Article

Design of the data-driven software application for identification, population monitoring, and risk assessment for lions in Serengeti Tanzania

Ambokile Okey[1], **Devotha Nyambo**[1], **Shubi Kaijage**[1], **Emmanuel Masenga**[2], **Matana Levi**[2]
[1]School of Computational and Communication Sciences and Engineering (CoCSE), Nelson Mandela African Institution of Science and Technology (NM-AIST), Tanzania
[2]Serengeti Wildlife Research Centre (SWRC), Tanzania Wildlife Research Insitute (TAWIRI), Tanzania
E-mail: ambokileo@nm-aist.ac.tz

Received 15 September 2024; *Accepted 20 October 2024*; *Published online 29 October 2024*; *Published 1 March 2025*

Abstract
This study presents a design of a Data-Driven software application for identification, population monitoring, and risk assessment for lions in Serengeti Tanzania. Lions' populations have been declining due to poaching, overhunting, and other ecosystem factors resulting in unmet demands for tourism and ecological balance. Data-driven techniques can lower the negative consequences by providing mechanisms for lions' management, risk assessment, and monitoring in selected wildlife reserves. Lion's whisker spots, poaching rates, prey availability, human-conflict incidences, and pride size are key elements for achieving management, identification, monitoring, and risk assessment for lions. The software application design aimed at providing conceptual and logical requirements for the development of the application that will enhance lions' monitoring and management efforts to protect their existence and contribution to the ecosystem. The study was conducted in the Serengeti ecosystem, including ecologists from the Tanzania Wildlife Research Institute Serengeti Wildlife Research Center, and information systems analysts. Through a mixed research methods approach, qualitative methods and incremental prototyping software development life cycle model were used to develop the specific requirements. Unified Modeling Language (UML) was used to model the requirements and led to the realization of design diagrams: application framework, database design, and artificial intelligence model workflows. The application should equip ecologists with tools to add and identify specific lions, monitor sightings, estimate population trends, assess risks for individual lions, and produce reports on monitoring and sightings. This design serves as a foundation for developing the data-driven software application for identification, population monitoring, and risk assessment for lions in Serengeti National Park Tanzania which will enhance monitoring and management activities of lions' population non-invasively.

Keywords database design; functional requirements; non-functional requirements; lion's identification; wildlife monitoring; conceptual design; unified modeling language; artificial intelligence; whisker spots.

Computational Ecology and Software
ISSN 2220-721X
URL: http://www.iaees.org/publications/journals/ces/online-version.asp
RSS: http://www.iaees.org/publications/journals/ces/rss.xml
E-mail: ces@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

In recent days of the 21st century, there have been a lot of innovations and technologies specifically in Information and Communication Technologies (ICTs) that have completely changed how our societies work, exchange knowledge, and consume natural resources (Chowdhary, 2020). The advancement of ICTs has affected many sectors in one way or another; the benefits outweigh their drawbacks (Albahri et al., 2020). The modern ICTs include blockchain, deep learning, machine learning, e-governance, data science, Internet of Things (IoT), quantum computing, robotics, virtual and augmented reality, embedded systems, Nanotechnology, and cloud computing.

Artificial Intelligence (AI) is one of the breakthroughs of ICT, and it has led to an increase in automated tools developed using computer vision techniques, machine learning, deep learning, and data science (Venkatesh, 2022). Adopting AI tools developed without local datasets may face challenges in their performance due to differences in data quality and data requirement which can cause misinformation and issues with responsible AI (Barocas et al., n.d.).

In wildlife management and conservation, AI solutions can greatly enhance management activities and help in decision-making. For example, A computer-assisted system for photographic mark-recapture analysis has been used at Tarangire National Park, in Tanzania for giraffe identification; further, it informed an estimate of the giraffe population (Bolger et al., 2012).

On the other hand, there is a need to develop data-driven software applications for identification, population monitoring, and risk assessment for lions in Serengeti because their current method is manual and tedious. This study was conducted at Serengeti National Park (SNP) where various stakeholders were involved in the co-creation of data requirements, conceptual design, and logical design. The outcome of this study is a conceptual and logical design towards developing a data-driven software application that will enhance lion monitoring and help the Tanzania Wildlife Research Institute (TAWIRI) Serengeti Wildlife Research Centre (SRWC) in conservation efforts.

Recently, the number of lions has been decreasing due to poaching, overhunting, and other ecosystem factors thus, proper identification and monitoring of lions are important to track their population and enhance decision-making (Pereira et al., 2022). Currently, radio collars are used for tracking and monitoring lions in SNP; the technique is invasive, requires replacement after some time, does not provide individual identification of lions as only a female lion in the group wears it, and is not durable because of the nature of lions, and can be easily destroyed. Traditionally, individual identification of lions involves experts who record and interpret the physical features of the lions: ears, nose color, mane size, and whiskers spot pattern. This traditional technique requires human expertise to interpret those physical features which takes time to get the desired results, furthermore, the method is prone to bias in lions' identification.

Radio collars, information retrieval systems, and AI-based systems have been used to identify and monitor lions' populations, proving the readiness to deploy newer technologies in the stated area of interest. Both of the mentioned approaches have different weaknesses: an information retrieval system requires expertise to interpret the physical features of lions (Maralions, 2023), while available AI-based systems do not make effective use of whisker spots pattern which are proven identification features in lionjust like a fingerprint on humans.

Hermon and Sharma (2020) developed a model on Asiatic lions that could use one side of the whisker spot pattern, the model was trained on a few datasets and faced challenges in using the whisker spots pattern on both sides of the Asiatic lions' face. Using both sides of whisker spots is important in identification as parent and offspring lions may have one highly correlated side. Furthermore, reviewed AI-based systems do not provide population estimation and risk assessment to warn wildlife conservation management about the danger to lion species (Hermon and Sharma, 2020; *LINC - Linc Lion*, n.d.; Teliolabs, n.d.).

A summary of the current methods for lions' identification and monitoring in SNP is shown in (Fig. 1). It contains a female lion wearing a radio collar, and one yellow card with the front and the back of it. The front of the card comprises the identification details using the traditional method, while the back contains general images of the individual.

Fig. 1 Current lion identification methods in Serengeti.

Camera traps were used to capture lions, and experts identified them by observing and interpreting the whisker spots pattern as a primary feature and other features such as scars and pelage patterns for cubs (Strampelli et al., 2022). The technique requires expertise and it is time-consuming to identify individual lions. On the other hand, the Mara Predator project uses the same method and stores the features in the database (Maralions, 2023). DNA sampling technique is also employed in identifying individual lions where experts collect lions' scat and analyze them to identify individuals (Soller et al., 2020). This method is expensive, requires expertise in collecting target species scat, and is time-consuming.

Deep Convolution Neural Networks (DCNN) have been employed to identify individual large felids, and individual tigers and leopards were identified using their coat markings (Pucci et al., 2020). This model did not work for lions because they do not possess coat markings. An automated tool for encoding uniqueness within the whisker spots pattern for individual lions was developed, the tool faced challenges in individual lion identification as it was trained on a few datasets, further, the tool used only one side of whisker spots for individual lions (Hermon and Sharma, 2020). The spatial mark-resight model was used to estimate lions' density where investigators were identifying lions using marks such as mane size, nose color, scars, and missing tail tips (Kane et al., 2015). The technique required investigators to compare their results because the methodology was prone to errors.

This study aims to enhance lion identification, population monitoring, and risk assessment by utilizing whisker spots from both sides of a lion's face. A whisker detection model will be developed to identify the left and right whisker spots by drawing bounding boxes to extract that region and give it as input to the identification model for generating unique codes for each side, which will be stored in a database along with other attributes. Population estimates will be made using a clustering technique that groups similar images, with the number of clusters representing the estimated lion population. Additionally, factors like age, pride location, reported injuries, pride leadership, pride size, and gender will be used to predict individual risk.

2 Methodology

This study was conducted in SNP Tanzania because it has the largest lion population in Tanzania, and it's the largest lion ecosystem in the world (Smitz et al., 2018). Four Ecologists at the TAWIRI research Centre in SNP were involved in this study. The study adopted a mixed research methodology for determining application requirements and creating its design (Mack et al., 2005). The incremental Prototyping technique was used for gathering the requirements, this software development life cycle model was employed because the requirements for developing the application were not clear in the beginning (Sommerville, 2011). Four ecologists were interviewed and engaged in focus group discussion, further, researchers observed the current methods used in identification, monitoring, and risk assessment for lions. After analysis, their results expanded the awareness of the problem, lions' ecosystem, and the requirements for developing the application. Unified Modeling. Language (UML) was used for modeling the requirement and MS Excel was used to present them in matrix format. Draw.io software tool for drawing UML diagrams was used to implement the Data-Driven Software Application (DDSA) framework diagram, class diagram for database design, and activity diagrams to present model workflow (Drawio Online, n.d.).

3 Results

This section presents results from the incremental prototyping, interview, focus group discussion, and observation as discussed in the methodology section. From the interview, focus group discussion, and observation; researchers could understand lions' behaviours and their features which served as a foundation for designing the application. After the analysis, researchers understood features to be included in the database design, and the DCNN and Machine Learning (ML) models. The Incremental prototyping technique improved the database and design of the processes for the DDSA.

3.1 Lion's behaviours and features

The study highlights that lions, as large felids, live in social groups called pride, where dominant males play a crucial role in protecting the pride and mating with resident females. Adult males weigh between 140 and 250 kg, while females range from 120 to 182 kg. Female lions remain in their pride for life, but young males are expelled around 2-3 years of age to avoid inbreeding, forming coalitions until they challenge resident males for dominance. Physical features such as whisker spots, ear notches, nose color, and manes in males are key for individual identification and age estimation, with both sides of whisker spots being particularly unique and essential for distinguishing lions, even among closely related individuals. A summary of the lion's features included in DDSA design is shown in (Table 1).

Table 1 Lion's features.

Feature	Description	Data type
Lion’s ID	This is like a name given to a lion, it is given by the abbreviation of Pride name, number for ordering, and abbreviation of age group e.g., CF for cub female. Example BAb01-AF; BA is Barafu pride, 01 is numbering, and AF is Adult Female.	Text
Pride name	This is the name of the social group of lions living together in the same territory	Text
Ear Notches	This is a cut on the lion’s ear	Integer
Age	This is the number of years respective to the year’s born	Integer
Nose color	This represents the color of the nose can be black or pink or pink with black dots	Text
Gender	This represents the sex of an individual	Character
Left Whisker spots	This represents the left whisker spots pattern on lion’s muzzles	Text
RightWhisker spots	This represents the right whisker spots pattern on lion’s muzzles	Text
Scar	This refers to the mark left on a lion’s body after an injury to body tissue	Text
Kinked tail	This refers to the bent in the tail	Text
Pride latitude	This refers to the latitude of the pride	Double
Pride longitude	This refers to the longitude of the pride	Double

3.2 Data-Driven application software functional and non-functional requirements

3.2.1 Functional requirements

Functional requirements refer to what a system must have to meet users’ needs, this includes features and functionalities. Further, they focus on what the system should do rather than how it should do it. Various requirements-gathering techniques include interviews, focus group discussions, Joint Application Development, prototyping, questionnaires, observation, document reviews, etc. This study employed incremental prototyping, observation, interview, and focus group discussion to gather the requirements for DDSA development. Unified modeling language was used to model the requirement, and the results summary is shown in (Table 2).

Table 2 DDSA Functional Requirement.

Requirement	Description	Actor
Application should allow user to be registered	All ecologists who interact with lions must be registered and given a user profile; also, they should be given username and password	Application administrator
Application should allow user to login	All ecologist, managers, and administrator must be given access to the application after successful login with username and password	All users
Application should allow user to control their account details	All users must be given capability to control their data: name, email, password, picture.	All users
Application should allow ecologist to add lion	All ecologists must be given the ability to add lion details: left whiskers image, right whiskers image, pride, name, pride membership status, age, general images, etc.	Ecologist
Application should allow ecologist to identify individual lion	Ecologists shall be allowed to identify lions by uploading images of left and right whiskers and search through the database	Ecologist
Application should provide details on number of sightings per each individual in the pride	Ecologists shall be given details of how many times an individual was identified in a given period of time	Ecologist
Application should provide estimate of lion population based on sightings after a given period of time	Ecologist shall be given estimate of the lion population from the sightings in a given period of time.	Ecologist
Application should predict risk associated with an individual	Ecologists shall be given with risk prediction details associated with identified individual	Ecologist
Application should keep record of lion behavior and healthy condition when sighted	Ecologists shall be given ability to record lion behavior and healthy status when sighted	Ecologist
Application should provide monitoring and sighting reports	Ecologists shall be given a report of monitoring and sightings	Ecologist, Centre Manager

3.2.2 Non-functional requirements

Non-functional requirements in contrast to functional requirements, describe quality in the operation of the system rather than its functions and services that are required. Non-functional requirements define how the system performs a certain action rather than what actions it performs. The techniques used to gather non-functional requirements are similar to those used for functional requirements discussed in section 3.2.1 of the functional requirements. A summary of non-functional requirements is presented in (Table 3).

Table 3 DDSA non-functional requirements.

Requirement	Description
Security	The application should be able to authenticate a user before giving them access to interact with system functionalities.
Availability	Application should have short recovery time from failures
Maintainability	The application should be able to accept changes such as adding or removing functionalities
Scalability	The application should scale easily
Robustness	The application should be able to operate normally in case of undesirable situations
Compatibility	The application should work in different software and hardware platforms
Response time	The application should have short response times

3.3 Conceptual design

This section provides the conceptual design of the application and the flow of the DCNN and ML models. The data-driven software application framework presents the interaction of different objects and actors during the operation of the system. The database design is presented using a class diagram that depicts the relationship between various objects that will be interacting during operations. The lion identification activity diagram presents the DCNN model for the individual lion identification process, the population estimation activity diagram presents the model workflow for the lions' population estimation, and the Risk Assessment activity diagram presents an ML model for predicting the risk associated with an individual lion using its features.

3.3.1 Data-driven software application framework

The Data-driven application software framework in (Fig. 2) presents an overall design of how systems components interact to achieve the study output. It is an application software capable of identifying individual lions, monitoring their population, and predicting the risk associated with an individual. In the DDSA framework, there is a DCNN model as a component, this model generates unique codes for the left-side and right-side whiskers images respectively, the unique codes will be stored in the database along with other features of an individual lion that the ecologist will provide. The risk prediction model is also integrated with the DDSA, its function is to predict the risk associated with an individual. The population estimation model estimates the number of lions for better management decisions. Ecologist interacts with a system for data entry and querying results generated by the application software through the integrated DCNN and ML models.

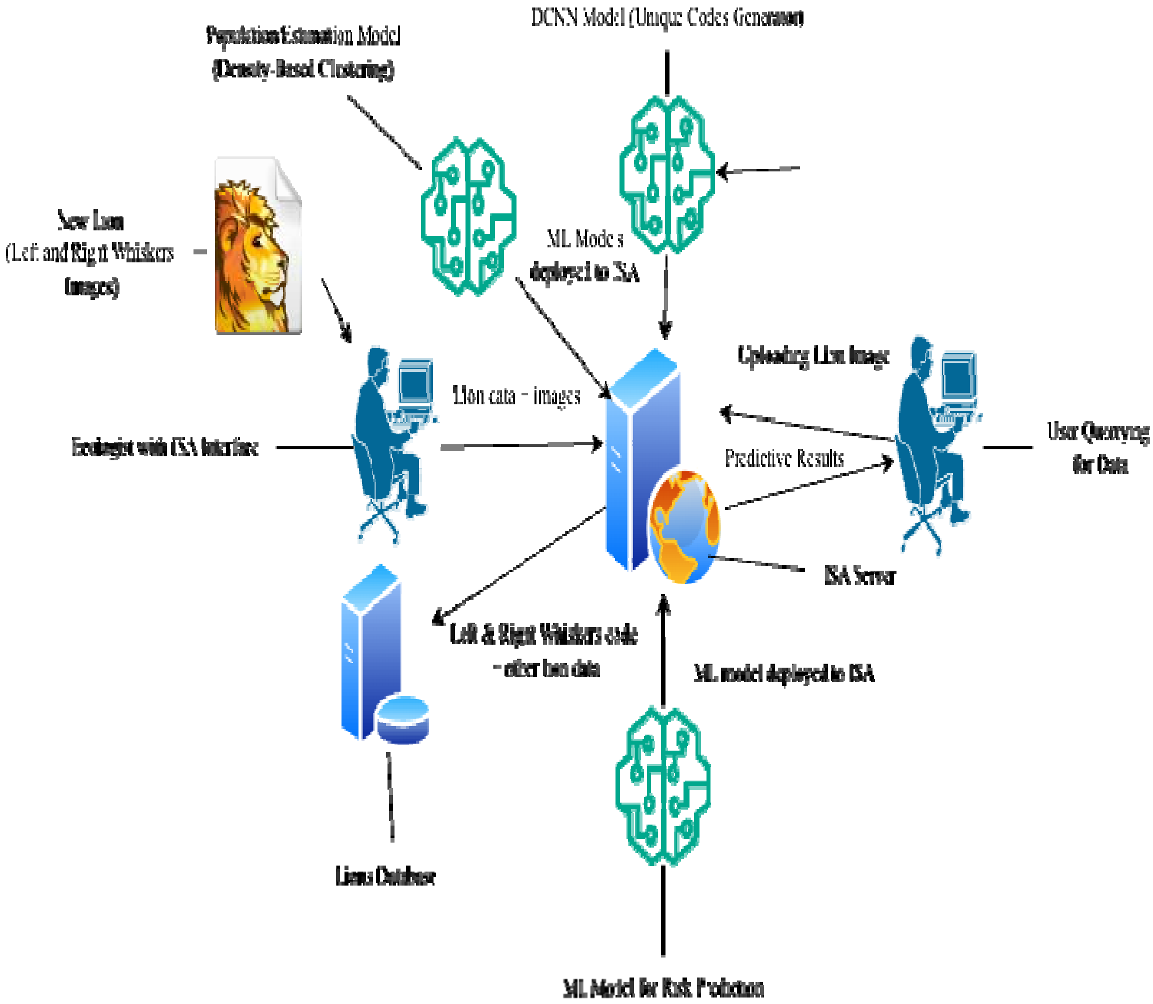

Fig. 2 Data-driven software application framework.

3.3.2 Database design

Database design is presented using a class diagram shown in (Fig. 3), the design shows different objects and the relationship between them. This presented design results from interview data analysis between the researchers and ecologists. It was found that a lion can belong to only one pride and a pride leader can own multiple prides. Further, the ecologist monitors multiple lions where they need to get different details regarding their existence. Lastly, each lion possesses in-born features and other developed features. A lion's scar is also a feature but treated differently because a single lion can have multiple scars and it's important to track when a scar was first seen.

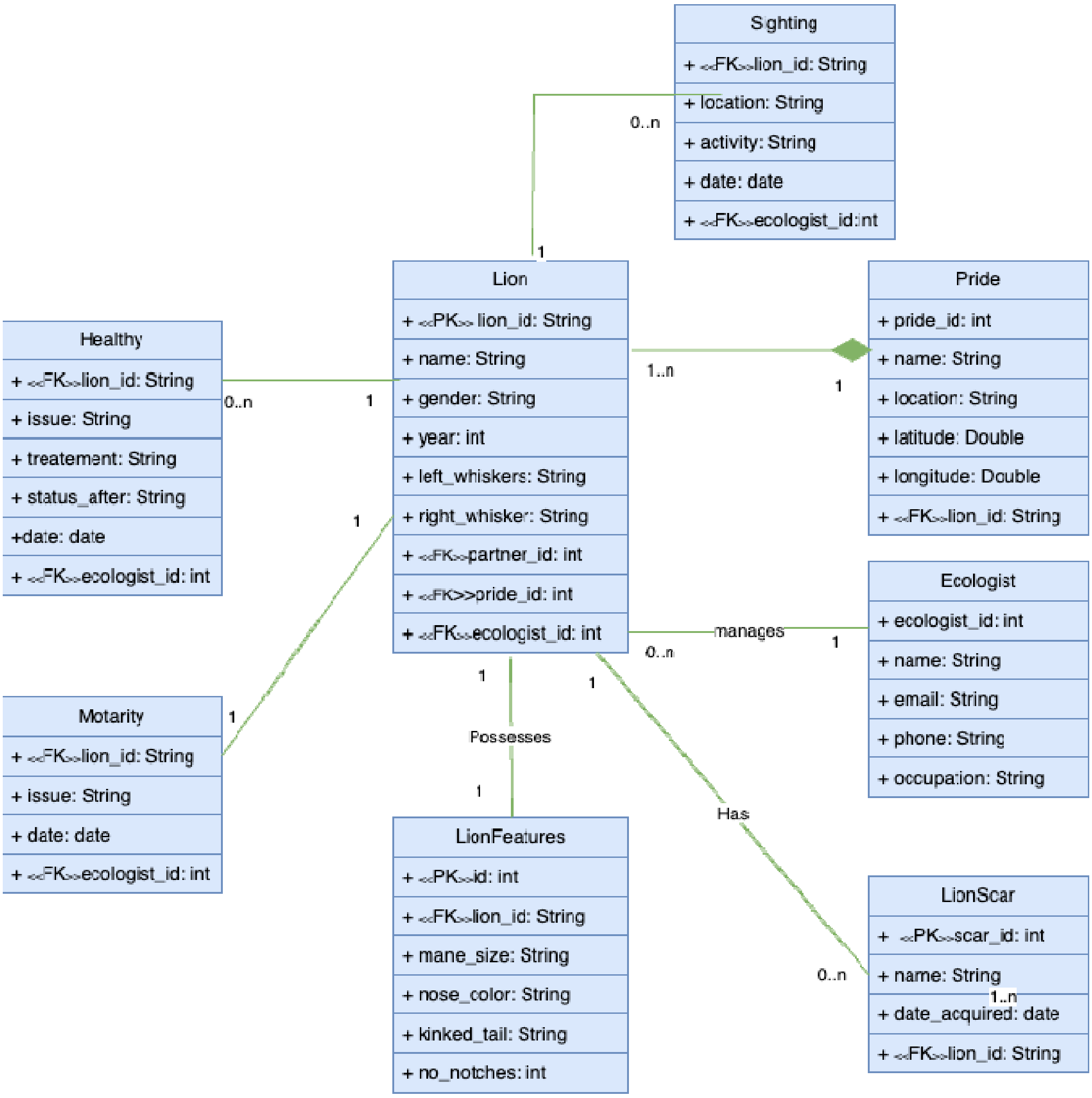

Fig. 3 Database design.

3.3.3 Lion Identification Activity Diagram

The lion identification activity diagram in Fig. 4 presents a series of processes undertaken toward identifying a lion. The first step is accepting lion images; containing right and left whiskers. The whisker detection model processes images for the presence of right and left whiskers; if not detected the application asks for the respective input images otherwise it proceeds to the next step of extracting the whiskers and identification which the DCNN identification model does. The identification model generates unique codes for the right and left whiskers, and searches the database using the codes; if a match is found the predictive results are provided otherwise the individual is registered as new.

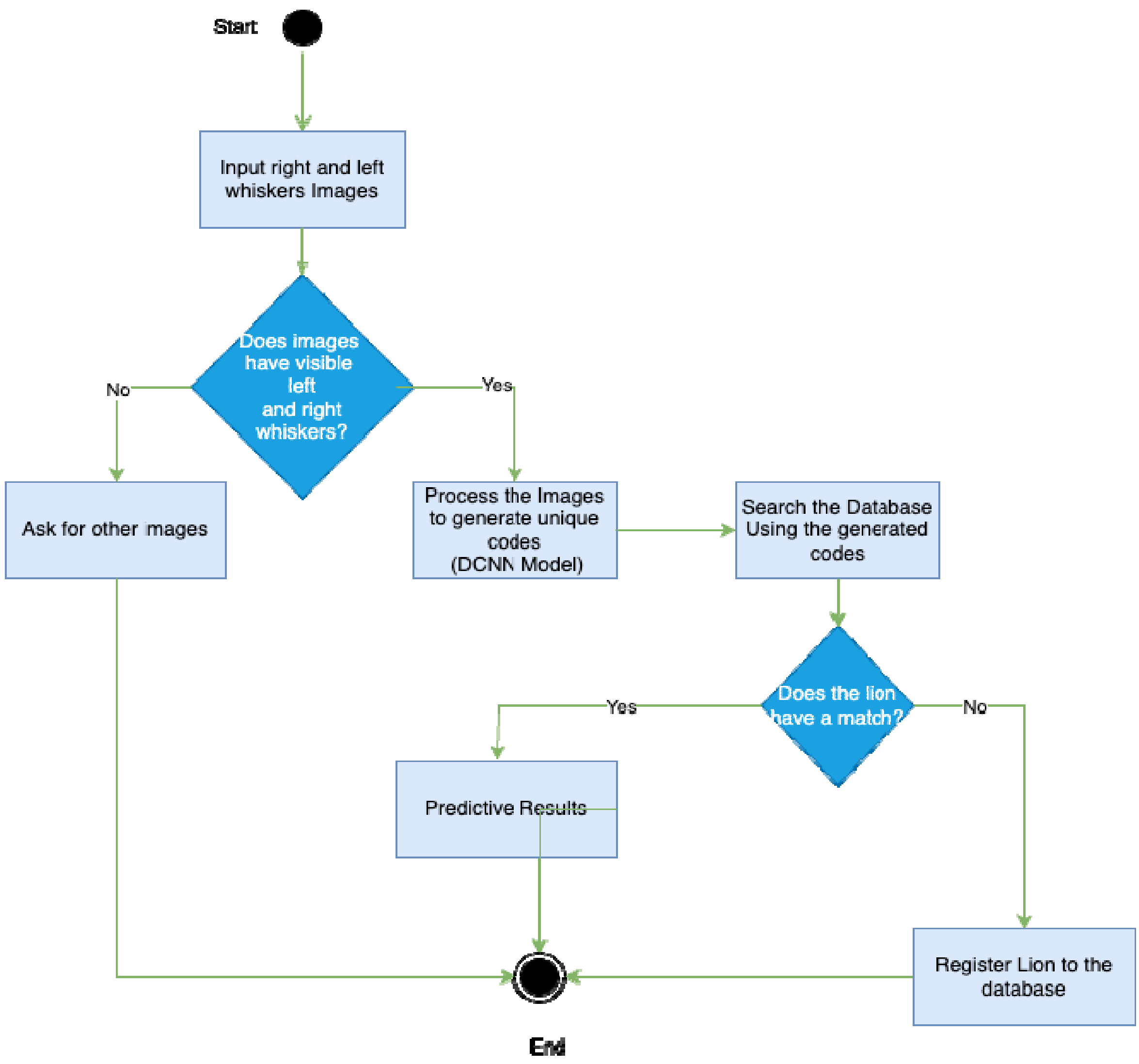

Fig. 4 Lion identification activity diagram.

3.3.4 Population estimation activity diagram

The population estimation activity diagram in Fig. 5 presents the process of estimating the lion population using images of lions. The population estimation model is based on density-based clustering techniques where given a batch of images clusters will be formed based on images of the same individual thus the resulting number of clusters will be an estimation of different individuals.

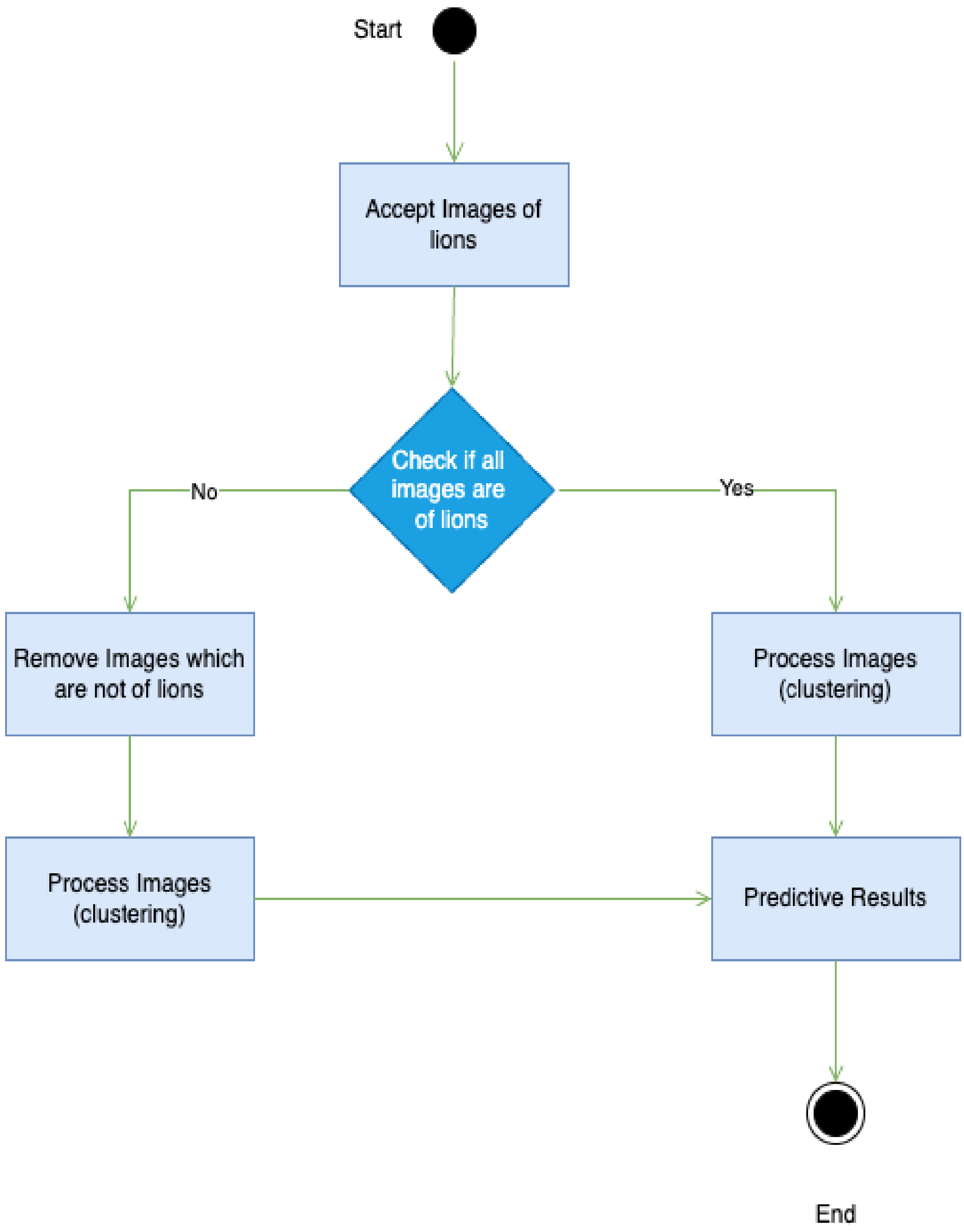

Fig. 5 Population estimation activity diagram.

3.3.5 Risk assessment activity diagram

The risk assessment activity diagram in Fig. 6 presents the process of assessing the risk associated with individual lions. The Risk prediction model will be based on classification where the risk classes will be possible poaching, death, isolation, and loss of pride; and they will be predicted based on features like age, pride location, reported injuries, pride leadership status, and gender.

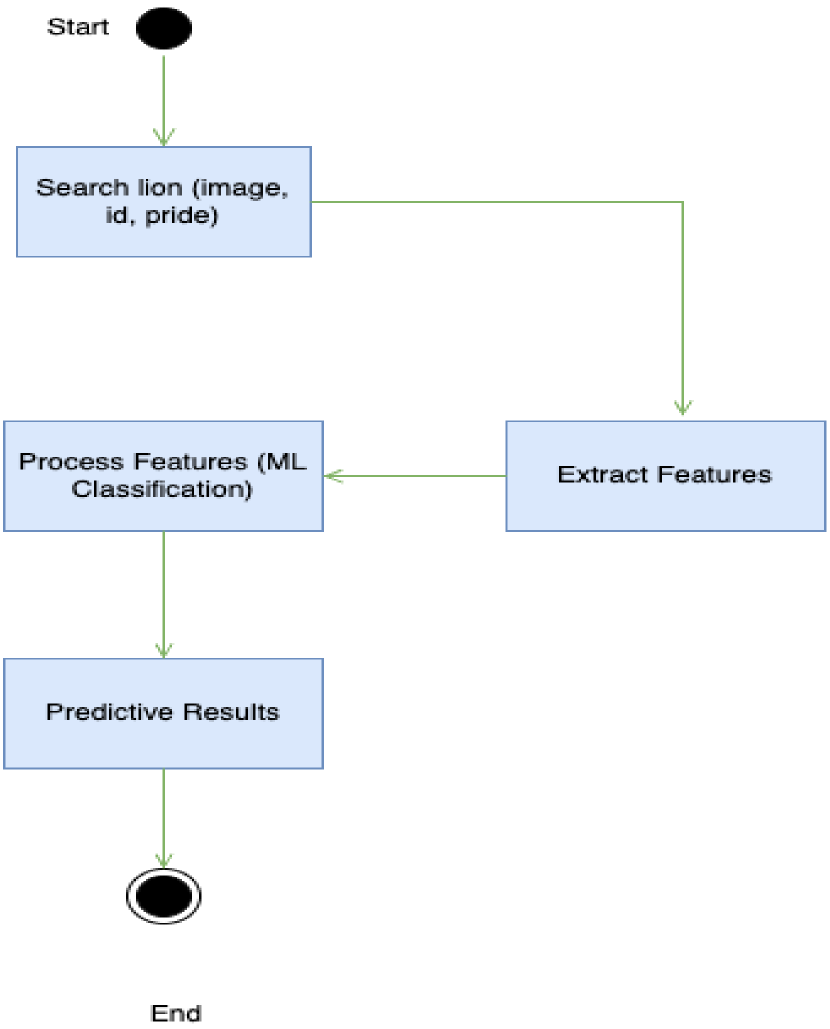

Fig. 6 Risk assessment activity diagram.

4 Discussion

This study aimed at designing a Data-Driven Software Application (DDSA), the findings in this study were significant for the design of the DDSA, which is a way forward for the described problem. Observation and the interview with ecologists at the Serengeti Wildlife Research Centre found that whisker spots from different individual lions with close gene relationships can have a highly correlated pattern that to human eyes looks like the same. This concept underscores the need for advanced lion identification, monitoring, and risk assessment techniques. Using one side of whisker spots can lead to identification challenges which was the case for the study (Hermon and Sharma, 2020) where it considered only one side of the whisker spots.

5 Conclusion

This study found that one side of whisker spots from different individuals can have a highly correlated pattern, making it difficult to distinguish them. The design of a Data-Driven Software Application accounts for both sides of the whisker spots of a lion’s face because they are vital to their identification. This design will be implemented in the next research phase to achieve the overall research goal of developing a Data-Driven Software Application for the identification, population monitoring, and risk assessment of lions in Serengeti National Park Tanzania. This study will enhance monitoring and management activities for the lion populations, helping conservation efforts.

Acknowledgment
We extend our heartfelt gratitude to the Serengeti Wildlife Research Centre for generously sharing their data. The gathered information has significantly enriched our research, allowing us to make informed analyses and drive conservation efforts. We appreciate the dedication and expertise of the Serengeti Wildlife Research Centre in promoting the protection of wildlife and supporting scientific inquiry.

References

Albahri AS, Hamid RA, Alwan JK, Al-qays ZT, Zaidan AA, Zaidan BB, Albahri AOS, AlAmoodi AH, Khlaf JM, Almahdi EM, Thabet E, Hadi SM, Mohammed KI, Alsalem MA, Al-Obaidi JR, Madhloom HT. 2020. Role of biological data mining and machine learning techniques in detecting and diagnosing the Novel Coronavirus (COVID-19): A systematic review. Journal of Medical Systems, 44(7): 122. https://doi.org/10.1007/s10916-020-01582-x

Barocas S, Hardt M, Narayanan A. Fairness and machine learning: Limitations and Opportunities. https://fairmlbook.org/

Bolger DT, Morrison TA, Vance B, Lee D, Farid H. 2012. A computer-assisted system for photographic mark-recapture analysis. Methods in Ecology and Evolution, 3(5): 813-822. https://doi.org/10.1111/j.2041-210X.2012.00212.x

Chowdhary KR. 2020. Fundamentals of artificial intelligence. In Fundamentals of Artificial Intelligence. Springer India. https://doi.org/10.1007/978-81-322-3972-7

Drawio Online. https://app.diagrams.net/. Retrieved June 28, 2024

Hermon GB, Sharma D. 2020. Unique lion identification using triplet loss and siamese networks. Revue d'Intelligence Artificielle, 34(6): 693-700. https://doi.org/10.18280/RIA.340603

Kane MD, Morin DJ, Kelly MJ. 2015. Potential for camera-traps and spatial mark-resight models to improve monitoring of the critically endangered West African lion (Panthera leo). Biodiversity and Conservation, 24(14), 3527–3541. https://doi.org/10.1007/s10531-015-1012-7

LINC - Linc Lion. https://linclion.org/. Retrieved October 19, 2023

Mack N, Woodsong C, Macqueen KM, Guest G, Namey E. 2005. Qualitative Research Methods: A Data Collector'S Field Guide Natashamack. Family Health International.

Maralions. 2023. Mara Predator Project. http://livingwithlions.org/mara/allprides/

Pereira KS, Gibson L, Biggs, D, Samarasinghe D, Braczkowski AR. 2022. Individual identification of large felids in field studies: common methods, challenges, and implications for conservation science. Frontiers in Ecology and Evolution, 10: e866403. https://doi.org/10.3389/fevo.2022.866403

Pucci R, Shankaraiah J, Jathanna D, Karanth U, Subr K. 2020. WhoAmI: An Automatic Tool for Visual Recognition of Tiger and Leopard Individuals in the Wild. http://arxiv.org/abs/2006.09962

Smitz N, Jouvenet O, Ambwene Ligate F, Crosmary WG, Ikanda D, Chardonnet P, Fusari A, Meganck K, Gillet F, Melletti M, Michaux JR. 2018. A genome-wide data assessment of the African lion (Panthera leo) population genetic structure and diversity in Tanzania. PLOS One, 13(11): e0205395. https://doi.org/10.1371/journal.pone.0205395

Soller JM, Ausband DE, Gunther MS. 2020. The curse of observer experience: Error in noninvasive genetic sampling. PLoS One, 15(3): e0229762. https://doi.org/10.1371/journal.pone.0229762

Sommerville I. 2011. Software Engineering. https://doi.org/10.1111/j.1365-2362.2005.01463.x

Strampelli P, Searle CE, Smit JB, Henschel P, Mkuburo L, Ikanda D, Macdonald DW, Dickman AJ. 2022. Camera trapping and spatially explicit capture–recapture for the monitoring and conservation management of lions: Insights from a globally important population in Tanzania. Ecological Solutions and Evidence, 3(1): e12129. https://doi.org/10.1002/2688-8319.12129

Teliolabs. Software with Intelligent Marking-Based Lion Identification. Teliolabs. https://www.teliolabs.com/itsolution/simba-software-with-intelligent-marking-based-identification/. Retrieved October 24, 2023

Venkatesh V. 2022. Adoption and use of AI tools: a research agenda grounded in UTAUT. Annals of Operations Research, 308(1-2): 641-652. https://doi.org/10.1007/s10479-020-03918-9