forest, water areas
- ii. Level Two Residential commercial, industrial, croplands, and pasture.
- iii. Level three Single-family units and multifamily units.

3. CLASSIFICATION ALGORITHMS

A computational methods or procedures used in machine learning and data analysis to categorize data into predefined classes or categories based on the features or attributes of the data. The main goal of a classification algorithm is to learn a mapping function from input data to output labels, enabling the algorithm to make accurate predictions or decisions about the class membership of new, unseen data points.

Common types of classification algorithms include;

1. Decision Trees and Random Forests

Decision trees are a widely used classification algorithm that partitions data based on feature values. They recursively split the dataset into subsets by selecting the feature that provides the best separation. Random Forests, an ensemble method of decision trees, aggregate the results of multiple decision trees to improve accuracy and reduce overfitting. Random Forests are particularly effective for handling high-dimensional and noisy data.

2. Support Vector Machines (SVM)

SVM is a powerful classification algorithm that finds a hyperplane that best separates different classes while maximizing the margin between them. SVM can handle linear and nonlinear classification tasks and has been successfully applied to tasks like text classification and image recognition.

3. k-Nearest Neighbors (k-NN)

k-NN is a simple instance-based classification algorithm that assigns a new data point to the majority class among its k-nearest neighbors in the training dataset. k-NN is sensitive to noise and requires careful selection of k and distance metrics. It's used in recommendation systems, image classification, and anomaly detection.

4. Naive Bayes

Naive Bayes is a probabilistic classification algorithm based on Bayes' theorem. It assumes that features are conditionally independent given the class label. Naive Bayes is computationally efficient and often performs well in text classification, spam filtering, and sentiment analysis.

5. Neural Networks and Deep Learning

Neural networks, particularly deep learning models, have revolutionized the field of classification. Convolutional Neural Networks (CNNs) excel at image classification by automatically learning relevant features from raw pixel data. Recurrent Neural Networks (RNNs) are used for sequence classification tasks like natural language processing.

6. Ensemble Methods

Ensemble methods combine multiple base classifiers to improve overall performance and robustness. Apart from Random Forests, Gradient Boosting (e.g., XGBoost, LightGBM) and AdaBoost are popular ensemble techniques that sequentially build strong classifiers from the mistakes of weak classifiers.

The basic steps in performing land classification typically include the following

1. Define the study area: This involves identifying the boundaries of the area of interest and selecting an appropriate scale for analysis. The study area should be large enough to capture the relevant spatial and temporal variability but not so large as to make data collection and analysis impractical (Jensen, 2013).

2. Select input data: Land classification requires data on the characteristics of the landscape, which can be obtained from a variety of sources, including remote sensing data, ground surveys, and existing maps. The selection of appropriate input data depends on the goals of the classification and the availability of data (Jensen, 2013).
3. Preprocess data: This step involves correcting, enhancing, and transforming input data to remove noise, reduce errors, and facilitate analysis. Preprocessing techniques include image correction, filtering, and normalization (Jensen, 2013).
4. Define classification scheme: A classification scheme is a set of rules or criteria used to assign pixels or areas to different land cover classes. Classification schemes can be based on various criteria, such as spectral reflectance, texture, shape, and context. The choice of a classification scheme depends on the characteristics of the input data and the goals of the analysis (Lu, 2018).
5. Train and validate classifier: The classification scheme is applied to the input data using a machine learning algorithm, such as maximum likelihood, decision tree, or support vector machine. The algorithm is trained using a set of ground truth data, which are samples of known land cover classes, and validated using an independent set of data. The accuracy of the classification depends on the quality and representativeness of the ground truth data, the appropriateness of the classification algorithm, and the validity of the assumptions underlying the classification scheme (Lu, 2018).
6. Post-classification processing: This step involves refining the classification results by filtering out the noise, filling gaps, and merging or splitting classes as needed. Post-classification processing techniques include majority filtering, contextual smoothing, and object-based analysis (Jensen, 2013).
7. Evaluate and interpret results: The classification results should be evaluated in terms of their accuracy, completeness, and consistency with the goals of the analysis. The results should also be interpreted in the context of the study objectives and the characteristics of the landscape. Interpretation of the results can be facilitated by visualization tools, such as thematic maps and change detection analyses (Lu, 2018).

2.5 Google Earth Engine (GEE)

Google Earth Engine (GEE) is a cloud-based geospatial processing platform for large-scale environmental monitoring and analysis. The free-to-use GEE platform provides access to petabytes of publicly available remote sensing imagery and other ready-to-use products with an explorer web app; high-speed parallel processing and machine learning algorithms using Google's computational infrastructure; and a library of Application Programming Interfaces (APIs) with development environments that support popular coding languages, such as JavaScript and Python.

Google Earth Engine provides a powerful and flexible platform for performing land classification and other geospatial analyses, particularly for large-scale and time-series datasets (Gorelick et al., 2017).

Features and capabilities

GEE offers a wide range of features and capabilities, including:

i. Access to a vast amount of geospatial data, including satellite imagery, weather data, and land cover data.
ii. The ability to analyze this data to track changes over time, identify patterns, and create maps and visualizations.
iii. A user-friendly interface that makes it easy to use even for users with no prior experience with geospatial data analysis.
iv. The ability to run complex analyses without the need for specialized software or hardware.

Applications

GEE has been used for a wide range of applications in remote sensing, including:

i. Monitoring environmental change, such as deforestation, desertification, and climate change.
ii. Tracking the spread of diseases, such as malaria and dengue fever.
iii. Studying the impact of human activities on the environment, such as urban development and agriculture.
iv. Planning for future development, such as infrastructure projects and disaster response.

Advantages and disadvantages

GEE offers a number of advantages for remote sensing, including:

i. Access to a vast amount of data
ii. The ability to analyze data at a variety of scales
iii. The ability to run complex analyses without the need for specialized software or hardware
iv. A user-friendly interface

However, GEE also has some disadvantages for remote sensing, including:

i. The data is not always accurate or up-to-date
ii. The platform can be slow at times
iii. The platform can be complex to use for users with no prior experience with geospatial data analysis

Google Earth Engine (GEE) performs classification process using a variety of machine learning algorithms, including:

1. Supervised classification: This algorithm uses labeled training data to classify new data. In the context of GEE, labeled training data consists of satellite images that have already been classified by a human expert. The GEE algorithm uses the labeled training data to learn the relationships between the spectral characteristics of different land cover types and their corresponding labels. Once the algorithm has learned these relationships, it can be used to classify new satellite images.
2. Unsupervised classification: This algorithm does not use labeled training data. Instead, unsupervised classification algorithms identify clusters of pixels in a satellite image that have similar spectral characteristics. The clusters are then labeled by the user. Unsupervised classification is often used to identify land cover types that are difficult to classify using supervised classification, such as wetlands and grasslands.

Here are some of the steps involved in the classification process in GEE:

i. Data preparation: The first step is to prepare the data for classification. This includes preprocessing the data, such as converting it to a format that is compatible with GEE, and creating training data.

   Code snippet

```
// Import the Sentinel-2 satellite imagery.
var sentinel2 = ee.ImageCollection("COPERNICUS/S2");
// Convert the Sentinel-2 satellite imagery to a format that is compatible with GEE.
var sentinel2_rgb = sentinel2.select('B2', 'B3', 'B4');
```

ii. Feature extraction: The next step is to extract features from the data. Features are the characteristics of the data that are used to classify it. For example, in the case of land cover classification, features could include the spectral reflectance of the land cover, the texture of the land cover, and the elevation of the land cover.

Code snippet

```
// Extract the spectral reflectance of the Sentinel-2 satellite imagery.
var sentinel2_features = sentinel2_rgb.reduceBands(ee.Reducer.mean());
```

iii. Model training: The third step is to train a model. The model is a mathematical function that learns to classify the data based on the features that have been extracted.

Code snippet

```
// Train a Random Forest model to classify the Sentinel-2 satellite imagery.
var sentinel2_model = sentinel2_features.randomForestClassifier(200);
```

iv. Model testing: The fourth step is to test the model. This is done by using the model to classify a set of data that was not used to train the model. This helps to ensure that the model is not overfitting the data.

Code snippet

```
// Test the Sentinel-2 model on a set of data that was not used to train the model.
var sentinel2_predictions = sentinel2_model.classify(sentinel2_rgb);
```

v. Model deployment: The fifth and final step is to deploy the model. This means making the model available for use by other users.

Code snippet

```
// Deploy the Sentinel-2 model so that it can be used by other users.
var sentinel2_classifier = ee.Classifier(sentinel2_model);
```

(Haifa Tamiminia et al,2020)

2.6 Random Forest

Random Forest is an ensemble learning method that combines multiple decision trees to make predictions. It utilizes the concept of "bagging" (bootstrap aggregating) to create diverse training datasets through random sampling with replacement. Each decision tree in the ensemble is built independently, and the final prediction is obtained through a majority vote or averaging of the individual tree predictions (Breiman, 2001).

Algorithm Advantages

i. Robustness: Random Forest is known for its robustness against overfitting, a common problem in machine learning. The ensemble of decision trees reduces the risk of overfitting by averaging out individual tree biases.

ii. Variable Importance: Random Forest provides a measure of variable importance, indicating the relative contribution of each input variable in the classification or regression task. This feature allows for feature selection and understanding of the key variables driving the prediction.

iii. Non-linearity Handling: Random Forest can effectively handle non-linear relationships between input variables and the response variable. It captures complex interactions and non-linear patterns in the data, making it suitable for various applications.

iv. Outlier Tolerance: Random Forest is relatively robust to outliers in the training data. The ensemble approach helps reduce the impact of individual outliers on the final prediction.

Applications

i. Remote Sensing: Random Forest has been widely applied in remote sensing for land cover classification, vegetation mapping, change detection, and object detection. Its ability to handle multi-dimensional data and exploit spectral, textural, and contextual information makes it a valuable tool in analyzing remotely sensed imagery.

ii. Ecological Modeling: Random Forest has shown promising results in ecological modeling, including species distribution modeling, biodiversity assessment, and habitat suitability mapping. It can effectively capture complex relationships between environmental variables and species occurrences or ecological processes.

iii. Data Science: Random Forest is extensively used in data science applications, such as customer segmentation, fraud detection, sentiment analysis, and recommendation systems. Its ability to handle high-dimensional data, missing values, and categorical variables makes it versatile in diverse data-driven tasks.

Limitations and Challenges

i. Interpretability: While Random Forest provides accurate predictions, the interpretability of individual decision trees within the ensemble may be limited. Understanding the underlying mechanisms and relationships driving the predictions can be challenging.

ii. Overfitting Risk with Noisy Data: Although Random Forest is robust against overfitting, noisy datasets can still affect its performance. Careful data preprocessing and feature selection are necessary to mitigate this issue.

iii. Computational Intensity: Random Forest can be computationally intensive, especially for large datasets with a high number of trees and input variables. Efficient implementation and parallelization techniques are required to handle large-scale problems.

(Hastie et al., 2009).

2.7 Generalized Linear Mixed Model

Generalized linear mixed model (GLMM) is a type of statistical analysis used to model data with both fixed and random effects. It is an extension of the generalized linear model (GLM) that takes into account the correlation between observations due to clustering or repeated measures. GLMMs are commonly used in fields such as biology, psychology, and economics to analyze data that exhibit non-normal distributions or non-constant variance. GLMMs are a powerful tool for analyzing complex data structures and are commonly used in longitudinal studies, cluster randomized trials, and observational studies (Zhang et al., 2020).

History and development

GLMMs were first developed in the early 1970s by John Nelder and Trevor Hastie. Nelder and Hastie were working on a problem in ecology, and they wanted to develop a model that could be used to analyze data that was both dependent and random. They developed GLMMs by combining the concepts of generalized linear models (GLMs) and mixed models.

Features and capabilities

GLMMs offer a number of features and capabilities that make them a powerful tool for data analysis. Some of the features and capabilities of GLMMs include:

i. Can be used to analyze data that is both dependent and random: GLMMs can be used to analyze data that is related to each other, as well as data that is not related to each other. This makes GLMMs a versatile tool that can be used to analyze a wide variety of data.

ii. Can be used to model a variety of response variables: GLMMs can be used to model a variety of response variables, including binary data, count data, and continuous data. This makes GLMMs a powerful tool for a wide range of applications.

iii. Can account for covariates: GLMMs can account for covariates, which are variables that may be related to the response variable. This allows GLMMs to provide more accurate estimates of the effects of the variables of interest.
iv. Can be used to make predictions: GLMMs can be used to make predictions about the response variable. This can be useful for decision-making or for planning future studies.

(Demidenko, 2013)

Applications

GLMMs have been used in a wide variety of applications, including:

i. Clinical trials: GLMMs have been used to analyze data from clinical trials, such as trials for new drugs or treatments. GLMMs can be used to assess the efficacy and safety of new treatments.
ii. Observational studies: GLMMs have been used to analyze data from observational studies, such as studies that investigate the relationship between smoking and lung cancer. GLMMs can be used to control for confounding factors, such as age and gender, in observational studies.
iii. Surveys: GLMMs have been used to analyze data from surveys, such as surveys that investigate the prevalence of a disease or the level of satisfaction with a service. GLMMs can be used to account for the clustering of data in surveys.

(Demidenko, 2013)

Advantages and disadvantages

GLMMs offer a number of advantages, including:

i. Can be used to analyze data that is both dependent and random: GLMMs are a versatile tool that can be used to analyze a wide variety of data.
ii. Can be used to model a variety of response variables: GLMMs can be used to model a variety of response variables, including binary data, count data, and continuous data.
iii. Can account for covariates: GLMMs can account for covariates, which are variables that may be related to the response variable. This allows GLMMs to provide more accurate estimates of the effects of the variables of interest.
iv. Can be used to make predictions: GLMMs can be used to make predictions about the response variable. This can be useful for decision-making or for planning future studies.

However, GLMMs also have some disadvantages, including:

i. Can be complex to understand and use: GLMMs are a complex statistical model, and they can be difficult to understand and use.
ii. Can be computationally demanding: GLMMs can be computationally demanding, especially for large datasets.
iii. Can be sensitive to model misspecification: GLMMs can be sensitive to model misspecification, which can lead to biased estimates.

In a GLMM, the response variable y is modeled as a function of the fixed effects x and the random effects u:

$y = f(x,u) + e$

where e is an error term, and f is a link function that relates the mean of y to the fixed and random effects. The random effects are assumed to have a distribution with a mean of zero and a covariance matrix that models the correlation among the observations (Zhang et al, 2020).

The GLMM is a powerful and flexible modeling framework that can handle a wide range of data structures, including clustered, longitudinal, and multilevel data. It can also handle a variety of response variable distributions, such as binomial, Poisson, and normal (Bates et al., 2015).

CHAPTER THREE

METHODOLOGY

3.1 Overview

This section explains and assesses all of the approaches and procedures used for the research. A mixed methodology that incorporates both qualitative and quantitative methodologies is the one that was used. As shown below in Figure 3-1 the research initiated with Data acquisition from online sources as well as provided data which in all included satellite images, digital elevation models, wildlife-vehicle collision data, road segment data and Traffic volume data. Then followed by processing of the data to obtain the outputs that were to be used for the analysis and achieving the objectives of the project.

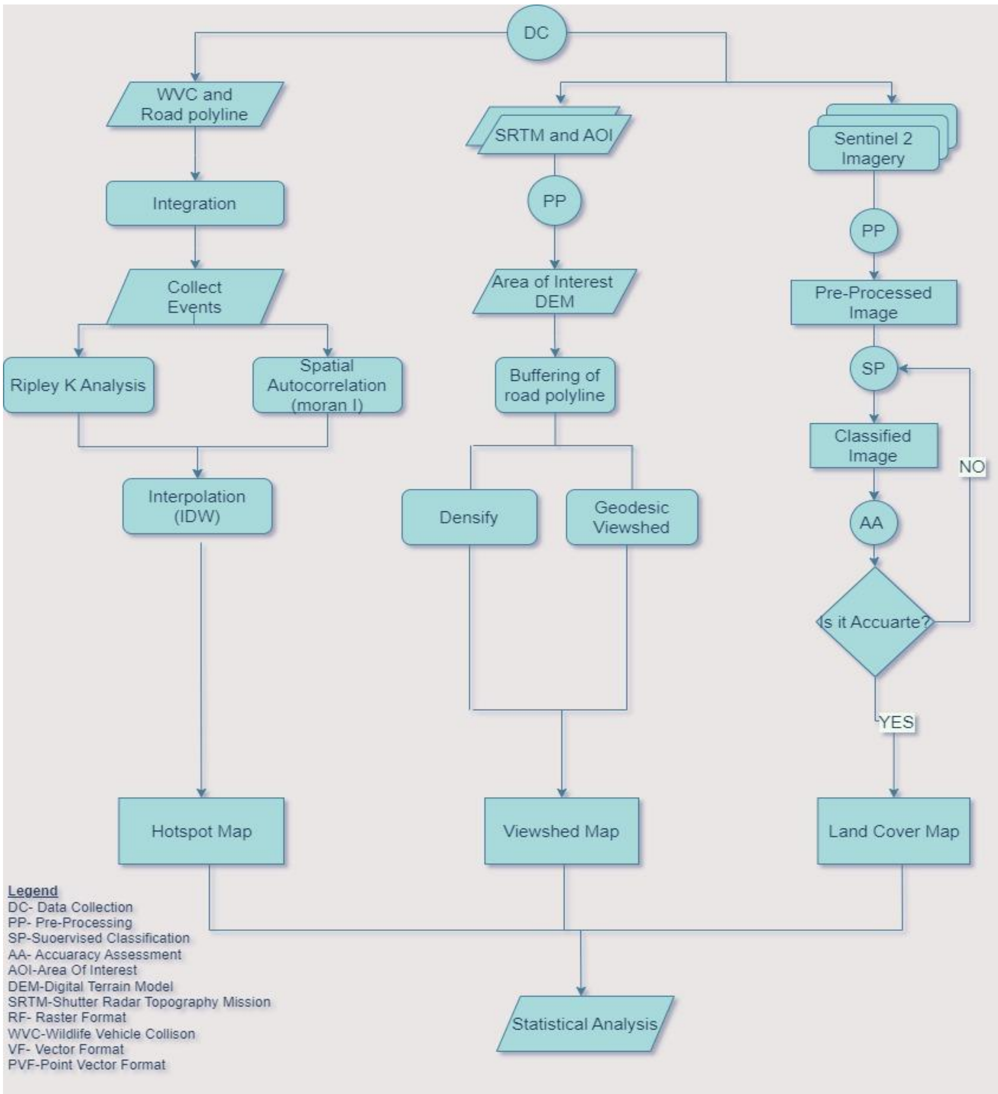

Figure 3-1: Methodology workflow

3.2 Data collection

Acquisition of data was the first step to be performed where in this study different types of data were collected from different sources which included sentinel-2 images downloaded from Google Earth Engine and the DEM images downloaded from USGS website as well as provided data which included, wildlife-vehicle collision data, road segment data and Traffic volume data.

Below shows a summarized detail of the data collected for the conduction of the research.

Table 3-1: Datasets, their source, resolution and year of acquisition

Datasets	Source	Format	Year	Use
Wildlife road killing data	MIKUMI NATIONAL PAK	Csv and Excel	2018-2022	For identifying the death and injury of wildlife
Road segment data	TANROAD	Shapefile		For determining the extent of the road
Sentinel-2 imagery	GEE	Tiff	2018-2022	To create land cover and change maps
SRTM DEM	USGS	Tiff	2014	To understand the topography of the road

3.3 Data collection of satellite images and preparation

The satellite images were obtained from open-source site where anyone from anywhere could easily access, GEE code editor was the engine at which provided free satellite images and allowed a download option of the desired quality. First, an account was created with GEE with all relevant instruction on how to activate the account since it does not allow data access without joining their site for further monitoring and prevent miss use of data, after an account was fully set GEE code editor provided full access to the site. In order to acquire data from GEE code editor the following steps were followed

i. Setting of search criteria

The first step taken was to add the zipped shapefile of Mikumi national park as the area of concentration where my data collection was based at, the selection of the region of interest that updates its coordinated automatically

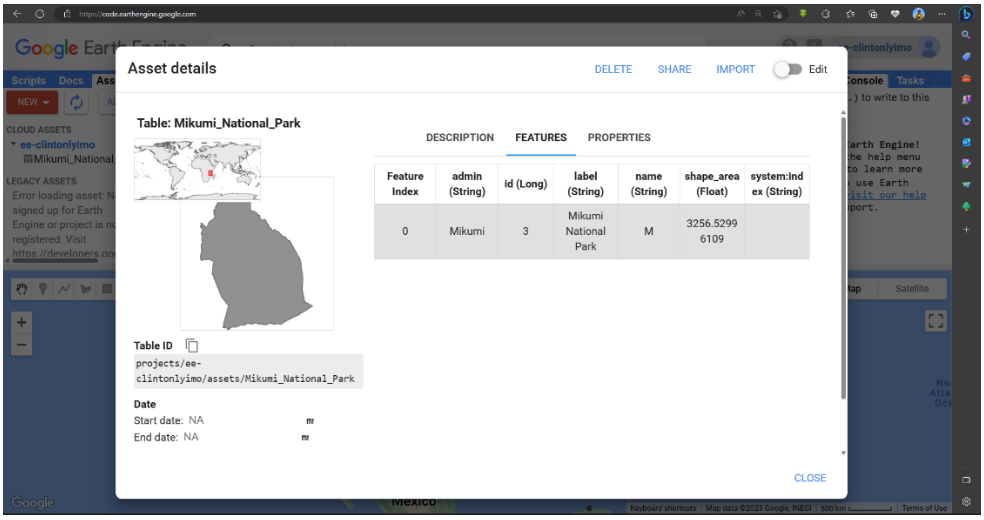

Plate 3-1: Imported AOI (area of interest) in GEE code editor

(Source: [Supervised Classification | Google Earth Engine | Author])

ii. Selection of data to be acquired

Data selection was the important part on requiring the necessary satellite images, there were various of specification that have to be made to archive a clean satellite image;

First, the specific time of data to be collected or the range at which data set should be collected was set by which was from 2018 to 2022, this also involved setting the necessary months at which cloud cover was at least to a minimum range. Then the amount of cloud cover was set to filter the data set to satellite images with fewer number of cloud cover at which the value that was less than or equal to ten percent.

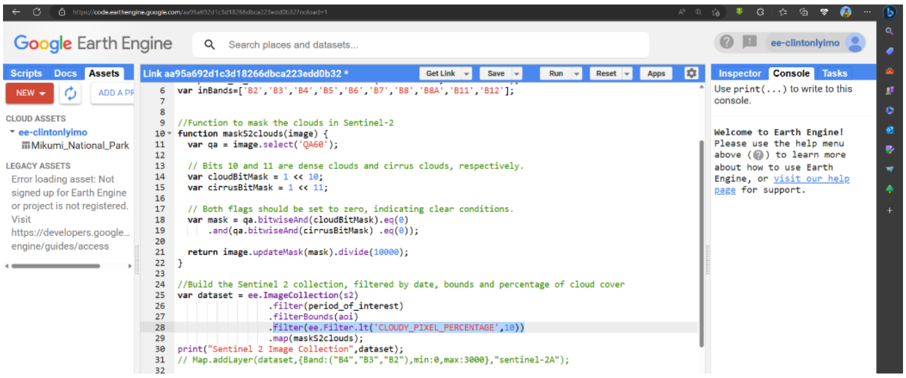

Plate 3-2: Cloud cover range value selected

(Source: [Supervised Classification | Google Earth Engine | Author])

iii. Set selection and additional filtering

A data set source was selected to justify the data location and type of satellite that was used for the collection of satellite images, since the data set time was focused from 2018 to 2022 the reasonable satellite to be selected was sentinel-2 having spatial resolution of 10m.

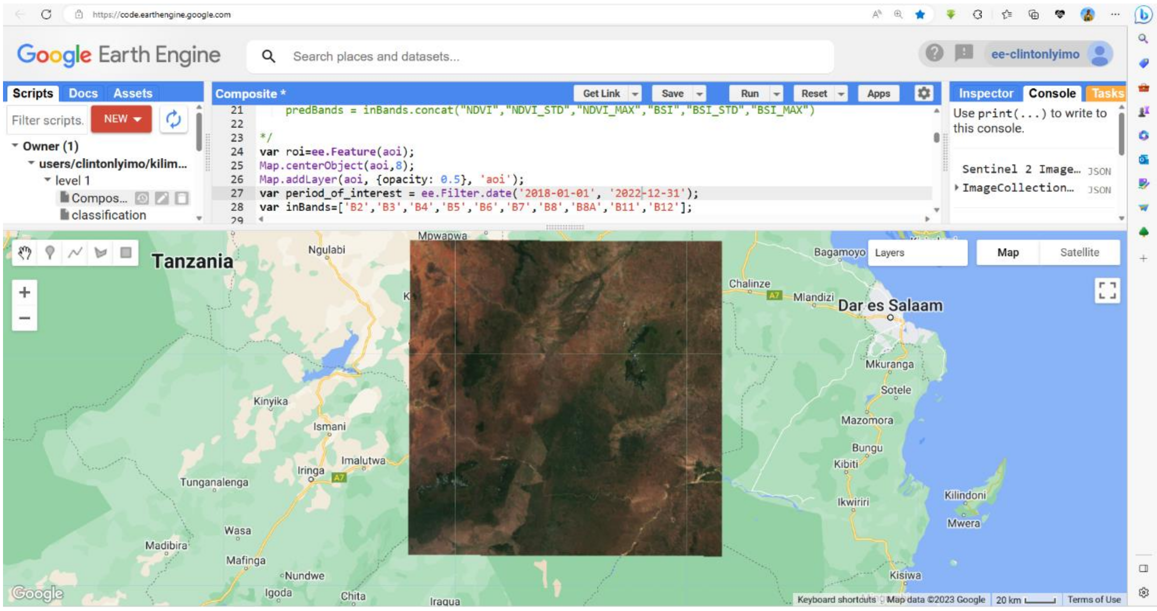

Plate 3-3: Qualified sentinel image of year 2018-2022

(Source: [Supervised Classification | Google Earth Engine | Author])

iv. Ground Truthing Data

Ground truthing data were obtained from Google Earth Pro, which provides access to high resolution current and historical satellite images. The data collected were of the dry season corresponding to the satellite images. Google Earth Pro was synchronized with GEE, which helped to relate the satellite image displayed at different color composites with the high-resolution images to help increase the accuracy of the collected ground truthing data.

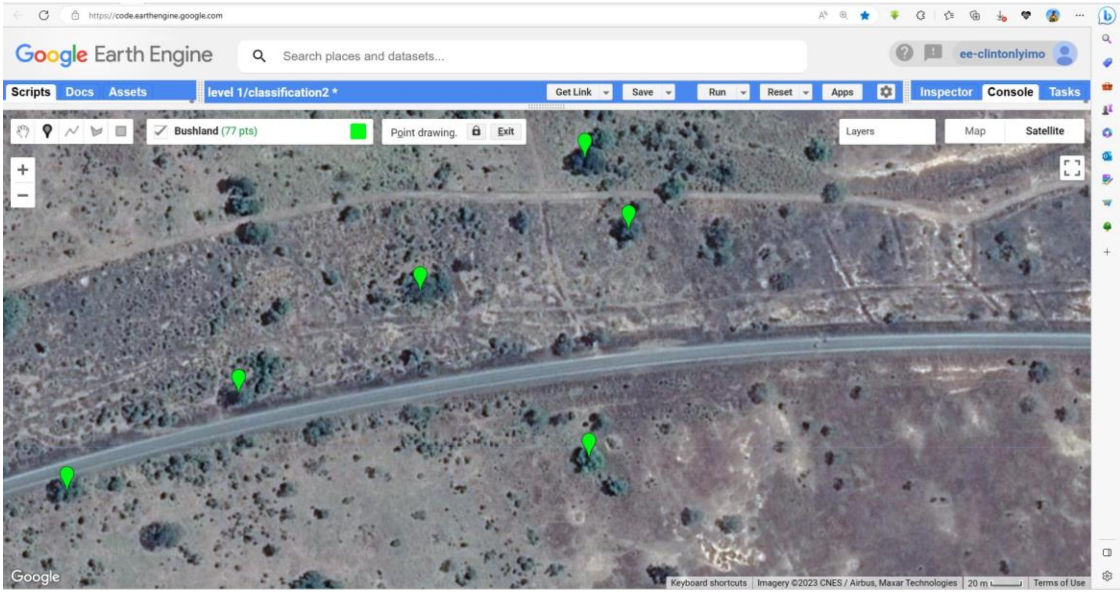

Plate 3-4: Ground truthing data collection

(Source: [Supervised Classification | Google Earth Engine | Author])

3.3.1 Image Pre-Processing and Procession

A field was created called "LandClass", this was used in reclassification with the Six categories chosen. The land classification categories selected for this analysis were Dense Vegetation, Waterbodies, Built-ups, Bushland, Grassland and Bareland. The dataset-specific catalogues and legends were used to reclassify the attribute numerical codes of each landcover type or non-forest feature into one of the respective Six categories.

Data pertaining to orchards, nurseries, crops, or vineyards were attributed to the Grassland category. All forested regions, including regenerating, commercial, or experimental forests were combined under the Dense Vegetation class. Greenspaces within urban centers, as well as infrastructural components such as railways, roads, highway stops, and bridges were included in the Built-up classification. The Water-bodies category was composed of lakes, rivers, retention/treatment ponds, and other surface water bodies, while empty spaces were fit by Bareland category and areas thats had shrubs and small bushes of trees were included in Bushland class.There were no areas that did not fit into these categories.

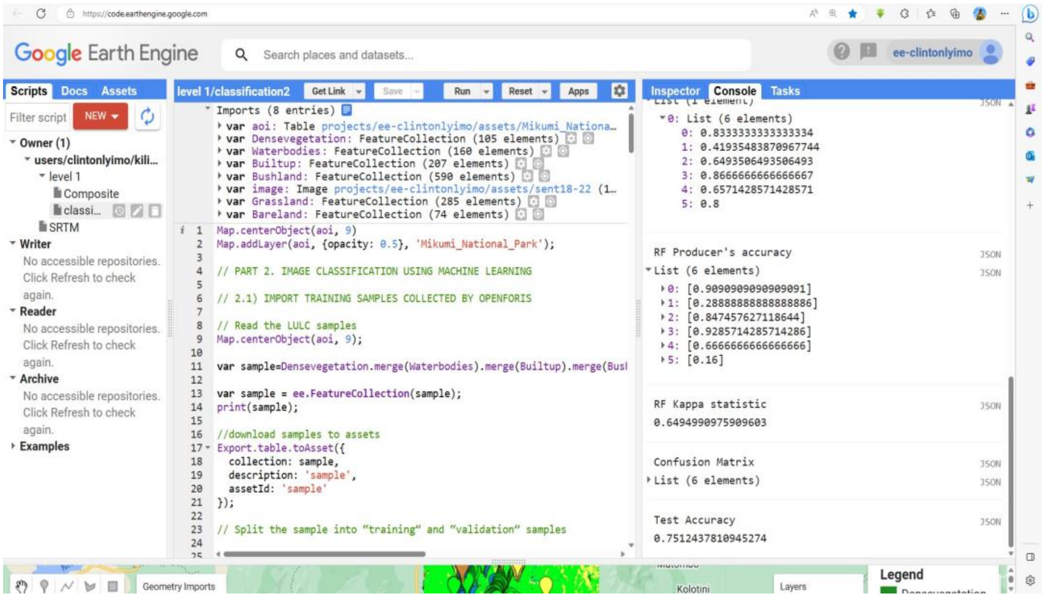

Plate 3-5: Land classes as mentioned above respectively

(Source: [Supervised Classification | Google Earth Engine | Author])

3.3.2 Image Classification based on Spectral Indices

After the Sentinel image were pre-processed to obtain bands with spectral information, and an image with filled gaps for the sake of 2018-2022 Sentinel image. The next step was to map the land cover of the study area. In order to achieve this goal, a number of procedures was followed.

i. Land Cover type determination

Land cover classes chosen to be classified based on the National Land Cover Database (NLCD) Classification System, this classification scheme helped to identify the land cover classes that can accurately classified using a Sentinel image. Before concluding the type of land cover to be mapped, high resolution images from google earth engine were used to gain familiarity with the area of interest.

ii. Calculation of spectral indices and layer stacking

Spectral indices have been developed in order to effectively separate a single feature from others effectively. SI's have been used in several classification studies and they have proven to be effective in improving classification accuracies. Lee et al (2018) used spectral indices derived from Sentinel and stacked the SI's to the original image, this proved to improve the classification accuracy for all six mapped classes (Waterbodies, Dense Vegetation,Bushland,Grassland, Built up and Bare land). Also Mete (2013) used a decision tree approach for land cover mapping while incorporating the use of vegetation indices, the use of vegetation indices together with the original image proved to increase the class separability of the trained pixels. Selection of the SI's strongly depends on the nature of the land cover in your area of interest, one has to choose SI's that will help separate the chosen classes from one another. In this study four land cover types were mapped and the spectral indices used were Normalized Vegetation Index (NDVI) and Bare Soil Index (BSI).

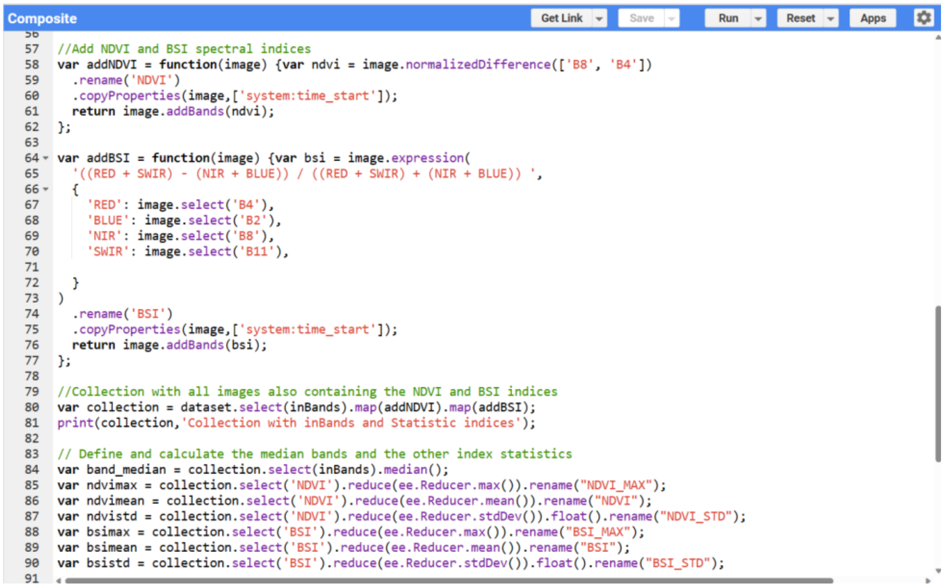

Plate 3-6: Showing both NDVI and BSI code representation

(Source: [Supervised Classification | Google Earth Engine | Author])

iii. Classification Algorithm

Creation of thematic maps showing the nature of land cover was done by classifying Landsat images using Random Forest (RF) classification algorithm. RF classifier was used due to its ability to treat individual pixels as a mixture of pure materials and end members in the classification process. The classifier sub-divided each individual pixel data to increase the spectral variance of different features within the pixels for superior and meaningful land covercomposition as well as improved classification accuracy. The RF library developed in GEE was used to execute the RF algorithm and the classification maps were prepared using the ArcGIS software version 10.8.

iv. Buffering

The analysis included the utilization of the Multiple Ring Buffer tool, where the Roads polyline layer was employed as the input, and linear unit distances of 0.5, 1.0, and 5.0 kilometers were selected. These distance values were chosen based on the recommendations of Benítez-López et al. (2010) to accommodate the distinct mobility and land use categories associated with each functional group (as outlined in Table 3.1). The default dissolve option was maintained, resulting in the consolidation of concentric rings into three distinct regions. Lastly, the LandClass Raster was clipped to the 5.0-kilometer buffer, aligning with the largest buffer distance used in the analysis.

Table 3-2: Road buffer distances associated with each functional group

Distance (Km)	Functional Groups
0.5	Amphibians and Reptiles
1	Small-Medium Vertebrates
	Birds
5	Carnivores
	Ungulates, Primates

3.4 Viewshed Analysis

To begin analysis of driver visibility, a 200m buffer was created for the Roads polyline layer composed one set of roads in Mikumi National Park. A 30m DEM for the study area was sourced from the https://earthexplorer.usgs.gov/

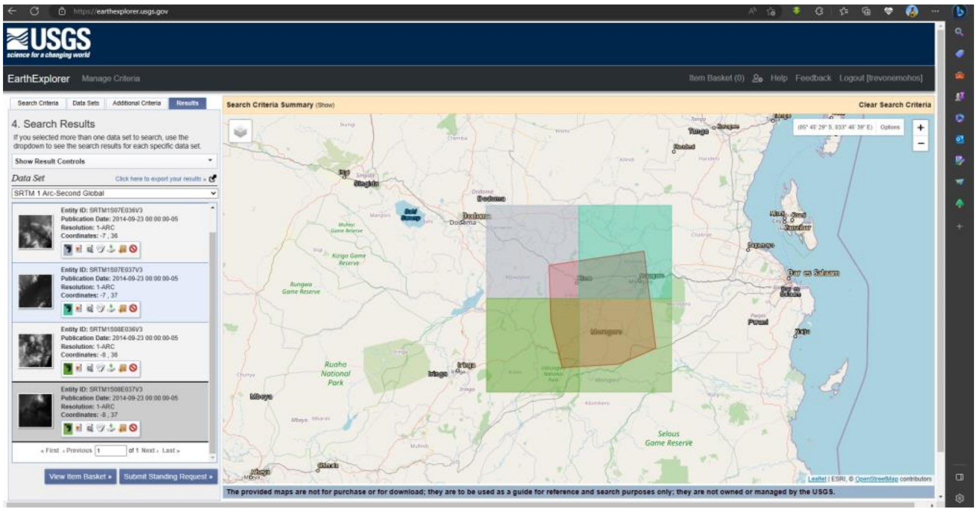

Plate 3-7: Mikumi national park DEM

(Source: USGS, 2000| Author)

The clipped DEM, named (RoadBufferDEM_200m_clipped), served as the input raster for the subsequent analysis. The Geodesic Viewshed tool provided by ESRI was employed to perform the analysis on the road segment. Prior to running the tool, the Densify tool was applied to each individual road layer to add vertices at 1-kilometer intervals along the entire length of the road. These newly added vertices, as well as any existing vertices from the original road dataset in Mikumi National Park, were used as the observation locations when the "Frequency" option was selected in the Densify tool.

In accordance with a GIS analysis of road visibility by Abdulhafedh (2019), the observer height offset in the ‘Geodesic Viewshed Tool’ was set at 1.08m to represent the average height of a driver in their vehicle. The surface offset was set at 0.6m, as this was determined in previous publications to be the point a stationary object is visible from average driver position Abdulhafedh (2019). The horizontal plane of visibility ranged from 0° to 180° with a full vertical ranger from -90° to 90°. The inner radius was 0, assuming immediate visibility of items directly in front of the driver, while the outer radius was applied as 560m, as standard with the American Association of State Highway and Transportation Officials Abdulhafedh (2019). The analysis was completed using “All Sightlines”. The output viewshed raster layers were labelled with “Viewshed”. The visibility was represented by a count showing the number of times each individual cell was visible from the input polyline.

Plate 3-8: Geodesic Viewshed tool (ESRI, 2021)

(Source: ArcGIS 10.8| Author)

3.5 Ripley’s K Analysis using Roads and WVC

To perform Ripley's K-analysis in ArcGIS 10.8, the following steps were followed:

Data Preparation:

1. The road polyline data for the study area was obtained and imported into ArcGIS 10.8.
2. The road polyline data was carefully reviewed and edited to ensure its quality and accuracy.
3. The provided RoadSegmentExportPointFile Python Script was used to convert the road features into TXT files, which were compatible for further analysis.

Wildlife-Vehicle Collision (WVC) Data:

1. WVC point data was collected from Mikumi National Park, including information on the latitude, longitude, and collision outcomes (e.g., mortality).
2. The Add Data tool in ArcGIS 10.8 was used to import the WVC point data into the project.
3. The WVC points were categorized into functional groups, such as Carnivore, Ungulate, Small-Medium Vertebrate, Amphibians and Reptiles, Primates and Birds, as well as different collision outcomes.

Selecting the Study Area:

1. The extent of the study area was defined, and specific road segments for analysis were identified.
2. The Select by Location tool in ArcGIS 10.8 was utilized to select WVC points located within a specified distance, such as 1.0 kilometer, from the study roads.
3. The selected WVC points were exported to Excel sheets using the Table to Excel conversion tool.
4. The exported data was reviewed in Microsoft Excel to ensure accuracy and identify any potential errors or inconsistencies.

3.5.1 Ripley's K-Analysis:

The Ripley's K Test was run for general WVC collisions using a 500m initial radius with 500m radius increases,90% confidence limits, and 100 simulations. This radius was selected due to the varying length of roads in this analysis, in correspondence with frequent radii used in road ecology literature, and to be representative of a scale in which mitigation measures may be implemented in the Mikumi National Park. This statistic was used as a preliminary analysis to assess the comparative feasibility between road sections as well as to determine the appropriate size of the spatial unit for assessing variation in WVC collision density.

1. A spatially referenced point feature class was created using the exported WVC data within ArcGIS 10.8.
2. The Ripley's K tool was selected from the available spatial statistics options in ArcGIS 10.8's Spatial Statistics Tools.
3. The WVC point feature class was inputted into the Ripley's K tool, and the desired spatial scale parameters were specified.
4. The analysis was executed, generating the Ripley's K results, which provided insights into the spatial clustering or dispersion patterns of wildlife-vehicle collisions within the study area.

3.6 Hotspot Identification in ArcGIS 10.8

The WVC point data and road network data were imported into ArcGIS 10.8 using the Add Data tool. The study area was defined to encompass the extent of interest for the hotspot analysis. WVC points falling within the study area were selected using the Select by Location tool.

To create a density surface representing the intensity of WVCs, the Kernel Density tool in the Spatial Analyst Tools was employed. The WVC point dataset was specified as the Input Point

Features, and parameters such as Cell Size, Search Radius, and Output Cell Values were adjusted accordingly. The Kernel Density tool was executed, generating a density surface capturing the distribution of WVCs across the study area.

Hotspot analysis was performed using the Hot Spot Analysis (Getis-Ord Gi*) tool in the Mapping Clusters section of the Spatial Statistics Tools. The density surface created in the previous step was designated as the Input Feature Class. Additional parameters, including Distance Method and Conceptualization of Spatial Relationships, were adjusted based on the nature of the data and research objectives. The Hot Spot Analysis tool was executed to identify statistically significant hotspots and coldspots of WVCs.

The results of the analysis included a hotspot/coldspot map displaying areas of high and low WVC density, as well as statistical measures such as Z-scores and p-values. These results provided valuable insights into the spatial patterns and clustering of WVC incidents in the study area.

Thematic maps were created to visualize the WVC hotspots and coldspots, employing appropriate symbology and labeling techniques. These visualizations effectively communicated the findings of the analysis.

3.7 Data Preparation for Statistical Modeling

To incorporate multiple factors as explanatory variables for collision risk, a comprehensive dataset was prepared. The dataset included collision intensity, land classification, topographic visibility, and traffic volume. In order to facilitate predictive model calculations, a feature layer with an attribute table was created to consolidate these factors.

To segment the roads into manageable units, the road was divided into 5km sections. For each road segment, a Multiple Ring Buffer feature layer was generated, containing buffer zones at three different distances. The attribute table of the feature layer also included information about the number and frequency of wildlife-vehicle collisions (WVCs) within each buffer zone. Additionally, a raster layer representing topographic visibility was created, specifically clipped to a 200m buffer surrounding each road segment.

The land classification information was available in the form of polygon layers, namely LandClass. These layers were not yet joined with any other feature layers at this stage of the analysis, but would be utilized in subsequent steps.

By combining these various data components into a single feature layer and attribute table, a comprehensive dataset was established, providing a foundation for further analysis and modeling to assess collision risk.

3.7.1 Land Classification

The land classification raster layers obtained in the previous steps were converted into polygon features using the Raster to Polygon tool in ArcGIS. To determine the intersection between the three buffer areas (0.5km, 1.0km, and 5.0km) and the land classification polygons, the Identity tool was applied. This allowed for the inclusion of land class attributes in the attribute table of the buffered road layers.

The attributes within the table were merged based on the shared road segment ID, buffer distance, and land classification category. This ensured that each road segment had only one polygon corresponding to each land classification. To quantify the relative contribution of each land classification within each buffer size and road segment, the total area of the land classification polygons was divided by the total area of the buffer region it intersected with. This calculation resulted in a proportion for each land classification type within each buffer size and segment.

The resulting layer, which now contained collision risk and land classification information for each buffer distance, was renamed as (Segment_Buffer_LC). It was then prepared for the incorporation of the next variable in the analysis.

3.7.2 Viewshed Analysis

The topographic visibility analysis was performed using the Geodesic Viewshed tool, as described earlier. However, the count data contained within each raster cell is not easily accessible or suitable for further analysis in a predictive model. The output of the tool is a raster layer that indicates the number of times each region can be seen from the input points, which are the 1-km spaced vertices created using the Densify tool.

For each road segment analyzed, there are different "observer regions" corresponding to each 1-km segment. However, overlapping regions are counted as separate "Value" or "Observer ID" in the attribute table of the raster. These raster layers contain count data for each cell, which is inversely proportional to the "Observer ID".

To make the count data more manageable and applicable for analysis, the raster layers were polygonised to create the (Segment_Poly_Vis) layer. This polygonization process was performed on 5-km segments, and the integer value representing the average Observer number for each section was assigned to the polygons using the mean value from Zonal Statistics.

3.7.3 Traffic Volume

Finally, a field was added to the layers from above (Segment_Buffer_LC_Vis) called Traffic Vol. Traffic volumes for the study road was measured in Annual Average Daily Traffic (AADT), According to the study conducted by the Park, the average speed of vehicles passing through the park was 77 km/h for buses 50 km/h for lorries, 67 km/h for min buses, 74 km/h for medium-sized, and 75km/h for Saloon categories of vehicles According to the data available, the Traffic density is 2208 vehicles per day an average of 92 vehicles per hours. These were entered manually after cross-referencing the road and segment location, with the resulting layer being renamed accordingly and exported to excel.

3.8 Generalized Linear Mixed Model

The objective of this investigation was to identify the key variables influencing the spatial distribution of wildlife vehicle collisions. The analysis considered seven functional groups, each with three buffer areas, six land classifications, topographic visibility, and traffic volume. Initial analyses revealed that the data for the Amphibians functional group was insufficient for further modeling.

The proportions of land classifications within 0.5 km, 1.0 km, and 5.0 km buffer regions were analyzed separately. When averaging segment proportions, the Built-up, Grassland, and Water bodies classifications showed negligible representation, accounting for less than 10% each of the total buffer area across all buffer zones.

To accommodate the scale of the model, the Traffic Volume variable was rescaled to range from 0 to 1. The Visibility Risk was expressed as a proportion of visibility, as described in (Section 3.6.2). Additionally, the log of Kilometer_Length was used as an offset in the model to standardize the number of collisions based on road segment length.

A generalized linear mixed model, specifically a Poisson model, was employed to analyze the effects of both random and fixed factors on the response variable, which in this case was the count of wildlife vehicle collisions.

The Poisson equation for the generalized linear mixed model (GLMM) used in R can be represented as follows:

$\log(\lambda) = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + ... + \beta_p * X_p + u + \varepsilon$

In this equation:

- $\log(\lambda)$ represents the logarithm of the expected count (λ) of the response variable, which follows a Poisson distribution.

- $\beta_0, \beta_1, \beta_2, ..., \beta_p$ are the fixed effects coefficients associated with the predictor variables $X_1, X_2, ..., X_p$.
- u represents the random effects, which account for the correlation within clustered or nested data.
- ε represents the error term, assumed to follow a normal distribution.

Model: glmerMod

Family: poisson (log)

Formula: NumFnctual ~ 1 + (1 | RoadID) + Vis_Risk + Trafficvol + Dense vegetation + Water bodies + Builtup + Bushland + Grassland + Bareland

Data: df1

Offset: log.km

The Poisson GLMM is used when the response variable is a count variable and the assumption of independence and equal variance in the data is violated. The model estimates the fixed effects coefficients (β's) and accounts for the random effects (u) to better capture the variability in the response variable and address potential correlations within the data. The logarithmic link function is employed to ensure that the predicted values remain positive, as counts cannot be negative.

In R, the GLMMs with a Poisson distribution can be fitted using packages such as 'lme4', 'glmmTMB', or 'MCMCglmm'.

CHAPTER FOUR

RESULTS AND ANALYSIS

4.1 Overview

This chapter presents outputs that have been obtained as a result of implementing the methodological work flow from Figure 3.1. It shows land cover map obtained from classifying a combination of Sentinel bands with some spectral indices, Topographic visibility map obtained from DEM, Hotspot analysis map obtained from integrating collision data and Statistical analysis as well as graphs showing the directional change of clustering and land covers. Finally, it determines which variables have had greatest impact on the spatial distribution of wildlife vehicle carnage.

4.2 Results

4.3 Land Classification (LC)

The highway was buffered at distances of 0.5 km, 1.0 km, and 5.0 km. The land within these buffers was then classified into six categories, and the proportional areas of each category were calculated. Among the three buffer sizes, the classification with the highest proportion of land coverage was Bushland, while Built-Ups accounted for less than 10% of the total areas on average. The classifications of Bareland and Grassland exhibited similar mean values and variances across the whole area.

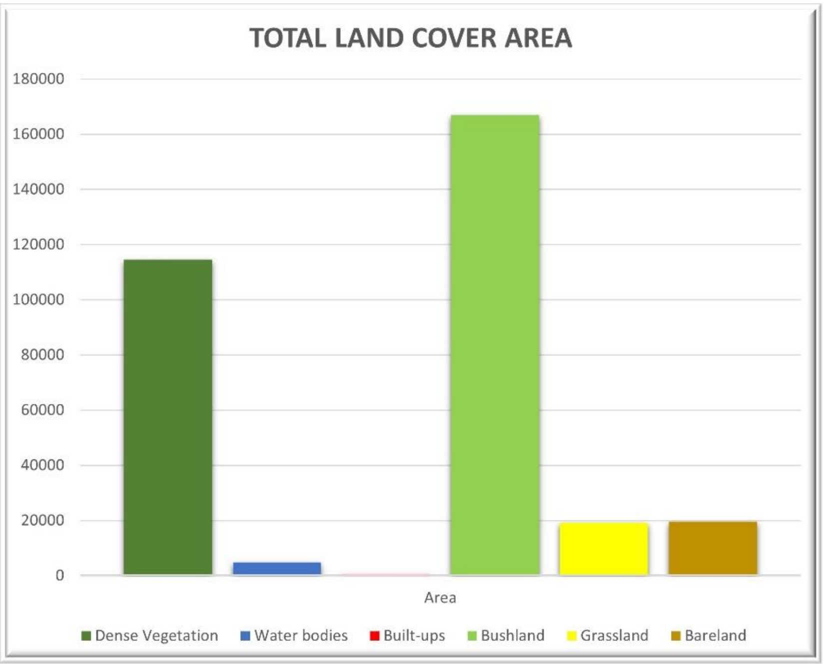

Figure 4-1: Graph of total land cover for the whole study area

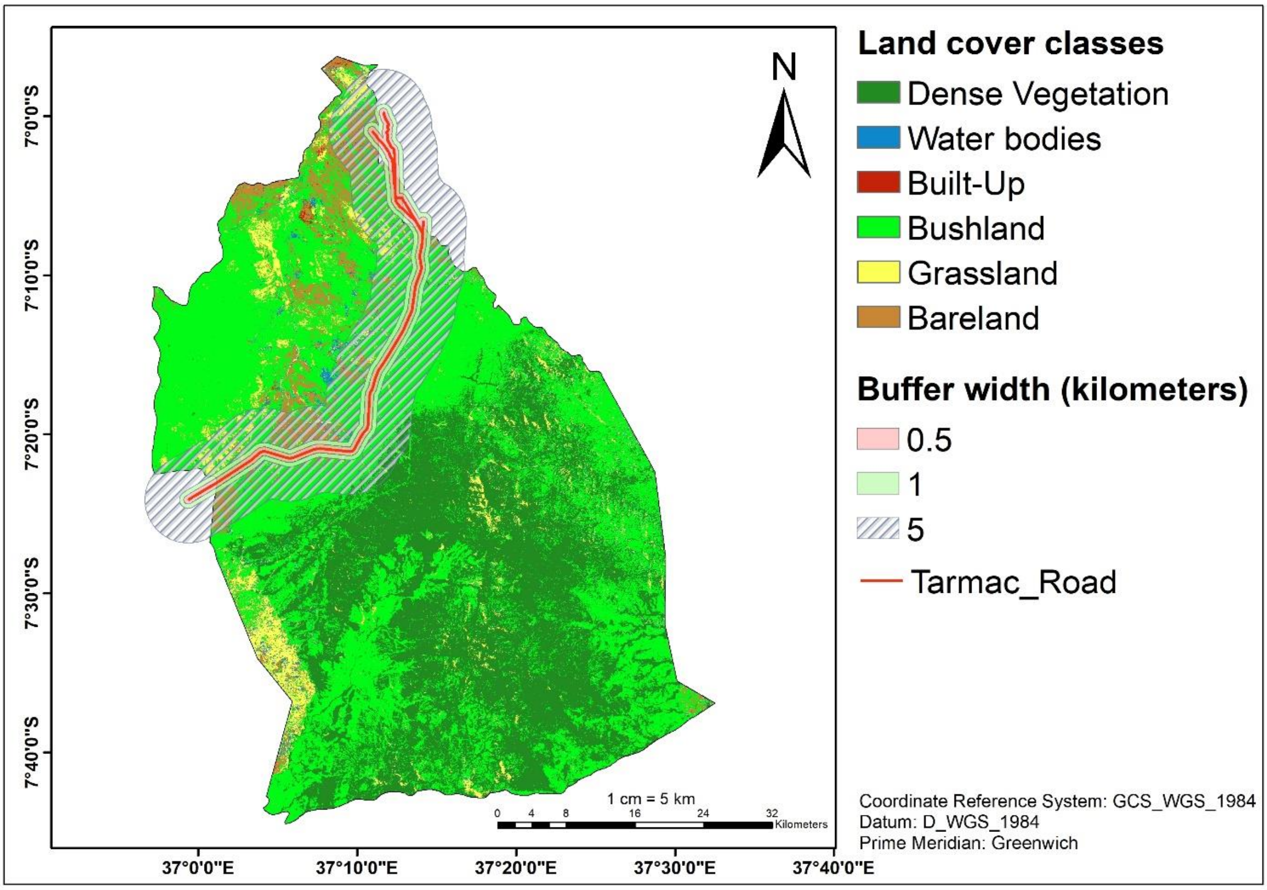

Figure 4-2: Classified land cover map of Mikumi national park with different buffer sizes along the highway

4.3.1 LC Proportions at the 0.5km Buffer Size

In the 0.5 km buffer size, the Bushland land classification had the highest proportion with a mean value of 0.35. The maximum proportion observed for Bushland coverage was 0.7. On the other hand, the Built-up, Grassland, and Water bodies categories had lower means of 0.01, 0.03, and 0.01, respectively. These means fell below the minimum threshold of 0.10 required for inclusion in modeling. The Dense vegetation and Bareland classifications had higher means of 0.19 and 0.12, respectively. These categories also showed comparable medians and distributions.

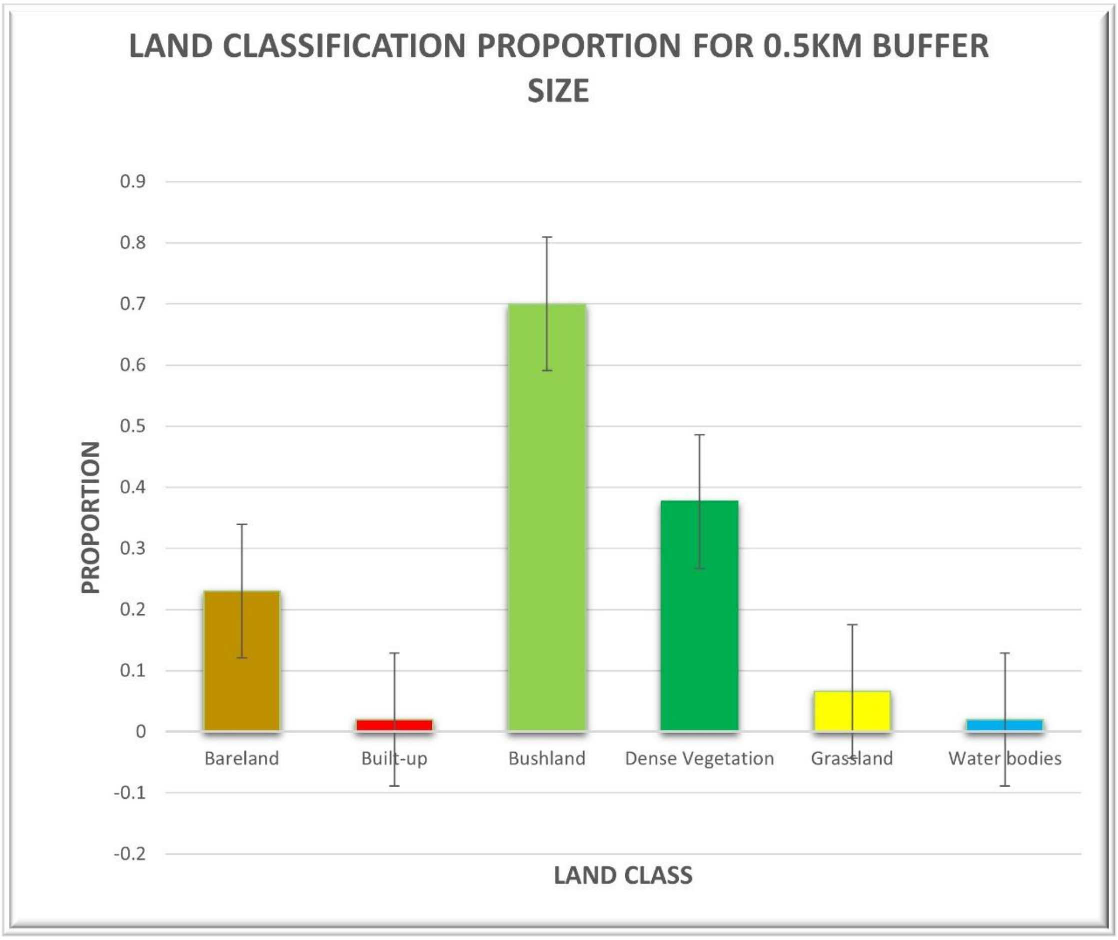

Figure 4-3: Land classification total proportion graph for the 0.5km buffer size

4.3.2 LC Proportions at the 1.0km Buffer Size

In the 1.0 km buffer size, similar patterns were observed as in the 0.5 km buffer size. The Built-up, Grassland, and Water bodies categories had proportions below the 0.05 threshold and were not included in the models. The Bareland and Dense Vegetation categories had mean proportions of 0.15 and 0.2, respectively, which justified their inclusion in the modeling process. In this buffer size, there was a noticeable increase in the difference between the Bareland and Dense Vegetation classifications compared to the 0.5 km buffer. The Bushland classification showed the greatest variation in proportional areas, with a mean of 0.36 and a maximum proportion of 0.72.

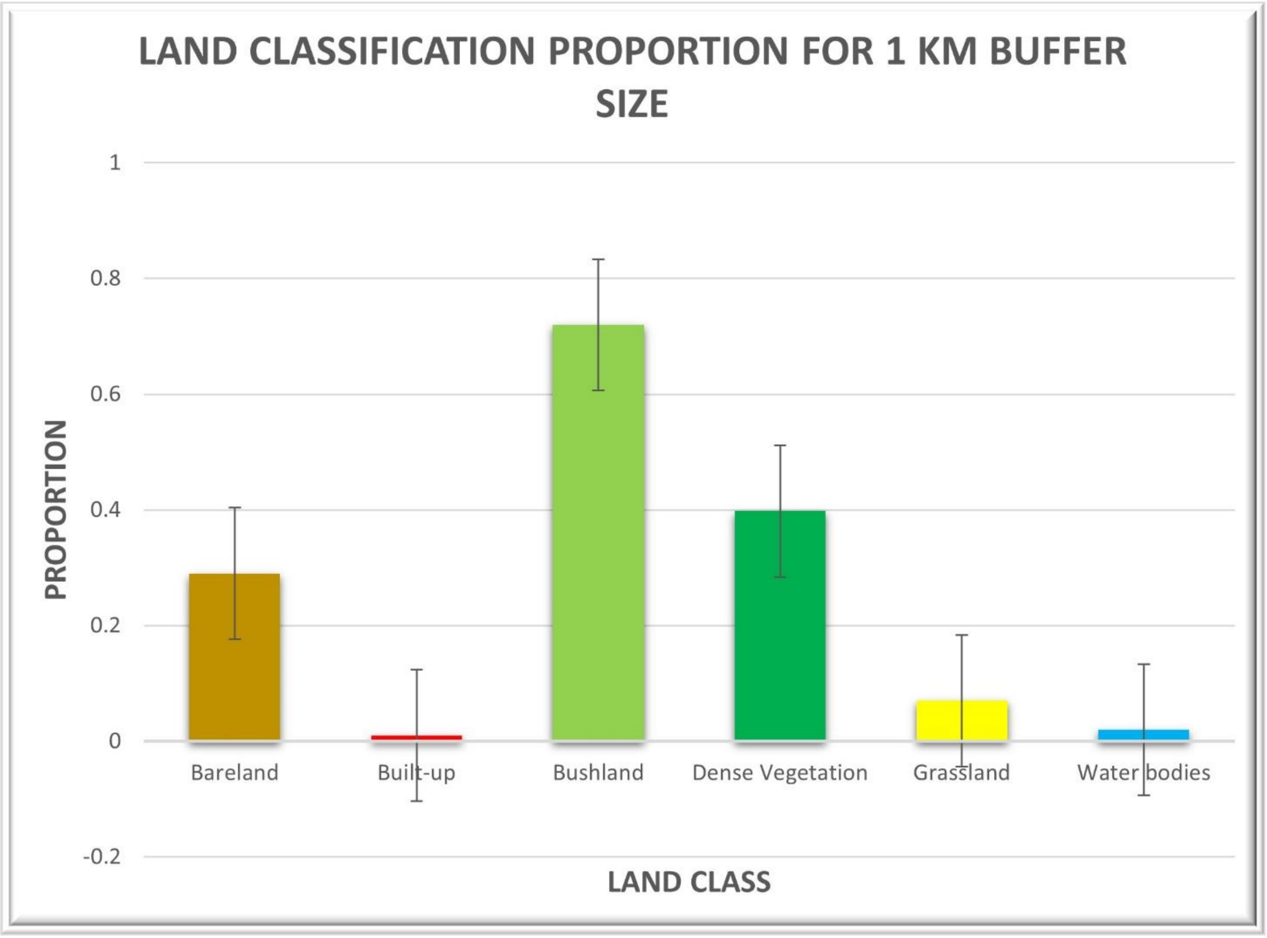

Figure 4-4: Land classification total proportion graph for the 1.0km buffer size

4.3.3 LC Proportions at the 5.0km Buffer Size

In the 5.0km buffer range there is a deviation from the trends of the 0.5 and 1.0 km buffer sizes. Bushland emerges as the dominant category and has no proportions below 0.35 in all buffer sizes, encompassing a substantial proportion of 0.8, indicating its extensive coverage within the study area. Additionally, Dense Vegetation exhibits a notable proportion of 0.33, while Bareland demonstrates a significant extent with a proportion of 0.2. These land classes are expected to exert substantial influence on the response variable in the GLMM framework. Conversely, Grassland, characterized by a proportion of 0.1, is likely to contribute to the variation in the response, albeit to a lesser degree. Built-up areas exhibit a minimal proportion of 0.0046, suggesting limited urban coverage, whereas Water bodies account for a proportion of 0.02, indicating a minor impact on the response variable.

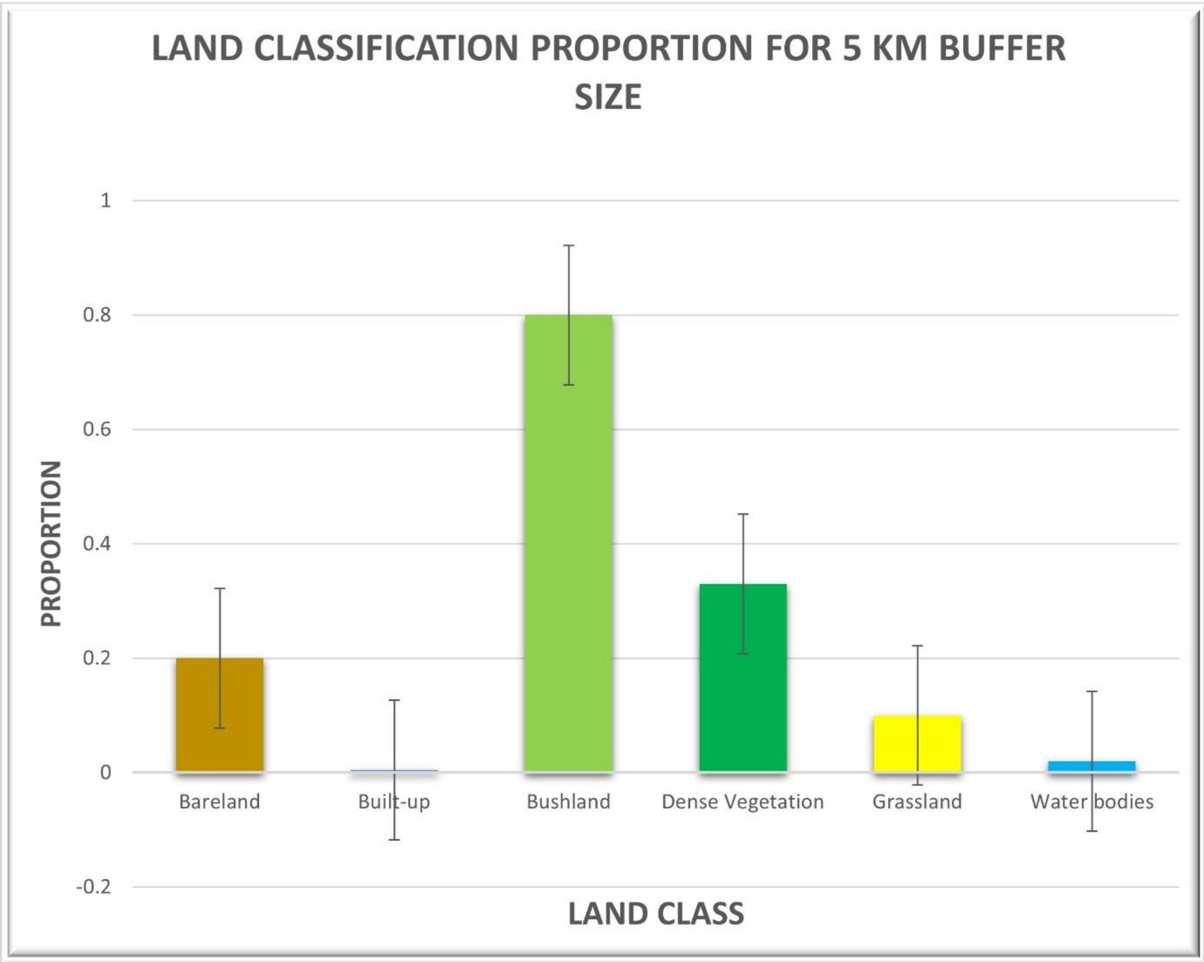

Figure 4-5: Land classification total proportion graph for the 5.0km buffer size

4.4 Topographic Visibility

Topographic driver visibility exhibited considerable variability along the roads within the study area. The visibility analysis measured the frequency of occurrence and visibility of values within the given dataset. The results revealed that certain areas exhibited visibility below 9, indicating limited visibility, while other areas demonstrated visibility exceeding 32, suggesting more visibility in these areas.

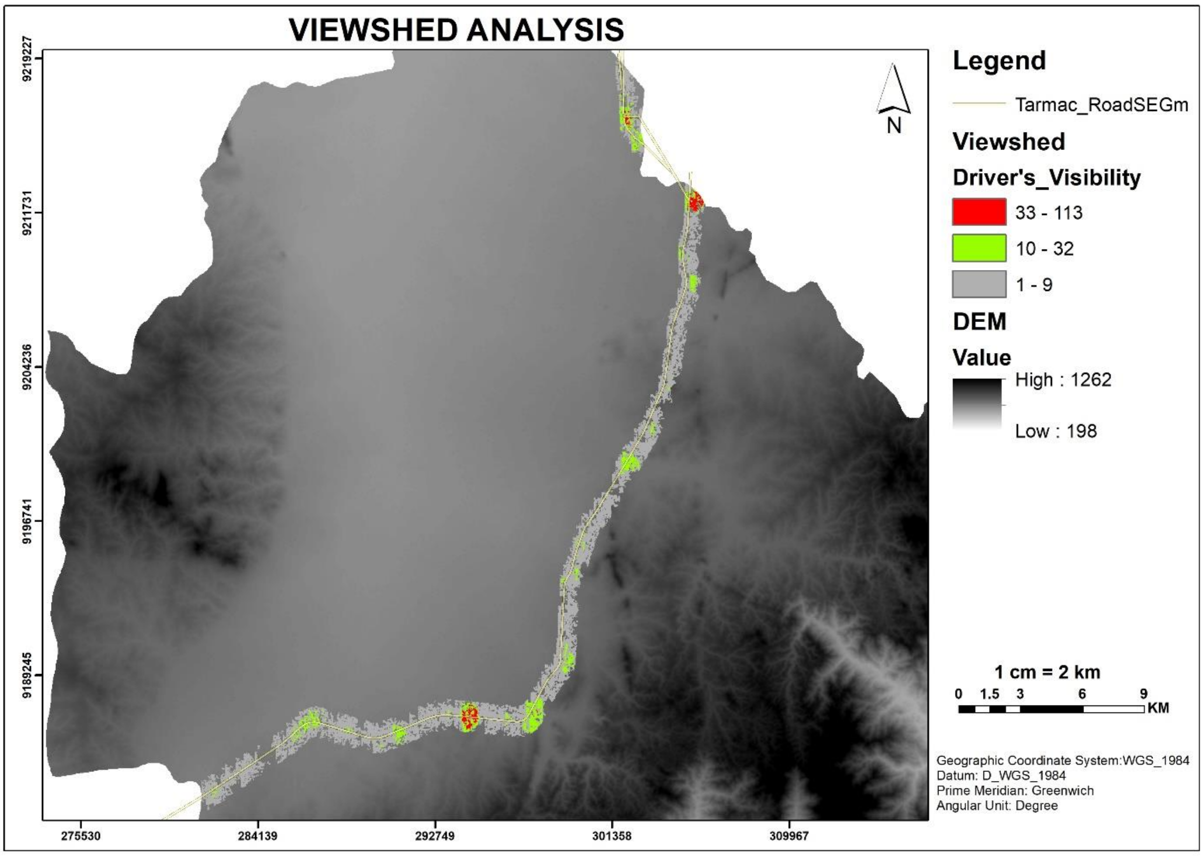

Figure 4-6: Topographic driver's visibility analysis

Higher counts are representing a greater risk and thus less driver visibility. Segment_ID 2 having a notably higher visibility risk, with ~0.99 of the road obscured from driver sight while segment_ID 4 has the lowest risks in terms of visibility in respect to the number of counts.

Additionally, From the study conducted by the park, the data indicates that the traffic density within the park amounted to 2208 vehicles per day, equivalent to an average of 92 vehicles per hour. These insights shed light on the diverse speeds exhibited by different vehicle types, along with the overall traffic volume present in the park. From this graph below, there is no apparent trend between road speed and driver visibility.

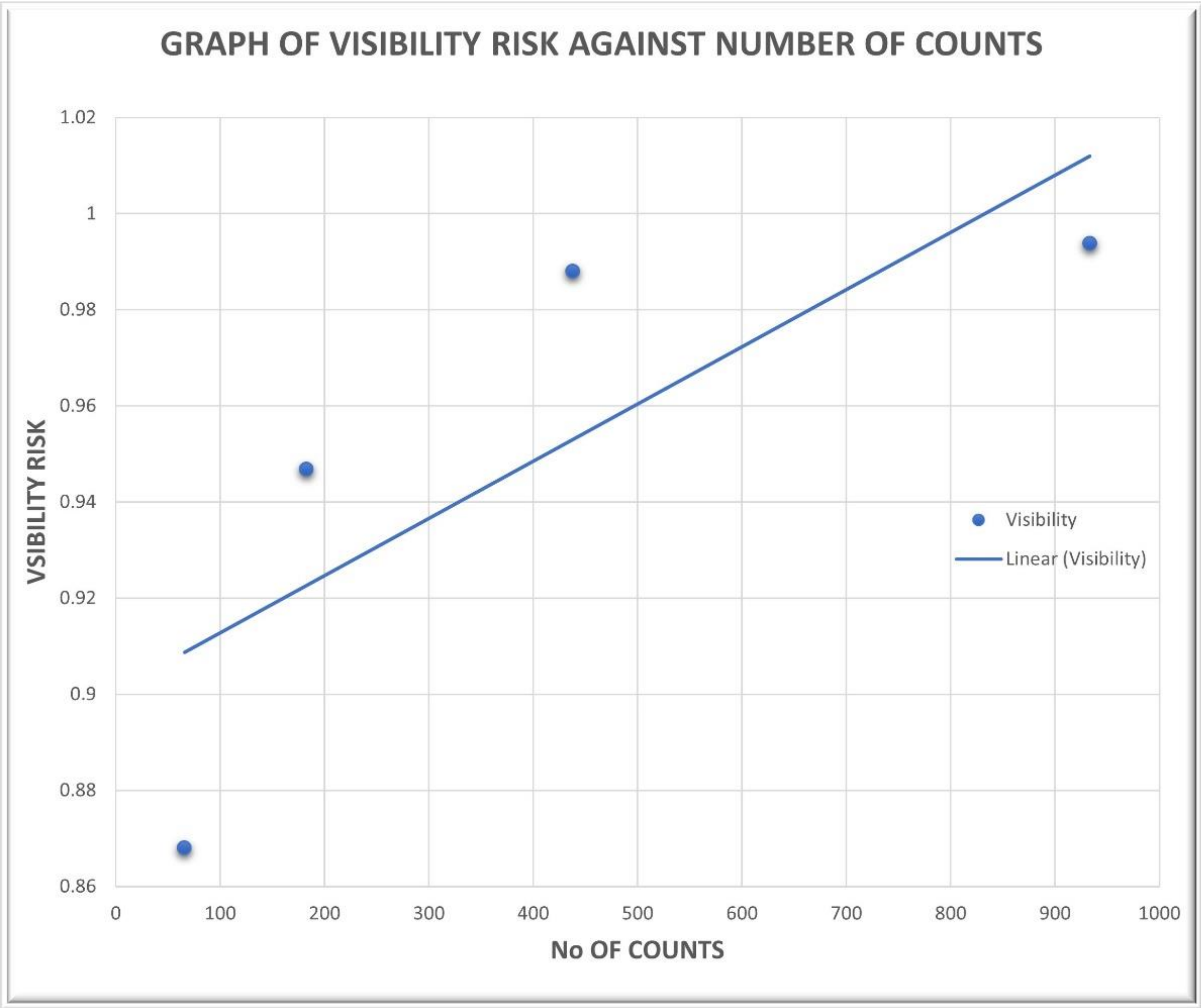

Figure 4-7: Visibility risk graph

4.5 Ripley's K Analysis

Ripley's K-Analysis was conducted for the road and functional group combinations. The results of the analysis are summarized in the graph below, which shows the scale at which clustering occurs for each distance. It is evident that the clustering patterns vary significantly between different road segments and functional groups. This suggests that the spatial dispersion of events, such as wildlife-vehicle collisions, is influenced by factors specific to the road and functional group. The results highlight the importance of considering these factors when analyzing clustering patterns and understanding the underlying dynamics of wildlife-vehicle collisions.

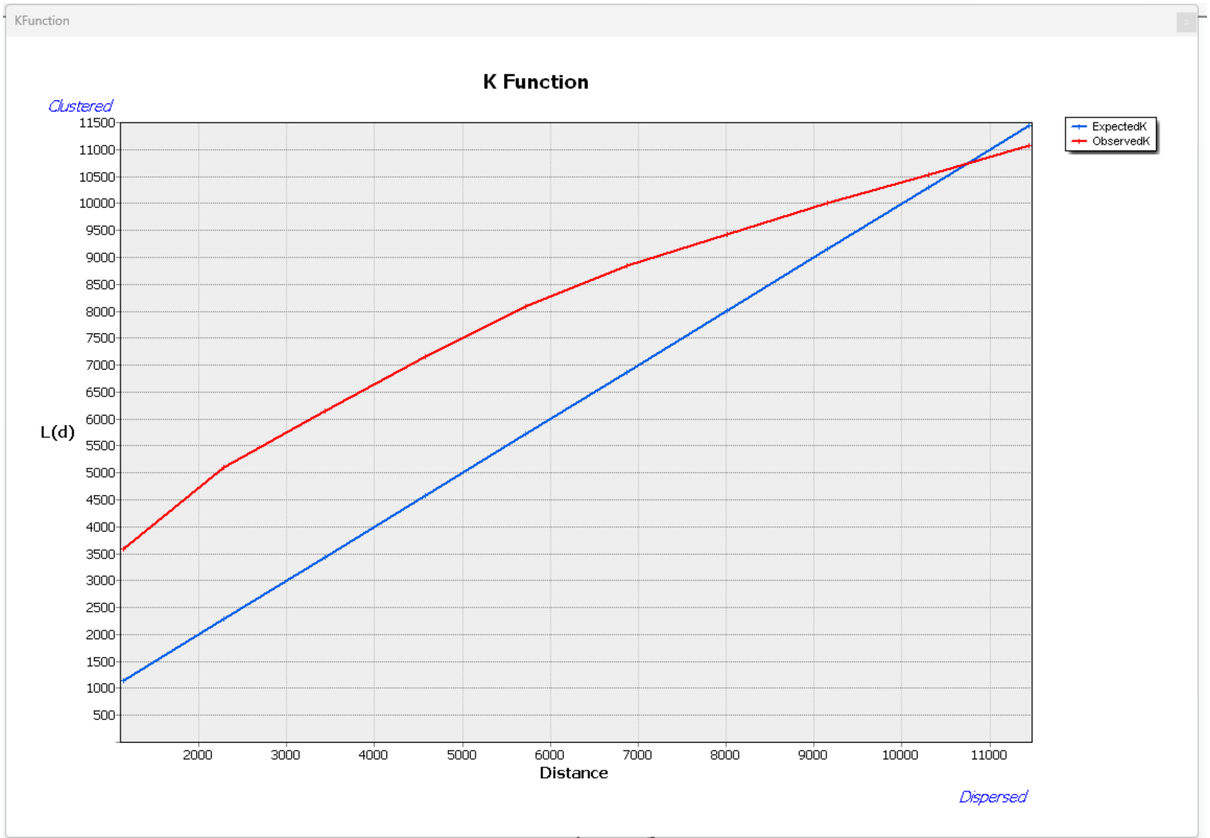

Figure 4-8: The result of Ripley K analysis

4.6 Spatial Autocorrelation

The spatial autocorrelation (Moran's I) analysis was conducted to examine the relationship between the locations of wildlife vehicle collisions and roads. The calculated z-score of 6.04891044713 indicated a significantly high value, suggesting a strong spatial correlation between these two variables. Thus, implying that wildlife vehicle collisions tend to cluster in close proximity to the road rather than being randomly distributed across the study area. Furthermore, the z-score's associated p-value, which is less than 1%, indicated that the observed clustering pattern is highly unlikely to occur due to random chance alone.

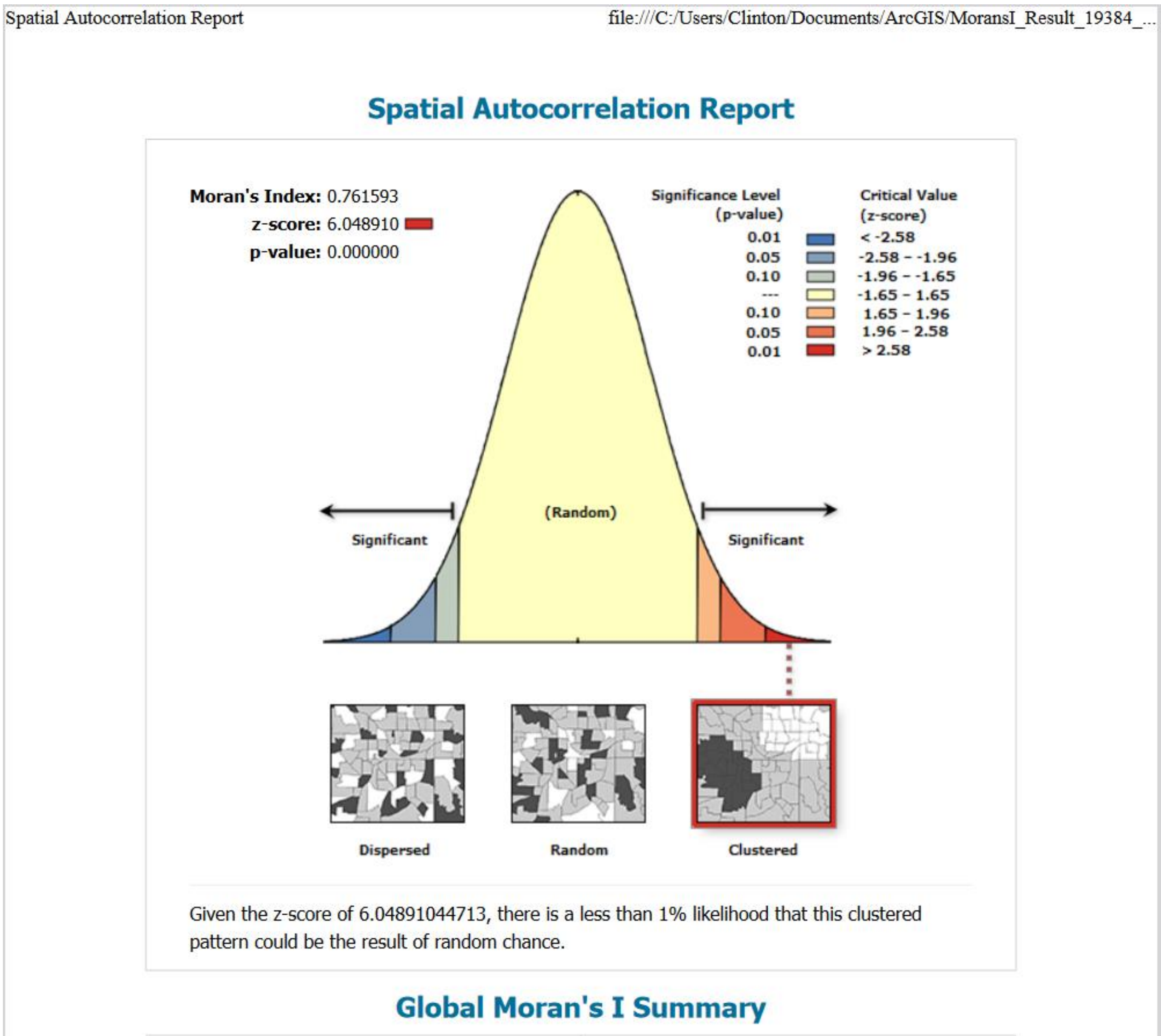

Figure 4-9: The Moran's I report

4.7 Hotspot Mapping

The hotspot analysis conducted on wildlife vehicle collisions (WVCs) revealed valuable insights into the spatial distribution of these incidents. A total of 1334 WVCs were analyzed, with the highest count observed in the Small Medium Vertebrates (SMV) functional group, accounting for 495 collisions or 37% of the total. Conversely, the Amphibians functional group had the lowest recorded collisions, with only 8 incidents, representing a mere 0.5% of the total collisions.

The hotspot analysis generated a visual representation of the collision frequency, with red spots indicating areas characterized by a high frequency of WVCs. These red spots signify regions where wildlife vehicle collisions are more concentrated, suggesting a higher risk and the need for targeted interventions to improve road safety. On the other hand, blue spots on the map denote areas with a lower occurrence of collision outcomes.

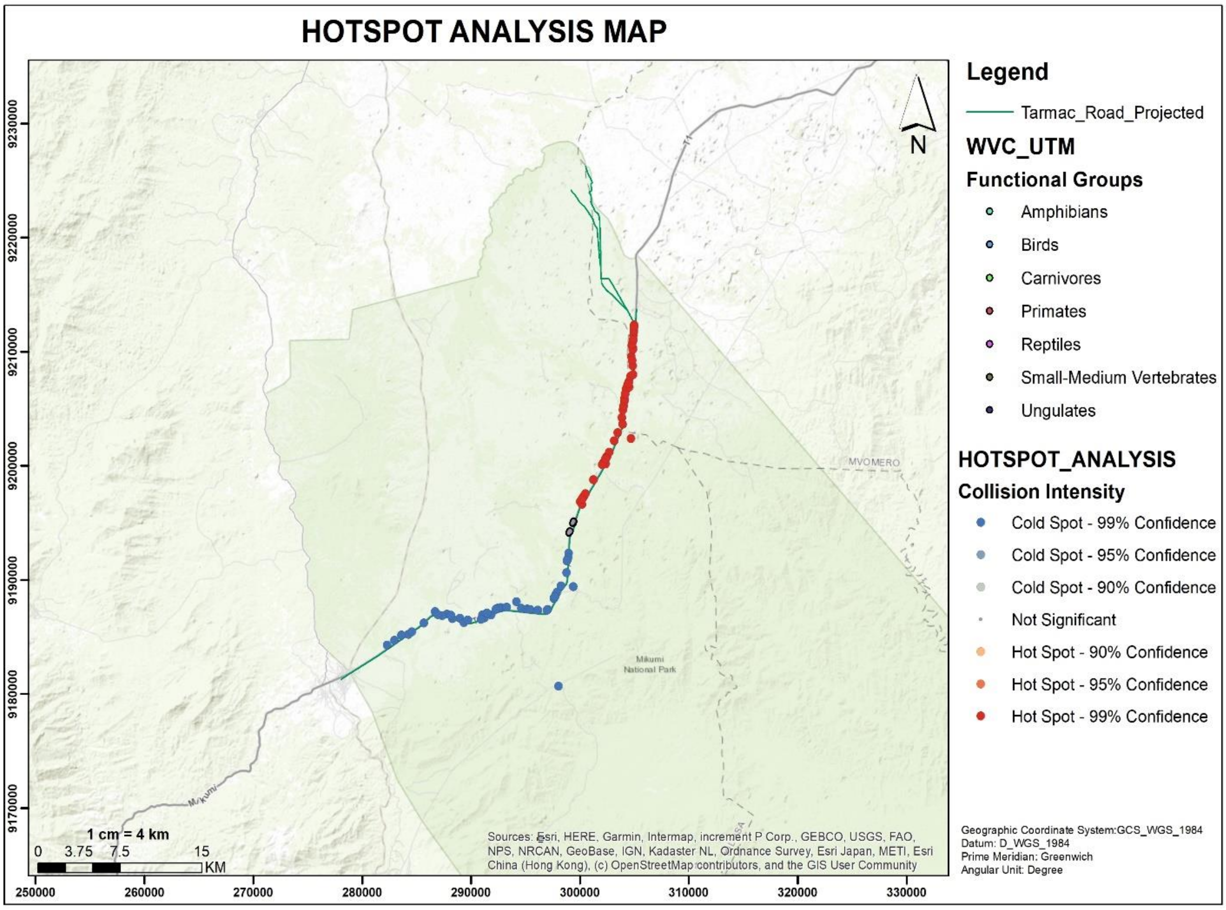

Figure 4-10: Wildlife vehicle collision hotspot mapping

4.8 Inverse Distance Weight

The resulting IDW analysis revealed areas of varying collision intensities across the study area. Regions with a higher density of WVCs were assigned higher values, indicating a greater concentration of collisions. Conversely, areas with a lower occurrence of collisions were assigned lower values, representing a lower collision density.

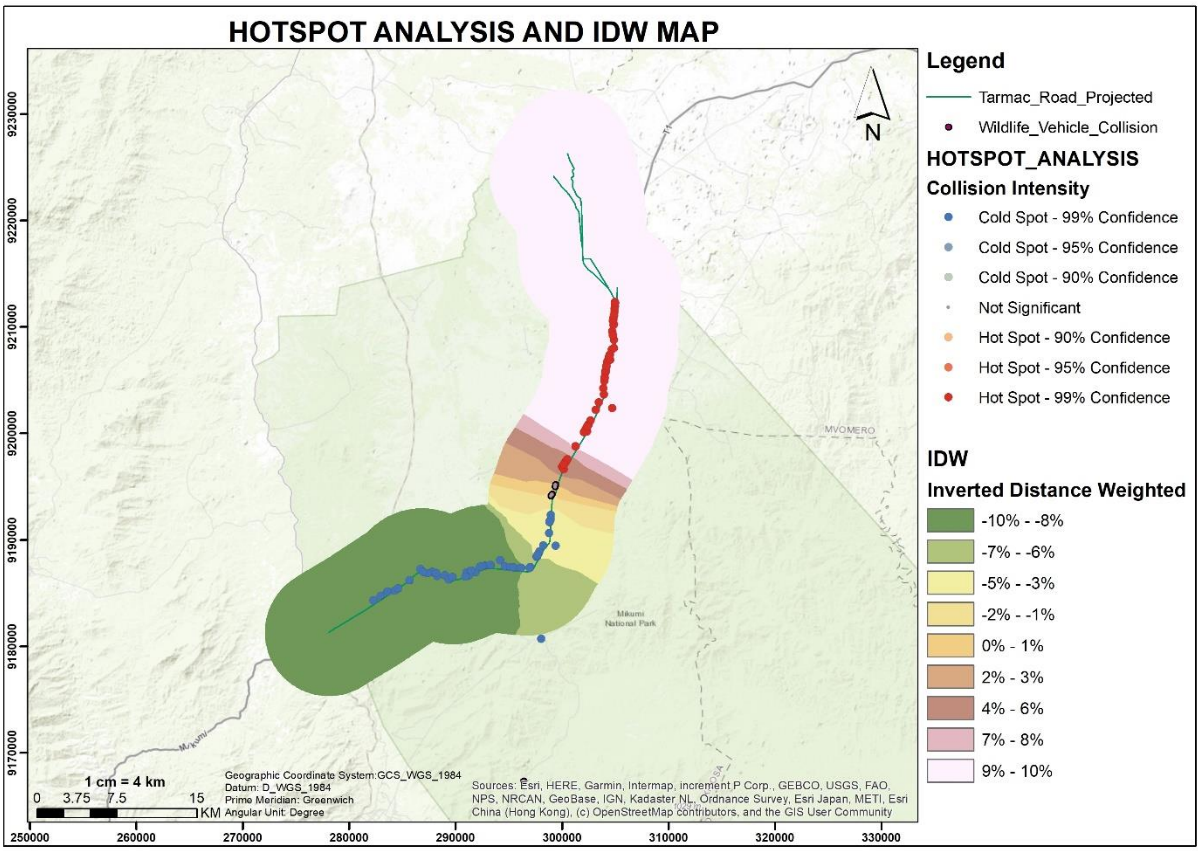

Figure 4-11: Inverse distance weight mapping

4.9 Generalized Linear Mixed-Effects Model for Amphibians and Reptiles in a 0.5 km Buffer size

A Generalized Linear Mixed-Effects Model (GLMM) was developed to analyze collisions involving Amphibians and Reptiles. The model incorporated variables such as topographic visibility risk, traffic volume, and six land classification categories. However, it was found that the classifications of Visibility Risk, Dense Vegetation, Grassland, and Bareland at the 0.5 km scale did not have a significant impact on predicting the collisions of Amphibians and Reptiles.

On the other hand, the land classifications of Built-up and Water bodies, along with the Traffic volume, showed high significance and a positive correlation with the occurrence of collisions. This suggests that areas with more built-up structures, water bodies, and higher traffic volumes are associated with an increased likelihood of collisions involving Amphibians and Reptiles.

When the developed model was used to predict collisions, the Root Mean Squared Error (RMSE) value was determined to be 0.5882. The RMSE provides an indication of the average difference between the predicted and observed collision counts, with a lower value indicating a better fit of the model to the data.

Table 4-1: Generalized Linear Mixed-Effects Model for Amphibians and Reptiles Collisions. Land Classification predictors from proportion of land cover at the 0.5 km buffer size. Significance codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 =.

Fixed effects					
Terms	Estimate	Std. Error	z value	P value	Significance
(Intercept)	0.327	0.1238	2.641	0.00823	**
Vis_Risk	0.0694	0.0403	1.724	0.08486	
Trafficvol	0.066	0.0257	2.583	0.01003	*
Dense vegetation	0.0473	0.042	1.122	0.2621	
Water bodies	0.2353	0.0963	2.447	0.01439	*
Builtup	0.183	0.0931	1.972	0.05014	
Bushland	0.1329	0.0814	1.636	0.1022	
Grassland	0.0735	0.0612	1.192	0.2329	
Bareland	-0.1394	0.0889	-1.573	0.1169	
Random effects					
Groups	Name	Variance	Std.Dev.		
RoadID	(Intercept)	0.2192	0.4705		
RMSE	0.5882				

4.10 Generalized Linear Mixed-Effects Model for Small-Medium Vertebrates and Birds in a 1.0 km Buffer size

A Generalized Linear Mixed-Effects Model (GLMM) was constructed to analyze collisions involving small-medium vertebrates (SMV) and birds. The model incorporated variables such as topographic visibility risk, traffic volume, and six land classification categories, following Equation 2 in Section 3.8 of the study.

The analysis revealed that the variables of Visibility Risk and Traffic Volume, measured at a 1.0 km buffer size, showed varying levels of significance in predicting SMV and Bird collisions. These variables had a significant impact on collision occurrence, although the strength of their influence differed.

On the other hand, the land classification categories included in the model were found to be insignificant predictors of SMV and Bird collisions. This suggests that the specific types of land cover examined in the study did not have a substantial impact on collision occurrences for the analyzed species.

When the developed model was employed to predict collisions, the Root Mean Squared Error (RMSE) value was calculated as 0.753. The RMSE provides a measure of the average difference between the predicted and observed collision counts, with a lower value indicating a better fit of the model to the data. In this case, the obtained RMSE value suggests a reasonable level of accuracy in estimating collision counts based on the model's predictions.

Table 4-2: Generalized Linear Mixed-Effects Model for Small-Medium Vertebrates and Birds Collisions. Land Classification predictors from proportion of land cover at the 1.0 km buffer size. Significance codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 =.

Fixed Effect					
Term	Estimate	Std. Error	z value	P value	Significance
(Intercept)	2.808	0.372	7.553	1.80E-14	***
Vis_Risk	0.563	0.216	2.621	0.00884	**
Traffic_Vol	0.103	0.03	3.436	0.00063	***
Dense Vegetation	0.186	0.127	1.468	0.1424	
Water bodies	-0.138	0.192	-0.723	0.4693	
Builtup	-0.128	0.125	-1.027	0.3058	
Bushland	-0.034	0.112	-0.304	0.7611	
Grassland	-0.048	0.116	-0.413	0.6779	
Bareland	-0.262	0.145	-1.807	0.0711	
Random Effects					
Groups	Name	Variance	Std.Dev.		
RoadID	(intercept)	0.0968	0.311		
RMSE:	0.753				

4.11 Generalized Linear Mixed-Effects Model for Ungulates, Primates, and Carnivores in a 5.0 km Buffer size

A Generalized Linear Mixed-Effects Model (GLMM) was developed to analyze collisions involving Ungulates, Primates, and Carnivores. The model incorporated variables such as topographic visibility risk, traffic volume, and six land classification categories, following Equation in (Section 3.8) of the study.

The analysis revealed that all variables, except for Bushland and Grassland, were highly significant predictors of collisions when measured at a 5.0 km buffer size. The variable with the greatest significance was Dense Vegetation, with a p-value of 1.55E-05 (***), indicating a strong relationship between collision occurrences and the presence of dense vegetation.

When the model was used to predict collisions, the Root Mean Squared Error (RMSE) value was determined to be 0.5156.

Table 4-3: Generalized Linear Mixed-Effects Model for Ungulates, Primates, and Carnivores Collisions. Land Classification predictors from proportion of land cover at the 5.0 km buffer size. Significance codes: 0 = ***, 0.001 = **, 0.01 = *, 0.05 =.

Fixed Effect					
Term	Estimate	Std. Error	z value	P value	Significance
(Intercept)	1.7121	0.2852	6.025	3.60E-09	***
Vis_Risk	0.0954	0.0293	3.254	0.00116	**
Trafficvol	-0.0274	0.0129	-2.12	0.03438	*
Dense vegetation	-0.1555	0.0357	-4.32	1.55E-05	***
Water bodies	0.084	0.0316	2.671	0.00745	**
Builtup	0.1133	0.0319	3.57	0.00036	***
Bushland	0	0.0277	0	1	
Grassland	0.0319	0.0268	1.204	0.22783	
Bareland	-0.0599	0.0275	-2.177	0.03016	*
Random Effect					
Group	Name	Variance	Std. Dev.		
RoadID	(Intercept)	0.2873	0.5409		
RMSE	0.5156				

4.12Analysis of Results

4.13Hotspot Analysis

Based on the results of the Ripley's K analysis as seen in (section 4.5) and the spatial autocorrelation (Moran's I) analysis in (section 4.6), it can be analyzed that hotspot patterns exist in the spatial distribution of wildlife vehicle collisions. The Ripley's K analysis revealed varying clustering patterns at different scales for different road segments and functional groups. This suggests that the occurrence of wildlife vehicle collisions is not randomly distributed but influenced by specific factors related to roads and functional groups. The spatial autocorrelation analysis further confirmed the presence of a strong spatial correlation. This combination of findings indicates the existence of hotspot areas where wildlife vehicle collisions are concentrated.

4.14Statistical Analysis

For collisions involving Amphibians and Reptiles, the GLMM indicated that variables such as topographic visibility risk, traffic volume, and land classifications of Built-up and Water bodies were significant predictors of collision occurrences. These findings suggest that areas with more built-up structures, water bodies, and higher traffic volumes are associated with an increased likelihood of collisions involving Amphibians and Reptiles. However, the land classifications of Visibility Risk, Dense Vegetation, Built-up, Grassland, and Bareland at the 0.5 km scale were not found to have a significant impact on predicting these collisions.

For collisions involving small-medium vertebrates (SMV) and birds, the GLMM revealed that topographic visibility risk and traffic volume at a 1.0 km buffer size were significant predictors of collision occurrences. However, the specific land classification categories examined in the study did not show a substantial impact on collision occurrences for these species.

In the case of collisions involving Ungulates, Primates, and Carnivores, the GLMM indicated that most variables, except for Bushland and Grassland, were highly significant predictors of collision occurrences when measured at a 5.0 km buffer size. The presence of Dense Vegetation was found to have the strongest relationship with collision occurrences among the variables tested.

The Root Mean Squared Error (RMSE) values obtained from the models can be used as an indication of the average difference between the predicted and observed collision counts. Lower RMSE values suggest a better fit of the model to the data and a more accurate estimation of collision counts based on the model's predictions. The obtained RMSE values in the study (0.5882 for Amphibians and Reptiles, 0.753 for SMV and birds, and 0.5156 for Ungulates, Primates, and Carnivores) suggest reasonably accurate predictions in estimating collision counts for each functional group.

The statistical analysis of the developed models provides evidence that specific factors such as topographic visibility risk, traffic volume, and certain land classifications play significant roles in predicting wildlife vehicle collisions for different functional groups. These findings can help inform conservation efforts and the implementation of targeted measures to reduce the risk of collisions and mitigate their impact on wildlife populations.

Table 4-4: Testing Breakdown

Assumptions	Test	Variables	Functional Groups
Wildlife-vehicle Collisions are not random events but rather spatial aggregation	Ripley's K Spatial Autocorrelation	WVC Collision Data	ALL
Collisions will occur more frequently in road areas with greater terrain or Traffic density	Hotspot Mapping GLMM	WVC Collision Data Viewshed Analysis Traffic Volume	ALL

CHAPTER FIVE

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

The comprehensive analysis conducted in this study presents compelling evidence of the existence of hotspot areas and the presence of non-random distribution patterns of wildlife vehicle collisions within the boundaries of Mikumi National Park. These findings have significant implications for understanding the underlying factors that contribute to collision occurrences and provide crucial insights for effective wildlife conservation and road safety management in the area.

The research underscores the importance of specific factors that influence collision events, with traffic volume emerging as a critical determinant. The observed correlation between higher traffic volumes and increased collision likelihood suggests the need for targeted measures to address the potential risks associated with vehicular movement through the park. Additionally, the study emphasizes the role of land classifications, particularly built-up structures, water bodies, and dense vegetation, in shaping collision patterns. These findings highlight the necessity of considering the impact of human settlements, water features, and vegetation density when designing mitigation strategies to minimize wildlife road carnage.

Overall, the research findings presented herein represent a significant advancement in the field of wildlife road carnage analysis. By demonstrating the existence of hotspot areas and non-random distribution patterns, the study provides essential insights into the spatial dynamics of collision events within Mikumi National Park. The identified factors of traffic volume and specific land classifications elucidate the underlying mechanisms driving collision occurrences, thereby enabling targeted interventions to reduce wildlife road carnage and enhance road safety along the highway crossing through the park. The developed GLMM models, with their robust predictive capabilities, serve as valuable tools for policymakers and wildlife conservationists in formulating effective management strategies aimed at preserving biodiversity and ensuring the coexistence of wildlife and road infrastructure in this ecologically significant region.

5.2 Recommendation

1. Implement targeted road infrastructure improvements: Given the significant influence of traffic volume on collision likelihood, it is crucial to assess and implement appropriate road infrastructure improvements along the highway crossing through the park. This may include measures such as the installation of wildlife-friendly fencing, overpasses, and underpasses strategically placed to facilitate safe wildlife movement across the road.
2. Enhance driver awareness and education: Raising awareness among drivers about the potential presence of wildlife and the importance of responsible driving within the park is vital. Implement educational campaigns and signage highlighting the need to adhere to speed limits, remain vigilant, and take necessary precautions to avoid collisions with wildlife.
3. Prioritize wildlife corridors and habitat connectivity: Identify and prioritize the establishment of wildlife corridors and habitat connectivity measures that enable animals to move safely across the park and mitigate the risks associated with road crossings. These corridors should consider the identified hotspot areas and the specific functional groups that are most affected by collisions.
4. Consider land use planning and management: Given the significant influence of land classifications, particularly built-up structures, water bodies, and dense vegetation, on collision occurrences, land use planning and management should be integrated into conservation strategies. Promote sustainable land use practices that minimize habitat fragmentation and enhance the availability of suitable habitats away from high-traffic areas.
5. Continuous monitoring and evaluation: Establish a systematic monitoring and evaluation framework to assess the effectiveness of implemented mitigation measures and identify any emerging trends or new hotspot areas. This will enable adaptive management strategies to be developed and refined over time.
6. Collaboration and stakeholder engagement: Foster collaboration between relevant stakeholders, including wildlife conservation organizations, transportation authorities, local communities, and researchers. Engage in ongoing dialogue and knowledge sharing to develop comprehensive strategies that address the complex challenges of wildlife road carnage in a multidisciplinary manner.

By implementing these recommendations, it is possible to reduce the occurrence of wildlife vehicle collisions within Mikumi National Park and safeguard the well-being of both wildlife populations and road users. It is essential to prioritize the conservation of wildlife and the preservation of their habitats while ensuring safe and sustainable transportation infrastructure within and around the park.

REFERENCES

Abdulhafedh, A. (2019). An Innovative GIS Method for Evaluating the Visibility of the Road Using the ArcMap-Tools. Open Access Library Journal, 6, e5586.

Alexander, S., & Waters, N. (2014). Road Ecology. In M. Garrett (Ed.), Encyclopedia of Transportation (pp. 1162–1165). essay, SAGE Publications

Aoife Moroney. (2018). The use of spatial and temporal analysis in the maintenance of road mortality mitigation measures for wildlife in Ireland**.** KTH ROYAL INSTITUTE OF TECHNOLOGY SCHOOL OF ARCHITECTURE AND THE BUILT ENVIRONMENT.

Baddeley, A. (2010). Spatial analysis of point patterns. John Wiley & Sons.

Barre, J., & Geldewhays, N. (2016). Impact of habitat fragmentation on the diet of Eurasian badgers (Meles meles) in Flanders, Belgium. European Journal of Wildlife Research, 62(3), 627-638.

Bates, D. M., Maechler, M., Bolker, B. M., & Walker, S. (2015). Fitting linear mixed-effects models using R: A tutorial. Journal of Statistical Software, 67(1), 1-48.

Chakravorty, S. (1995). Identifying crime clusters: The spatial principles. Middle States Geographer, 28(1), 53-58.

Cibot, M., Goswami, V. R., Pouzols, F. M., & Blanchet, F. G. (2015). Road kill hotspots in protected areas: a case study in Western Ghats, India. Environmental management, 56(1), 171-182.

Cramer, P., Hobbs, R., & Higgs, E. (2015). Habitats through time: the past two decades of road ecology. Ecology and Society, 20(4), 33.

Demidenko, E. (2013). Mixed models: Theory and applications. CRC Press.

DREWS. e. (1991) Road kills in Mikumi National Park. *Miombo, Newsl. Wildl. Cons. Soc. Tanzania 7,6-7.*

Favilli, F., Dapporto, L., Brilli, M., Fanfani, A., & Marconi, S. (2018). Toad mortality on roads: an experimental investigation of the effects of motor vehicle speed and tire composition. PeerJ, 6, e4392.

Forman, R. T. T., 2000. Estimate of the Area Affected Ecologically by the Road System in the United States. *Conservation Biology, Vol. 14, No. 1 (Feb., 2000), pp. 31-35.*

Geldewhays, N., Barre, J., & Keersmaekers, B. (2018). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. Journal of Applied Ecology, 55(1), 109-117.

Geological Survey (U.S.), 2000. *USGS EARTH EXPLORER*. [online] EarthExplorer. Available at: <https://earthexplorer.usgs.gov/>.

Google Earth Engine for geo-big data applications: A meta-analysis and systematic review https://www.sciencedirect.com/science/article/pii/S0924271620300927.

Gorelick, M. B. (2017). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. MSc thesis, University of York.

Haifa Tamimnia, I. (2020). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. MSc thesis, The Hebrew University of Jerusalem.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning (2nd ed.). Springer.

Ho, T. K. (1995). Random decision forests. In Proceedings of the 3rd international conference on machine learning (pp. 244-251).

Huijser, M. P. (2009). Predicting wildlife collisions using viewshed analysis and generalized linear mixed models. Landscape Ecology, 24(10), 1337-1351.

Jasiewicz, P., & Stepinski, T. F. (2013). Geomorphons: a new approach to terrain analysis using visibility analysis and computer vision. Geomorphology, 194, 15-27.

Jense, S. (2013). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. MSc thesis, University of Copenhagen.

Johansson, M., Hedlund, J., Wallin, H., & Liberg, O. (2019). The impact of forest fragmentation on the diet of Eurasian badgers (Meles meles) in southern Sweden. Landscape Ecology, 34(11), 2359-2372.

Lee, J., Acharya, S., & Lee, J. (2018). Land cover classification using decision tree and spectral indices derived from Sentinel-2 imagery. ISPRS Journal of Photogrammetry and Remote Sensing, 139, 17-27.

Leo Breiman. (1995). Random Decision Forests. In Proceedings of the 3rd International Conference on Machine Learning (ICML), pp. 244-251.

Lin, X., Zhang, H., & Zhang, J. (2017). Generalized linear mixed models: A practical guide. CRC Press.

Lu, Y. (2018). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. MSc thesis, Beijing Forestry University.

Mather, M. C., & Koch, P. L. (2011). Using viewshed analysis and generalized linear mixed models to identify areas of high wildlife collision risk. Wildlife Biology, 17(4), 309-318.

Mc Garigal, K. J., & Noon, B. R. (2016). Spatial patterns in wildlife populations: analysis and interpretation. Springer.

McLennan, M., & Asiimwe, P. (2016). Road-kills in Uganda's national parks: the impact of protected areas on large mammal populations. Oryx, 50(1), 141-148.

Mete, O., & K, S. (2013). Land cover mapping using decision trees and vegetation indices in a Mediterranean environment. International Journal of Applied Earth Observation and Geoinformation, 22, 116-124.

Needham, J. G., Faust, L. J., Zollner, P. A., & Bissonette, J. A. (2020). Spatial patterns of animal-vehicle collisions and associated risk factors in Indiana, USA. Journal of Environmental Management, 262, 110323.

Rehna, M., & Natya, N. (2016). Identifying wildlife collision hotspots using viewshed analysis and generalized linear mixed models. Journal of Threatened Taxa, 8(3), 7392-7398.

Valerio, A., Ciolli, M., Zaccarelli, N., & Ricotta, C. (2021). Carcass removal patterns in road-kill surveys: A case study with slow-moving turtles. Ecological Indicators, 122, 107303.

Wheatley, H. J. (1995). The use of viewshed analysis to identify areas of high wildlife collision risk. MSc thesis, University of Birmingham.

Wood, J. J. (2018). The use of viewshed analysis and generalized linear mixed models to identify and assess the risk of wildlife collisions at road-rail crossings. MSc thesis, University of Exeter.

Zhang, J., & Ye, J. (2016). Random forest: A powerful tool for classification and regression. In Data mining with big data (pp. 161-187). Springer, Cham.