

# Footprint Biometrics for Reliable Wildlife Identification and Population Estimation

A Case Study on Eurasian Otters



Zur Erlangung des akademischen Grades eines  
**Doktors der Naturwissenschaften (Dr. rer. nat.)**

von der KIT-Fakultät für  
Bauingenieur-, Geo- und Umweltwissenschaften  
des Karlsruher Instituts für Technologie (KIT)

genehmigte

**DISSERTATION**

von

Frederick Sebastian Kistner

Tag der mündlichen Prüfung: 03.02.2026

Referent: Prof. Dr.-Ing. habil. Stefan Hinz

Korreferentin: Zoe Jewell, M.A., M.Sc., Vet., M.B. (Cantab.)

Karlsruhe (2026)





*It all started with a trail*

## Abstract

The world is facing a biodiversity crisis, as the loss of species and natural habitats threatens ecosystem services, climate regulation, and human well-being. Addressing this challenge requires data-driven, scalable, and non-invasive monitoring methods that can be applied reliably by local conservationists without costly equipment. The *Footprint Identification Technology (Footprint Identification Technology (FIT))*, developed by WildTrack, meets these requirements by combining traditional tracker expertise with modern *Machine Learning (ML)* and operating with only a smartphone and a ruler.

This dissertation advances *FIT* using the Eurasian otter (*Lutra lutra*) as a model species of conservation concern. It develops algorithms for species classification, sex determination, individual identification, and population estimation from footprint images, adapting the method to otter-specific challenges while ensuring transferability to other taxa.

We established the largest annotated dataset of Eurasian otter footprints from known individuals, refined data collection protocols, and contributed to app-based and cloud-enabled data workflows. Complementary datasets were created for Asian small-clawed (*Aonyx cinereus*) and smooth-coated otters (*Lutrogale perspicillata*), accompanied by professional reference illustrations.

Methodologically, we implemented morphometric classifiers and the first end-to-end *Convolutional Neural Network (CNN)* for footprint images, both achieving accuracies above 90 %. For Eurasian otters, sex was predicted from single prints with >80 % accuracy and from grouped trails with >90 %, fully consistent with independent *Deoxyribonucleic Acid (DNA)* data. Modified pairwise-comparison workflows improved individual identification to >95 % *balanced accuracy (BA)*, while automated clustering and sequential hold-out validation enhanced population estimation (low *Dendrogram Error Rate (DER)*, high *harmonic mean of homogeneity and completeness (V-measure)*).

To address data scarcity, we evaluated missing-value imputation as a preprocessing step, demonstrating that models including incomplete prints performed comparably to complete-data models (e.g., *Multiple Imputation by Chained Equations (MICE)/Predictive Mean Matching (PMM)* as applicable). Finally, *FIT* was ported from *JMP statistical software package (JMP)* to a modular, *Python* open-source framework, enabling reproducible and automated analyses.

# Zusammenfassung

Die Welt befindet sich in einer Biodiversitätskrise: Der Verlust von Arten und Lebensräumen gefährdet Ökosystemdienstleistungen, Klimaregulation und die menschliche Gesundheit. Um diesen Herausforderungen zu begegnen, braucht es datenbasierte, skalierbare und nicht-invasive Monitoringmethoden, die sich ohne teure Ausrüstung von lokalen Naturschützer:innen anwenden lassen. Die von WildTrack entwickelte *Footprint Identification Technology (FIT)* erfüllt diese Anforderungen, indem sie traditionelles Spurenwissen mit modernen Verfahren des *MLs* kombiniert.

Diese Dissertation entwickelt die *FIT*-Methode am Beispiel des Eurasischen Fischotter (*Lutra lutra*) weiter. Sie umfasst Algorithmen zur Artklassifikation, Geschlechtsbestimmung, Individualerkennung und Populationsschätzung auf Basis von Fußabdruckbildern und passt die Methode gezielt an die Besonderheiten von Ottern an, wobei die Übertragbarkeit auf andere Arten gewährleistet bleibt.

Es wurde der bislang größte annotierte Datensatz bekannter Otterindividuen aufgebaut, Feld- und Zoodaten wurden vereinheitlicht, und app-basierte sowie cloudgestützte Erfassungsprozesse wurden integriert. Ergänzende Datensätze für Zwerg- (*Aonyx cinereus*) und Glattotter (*Lutrogale perspicillata*) sowie professionelle Referenzzeichnungen vervollständigen die Grundlage.

Methodisch wurden morphometrische Klassifikatoren und erstmals ein *CNN*-basierter End-to-End-Workflow für Fußabdruckbilder entwickelt; beide erzielten Genauigkeiten über 90 %. Für Eurasische Fischotter konnte das Geschlecht anhand einzelner Abdrücke mit über 80 % und anhand ganzer Spuren mit über 90 % Genauigkeit bestimmt werden, vollständig konsistent mit genetischen Referenzdaten (*DNA*). Modifizierte paarweise Vergleichsverfahren erreichten bei der Individualerkennung über 95 % *BA*, und automatisierte Clusteranalysen verbesserten die Populationsschätzung (niedrige *DER*, hohe *V-measure*).

Durch den Einsatz fehlwertbasierter Imputation (z. B. *MICE/PMM*) konnten zudem unvollständige Abdrücke in die Analyse einbezogen werden, ohne die Modellleistung zu beeinträchtigen. Abschließend wurde *FIT* von *JMP* in eine modulare, offene *Python*-Struktur überführt, die reproduzierbare und skalierbare Analysen ermöglicht.

## Acknowledgements

This dissertation would not have been possible without the support of many people and institutions to whom I am deeply grateful. The **Deutsche Bundesstiftung Umwelt (DBU)** funded the major part of my PhD through a doctoral scholarship. I further thank **GRACE – Graduate and Research School for Energy, Climate and Environment** for enabling professional development through trainings, conferences, and international stays, as well as a follow-up scholarship after the DBU fellowship had ended. The **Karlsruhe House of Young Scientists (KHYS)** kindly supported this work through a Remote Collaboration Grant that facilitated data collection on additional Asian otter species.

My sincere thanks go to **Prof. Dr.-Ing. Stefan Hinz** for his continuous supervision, valuable feedback, and unwavering support throughout all stages of this work. His empathy and reliability, especially during challenging times, and his effort to ensure continuous funding are gratefully acknowledged.

I thank **Prof. Dr. Markus Pauly** and his team for methodological advice on the imputation of missing values and their valuable input during our joint efforts to handle footprints of varying quality.

I owe special gratitude to **Prof. Zoe Jewell** and **Prof. Sky K. Alibhai** (Duke University) for more than a decade of inspiring collaboration. When I was a much younger student in Portugal, eager to study otters but without funding for more complex monitoring approaches, they introduced me to the FIT method — a simple, precise, and cost-effective technique that became the foundation of this work. Their lifelong dedication to wildlife conservation and to advancing non-invasive, footprint-based monitoring — often in combination with Traditional Ecological Knowledge — has been deeply inspiring. I have always admired their persistence, integrity, and commitment to ethical, evidence-based conservation. FIT would not have evolved into the method and global network it is today without their dedication and the community of like-minded conservationists that has grown under the umbrella of *WildTrack*. I am profoundly grateful for their openness to new ideas, their thoughtful feedback, and the trust and encouragement they have shown throughout this journey. A special thanks goes to Zoe for kindly agreeing to serve as my second reviewer and for her continued support and patience during the many stages of this work.

My thanks also go to all colleagues from the *Otter Specialist Group* and the *WildTrack Specialist Group*, in particular **Larissa Slaney**, **Assaf Ben David**, and **Dr. Amy Fitzmaurice**, whose friendship and exchange have been a steady source of motivation and community. I am equally grateful to all my fellow trackers and tracking teachers who have helped me grow into the tracker I am today — and to those I have yet to meet on my journey to further deepen my understanding of this ancient and remarkable craft.

I am deeply grateful to **Dr. Axel Bamberger** and **Dr. Astrid Blum**, founders and directors of the **CERES Field Centre**, and to Axel, who first sparked my fascination with otters and introduced me to Zoe and Sky. For nearly fifteen years, *CERES* has been far more than a place of collaboration — it has become a second home, shaped by friendship, shared values, and a deep common passion for nature and learning. Axel and Astrid have become true friends and role models to me in many ways. They often choose to do things differently — and it is precisely this authenticity and courage to follow their own path that make them, and *CERES*, such a unique and inspiring place.

I would like to thank **Dr. Susanne Benz** and **Priv.-Doz. Dr. Sina Keller**, as well as all colleagues at the **Institute of Photogrammetry and Remote Sensing (IPF)** at **KIT**. Even though I worked mostly remotely, I greatly appreciated the helpful advice and inspiring conversations during my visits and coffee breaks at the institute. Susanne's guidance during the final phase of my PhD helped me stay focused, while Sina's efficiency and sense of organisation have been truly inspiring. The IPF has been my academic home within KIT — first as a docking station during independent research, and now as part of Susanne's **GRUSS** group.

I am also grateful to **all zoos and wildlife centres** whose otters contributed to this research, and to **all the otters and other animals** whose footprints I was fortunate to follow — across many beautiful places where their presence quietly endures in tracks and traces, each one telling its own small part of a larger story. Finally, I dedicate a special thought to the unknown otter spirit whose trail I have not yet crossed, but to whom I may one day return — perhaps at that very place where the final photo of this dissertation was taken.

Last but not least, my family — my wonderful daughters **Malou** and **Emmi**, and above all my wife **Rebecca**. Thank you for enduring all the otter talk and the phases of hyperfocus that at times interfered with my ability to be the father and husband you all deserve. If someone had told us beforehand everything that would happen during this journey, I might never have started it — and without you, it would never have been completed the way it has. Thank you!

# Contents

Abstract . . . . .	iv
Zusammenfassung . . . . .	v
<b>Part I Background and Motivation</b>	<b>1</b>
<b>1 Introduction: Rationale and Objectives</b>	<b>3</b>
1.1 Biodiversity Crisis and Monitoring Gaps . . . . .	3
1.2 Non-Invasive Monitoring and the Role of Citizen Science . . . . .	3
1.3 Otter Conservation as a Framework for Non-Invasive Monitoring . . . . .	4
1.4 Research Objectives and Dissertation Structure . . . . .	6
<b>2 State of the Art</b>	<b>9</b>
2.1 State of the Art in Otter Monitoring . . . . .	9
2.2 Footprint Identification Technology (FIT) . . . . .	15
2.3 Current Challenges in the Field of Footprint Monitoring: Particular Focus on FIT . . . . .	16
2.4 Positioning within Citizen Science and Biodiversity Monitoring . . . . .	25
2.5 Chapter Summary . . . . .	25
<b>Part II Primary Data Collection, Field Campaigns, and Data Annotation</b>	<b>27</b>
<b>3 Data Collection in Zoos and Field Campaigns</b>	<b>29</b>
3.1 Ethics Statement . . . . .	29
3.2 Overview and Objectives . . . . .	29
3.3 Development of Data Collection Protocols and Tools . . . . .	30
3.4 Data Collection . . . . .	31
3.5 Morphometric Benchmark Datasets for Evaluating FIT Pipeline Im- provements . . . . .	36
3.6 Additional Datasets from Publications Originating in this Dissertation	38
3.7 Chapter Summary . . . . .	39

<b>4</b>	<b>Methods A: Data Annotation and Exploration</b>	<b>41</b>
4.1	Generation of Reference Track Drawings . . . . .	41
4.2	Annotation . . . . .	42
4.3	Feature Generation . . . . .	43
4.4	Data Exploration . . . . .	44
4.5	Chapter Summary . . . . .	48
<b>Part III</b>	<b>Analytical Workflow and Results</b>	<b>49</b>
<b>5</b>	<b>Methods B: Analytics Pipeline and Experimental Design</b>	<b>51</b>
5.1	Original Pairwise Method in JMP for FIT . . . . .	51
5.2	Modular Modelling Framework for Classification and Population Esti- mation . . . . .	51
5.3	Evaluated Methodological Improvements to Morphometric FIT . . . . .	63
5.4	Chapter Summary . . . . .	64
<b>6</b>	<b>Results</b>	<b>65</b>
6.1	Professional Tracker-Based Analysis and Description of Footprints . . . . .	66
6.2	Morphometric Analyses . . . . .	67
6.3	Benchmarking with Other Species . . . . .	76
6.4	Improvements to the <i>FIT</i> Workflow . . . . .	77
6.5	Chapter Summary . . . . .	79
<b>Part IV</b>	<b>Synthesis, Critical Discussion, and Future Perspectives</b>	<b>81</b>
<b>7</b>	<b>Discussion</b>	<b>83</b>
7.1	Summary of Main Achievements . . . . .	83
7.2	General Improvements of FIT and Remaining Challenges When Work- ing With Footprints . . . . .	89
7.3	Synthesis and Final Discussion . . . . .	91
<b>8</b>	<b>Conclusion and Outlook</b>	<b>93</b>
8.1	Future Work and Outlook . . . . .	93
8.2	Final Reflections – On Tracking and Knowing the World . . . . .	100
	<b>References</b>	<b>103</b>
	<b>Appendix</b>	<b>115</b>



# Glossary

**3×4 lope gait** A characteristic quadrupedal gait pattern of otters, consisting of three front and four hind print groupings per stride sequence. It typically produces paired tracks with overlapping fore- and hind-foot impressions, distinguishing otters from similar-sized carnivores such as dogs or martens.. 36

**Python** Open-source programming language widely used for data science, machine learning, and reproducible research pipelines. iv, v, 24, 51, 55, 63, 64, 65, 83, 89, 90, 91, 97, 114

**spraint** A faecal deposit of otters, typically placed at conspicuous sites such as rocks, bridge bases, or riverbanks. Spraints serve as territorial markers and communication cues and are key indicators in otter monitoring, often used for presence confirmation and DNA-based analyses.. 36, 39

**WildTrack** An international non-profit organisation dedicated to developing non-invasive monitoring approaches for endangered species, including the Footprint Identification Technology (*FIT*).. 21, 24

**WildTrackAI** A cloud-based extension of the WildTrack initiative that integrates artificial intelligence (*Artificial Intelligence (AI)*) and machine learning (*ML*) for scalable footprint data collection, annotation, and analysis.. 21, 25, 86



# Acronyms

- AI** Artificial Intelligence. xi, 6, 7, 9, 14, 16, 18, 21, 23, 25, 39, 86, 87, 91, 95, 98, 100
- ANOVA** Analysis of Variance. 45, 48
- BA** balanced accuracy. iv, v, 61, 64, 69, 73, 76, 77, 78, 79
- CERES** Conservation and Environmental Research & Education – International Project (field centre, SW Portugal). 35, 39, 99
- CNN** Convolutional Neural Network. iv, v, 15, 19, 23, 38, 41, 83, 89
- COCO** Common Objects in Context (computer vision dataset / annotation format). 86
- CV** Cross-Validation. 53, 55, 57, 58, 60, 63, 64, 65, 67, 69, 73, 76, 77, 79, 113
- DER** Dendrogram Error Rate. iv, v, 61, 64, 76, 78, 79
- DL** Deep Learning. 63, 86, 88, 89, 95
- DNA** Deoxyribonucleic Acid. iv, v, 4, 6, 7, 9, 12, 15, 25, 31, 32, 37, 39, 42, 53, 62, 64, 83, 84, 93
- eDNA** Environmental DNA. 3, 9
- FAIR** Findable, Accessible, Interoperable, Reusable (data principles). 85, 86, 97
- FFH** Fauna–Flora–Habitat Directive of the European Union. 9, 32, 39
- FIT** Footprint Identification Technology. iv, v, ix, xi, xvii, 4, 6, 7, 14, 15, 16, 18, 19, 21, 23, 24, 25, 29, 30, 31, 32, 36, 37, 38, 39, 41, 42, 43, 44, 48, 51, 53, 56, 57, 58, 60, 63, 64, 65, 67, 69, 77, 79, 83, 84, 85, 87, 88, 89, 90, 91, 93, 97, 99, 112, 114
- GSD** Ground Sampling Distance. 39, 97

**GUI** Graphical User Interface. 15

**ID** Identification. 5, 31, 46, 48, 53, 55

**IOSF** International Otter Survival Fund. xvii, 93

**IUCN OSG** International Union for Conservation of Nature Otter Specialist Group. 9, 16, 93

**JMP** JMP statistical software package. iv, v, 24, 36, 42, 43, 48, 51, 55, 57, 58, 60, 63, 64, 65, 83, 86, 89, 90, 91, 97, 114

**JSL** JMP Scripting Language. 24, 43, 48

**LASSO** Least Absolute Shrinkage and Selection Operator. 57, 63, 77, 79

**LDA** Linear Discriminant Analysis. 14, 51, 56, 57, 58, 60, 63, 64, 79, 87, 88

**LF** Left Front (foot). 41, 42, 48

**LH** Left Hind (foot). 41, 48

**MCAR** Missing Completely At Random. 77, 78

**MICE** Multiple Imputation by Chained Equations. iv, v, 19, 78

**ML** Machine Learning. iv, v, xi, 6, 9, 14, 15, 23, 35, 36, 37, 39, 55, 62, 63, 83, 87, 100

**MobileNet** MobileNet convolutional neural network architecture (lightweight model optimised for mobile and embedded vision applications). 95

**OOF** Out-of-Fold (validation setting). 63

**PCA** Principal Component Analysis. 45, 58, 60, 63, 64

**PMM** Predictive Mean Matching. iv, v, 19, 78

**QA/QC** Quality Assurance / Quality Control. 39, 97

**QC** Quality Control. 21, 23, 36, 62, 84, 97

**RCV** Reference Centroid Value. 51, 53, 55, 58

**RF** Random Forest (model context). 57, 63, 77

**RFF** Right Front (foot). 42

**ROC–AUC** Receiver Operating Characteristic – Area Under the Curve. 62

**SD** Standard Deviation. 65, 76, 77, 78

**TEK** Traditional Ecological Knowledge. 6, 87

**TNR** true negative rate. 61

**TPR** true positive rate. 61

**UAV** Unoccupied Aerial Vehicle (drone). 97

**UMAP** Uniform Manifold Approximation and Projection. 45, 48

**V-measure** harmonic mean of homogeneity and completeness. iv, v, 61, 64, 79

**WSG** WildTrack Specialist Group. 18

**XAI** Explainable Artificial Intelligence. 18

**XGBoost** eXtreme Gradient Boosting. 15, 60, 64, 79

**YOLO** You Only Look Once (object detection architecture). 95, 96



## Preamble

This dissertation builds upon research undertaken during the doctoral studies, focusing on the advancement of the *FIT* and related methods for non-invasive population monitoring. Its main emphasis lies on new, previously unpublished material that integrates and extends the results of earlier research within a coherent framework.

Three first-author, peer-reviewed papers have been published one of these received the *International Otter Survival Fund (IOSF)* “Otter Oscar” in 2022. Additional unpublished material presented in this dissertation will be submitted for future peer-reviewed publications.

For coherence and transparency, some sections contain verbatim excerpts from the published papers listed above. These passages are visually marked in the margins with colour-coded change bars indicating their source.

All peer-reviewed publications are included in full in the **digital version** of this dissertation as part of the appendix, while the **printed version** omits these publications to reduce length and focus on the integrative synthesis.

- **Blue** – *It’s Otterly Confusing – Distinguishing Between Footprints of Three of the Four Sympatric Asian Otter Species Using Morphometrics and Machine Learning*, *IUCN/SCC Otter Specialist Group Bulletin*, Kistner et al. 2022
- **Purple** – *Can You Tell the Species by a Footprint? – Identifying Three of the Four Sympatric Southeast Asian Otter Species Using Computer Vision and Deep Learning*, *IUCN/SCC Otter Specialist Group Bulletin*, Kistner et al. 2023
- **Red** – *Enhancing Endangered Species Monitoring by Lowering Data Entry Requirements with Imputation Techniques as a Preprocessing Step for the Footprint Identification Technology (FIT)*, *Ecological Informatics*, Kistner et al. 2024

## Terminology Used in This Dissertation

- **Track.** A single impression of an animal's foot (also called a *footprint* or *spoor*).
- **Trail.** A continuous sequence of successive tracks (footprints) made by an individual moving across a substrate; one trail typically consists of multiple tracks forming a measurable pattern.
- **Track&Sign.** A broader term referring to any sign of an animal's presence, including footprints, scat, or hair.
- **FIT.** The Footprint Identification Technology, a framework for individual and population-level identification based on geometric measurements extracted from digital footprints.

# Part I

---

Background and Motivation



# Introduction: Rationale and Objectives

## 1.1 Biodiversity Crisis and Monitoring Gaps

Earth is facing a biodiversity crisis, with current extinction rates estimated to be up to 1,000 times higher than natural levels and projected to rise, impacting species and ecosystems [1]. According to the Kunming-Montreal Global Biodiversity Framework, 195 countries have pledged to stop biodiversity loss by 2030 and reverse it by 2050 [2]. Achieving these targets requires reliable, comprehensive, and cost-effective data on threatened species populations [3–5]. Various monitoring methods exist, each with its own pros and cons, but non-invasive approaches should be prioritised for ethical and animal welfare reasons [6].

## 1.2 Non-Invasive Monitoring and the Role of Citizen Science

Traditional monitoring methods such as telemetry, genetic sampling, and camera trapping provide detailed data but are costly, complex, and often invasive [7, 8]. Non-invasive approaches—including *Environmental DNA (eDNA)* and track or sign analysis—offer scalable and sustainable alternatives [6, 9–11]. Footprints are particularly abundant and information-rich and can be systematically collected by researchers and citizen scientists. Recent advances in mobile applications [12, 13] and cloud-based platforms [14] now enable large-scale participation by local and indigenous communities, contributing diverse datasets that enhance biodiversity monitoring and conservation planning [15–17].

## The Footprint Identification Technology (FIT)

The *FIT*, developed in 2001 [18], is a non-invasive technique for wildlife monitoring based on morphometric analysis of footprints. It has been applied to more than 25 species worldwide with high cross-validated accuracy [19–23]. The method uses geometric measurements derived from anatomical landmarks on digital footprints to classify individuals or species, offering a low-cost complement to *DNA*-based approaches [24, 25]. This dissertation builds on a pilot study for the Eurasian otter (*Lutra lutra*) to advance and adapt *FIT* for broader, scalable application. By refining analytical workflows and improving accessibility, *FIT* can contribute to a unified framework for non-invasive monitoring across diverse taxa and environments.

### 1.3 Otter Conservation as a Framework for Non-Invasive Monitoring

The Eurasian otter declined severely across Europe during the 20th century [26]. Since the early 2000s, improved water quality, habitat restoration, and legal protection have enabled recolonisation. As populations recover, conflicts with aquaculture and fisheries have re-emerged [27, 28]. Monitoring remains challenging as otters are elusive, semi-aquatic, and nocturnal. Standard large-scale monitoring provides only presence data [29], whereas *DNA*-based methods offer higher taxonomic precision but are costly and labour-intensive [30]. Reliable, scalable monitoring solutions are therefore essential to support conservation management and conflict mitigation across Europe.

Thirteen otter species occur worldwide, many threatened by habitat loss, pollution, and human conflict [31]. In Southeast Asia, four species overlap geographically, complicating field identification due to morphological similarity and cryptic behaviour [32]. In such contexts, the *FIT* provides a practical, non-invasive tool for reliable species identification where visual or *DNA*-based evidence is scarce [25]. Establishing *FIT* within otter research thus offers a transferable model for improving non-invasive monitoring in other elusive carnivores.

### Rationale for Using Otters as a Model System

Otters possess several characteristics that make them particularly suitable for the development and evaluation of non-invasive, footprint-based monitoring approaches:

**Wide distribution:** The Eurasian otter ranges from Western Europe to East Asia, spanning diverse ecological gradients, substrate types, and management regimes [26].

**High conservation relevance:** All otter species are of conservation concern, and recurring conflicts with human stakeholders make reliable monitoring data both ecologically and politically significant [26].

**Low detectability:** As elusive, semi-aquatic, and nocturnal carnivores, otters are difficult to monitor through direct observation or camera trapping [26, 31].

**Rich track information:** Otters regularly leave footprints in soft substrates near waterbodies, providing abundant and information-rich material for track-based *Identification (ID)* and morphometric assessment [26].

**Methodological transferability:** Monitoring challenges—such as data scarcity, field constraints, and species misidentification—closely resemble those encountered in other elusive carnivores, facilitating the transfer of developed techniques to broader applications.

## 1.4 Research Objectives and Dissertation Structure

The central aim of this dissertation is to advance the *FIT* as a robust, non-invasive tool for wildlife monitoring, using the Eurasian otter as a focal species and model system. The work addresses the urgent need for scalable, cost-efficient, and reliable monitoring methods in conservation biology and is embedded in an interdisciplinary framework that combines ecology, statistics, *AI*, and *Traditional Ecological Knowledge (TEK)*.

### Research Objectives and Scope

(1) **Developing and field-testing *FIT* for Eurasian otters.** The primary goal is to establish *FIT* as an operational monitoring tool by developing classification models for *species*-, *sex*-, and *individual identity*-level analyses, and by deriving population estimates through clustering. Field trials provide evidence-based recommendations for applying *FIT* in practice.

(2) **Advancing *FIT* methodology beyond a single species.** The second goal is to expand the general *FIT* framework, using otters as a model for developing transferable innovations across taxa. This includes reducing data requirements, embedding *FIT* in modern *ML* workflows, and benchmarking against applications in other threatened species.

To meet these aims, the dissertation pursues the following specific objectives:

- **Dataset development and data collection:** Create expert-labelled footprint datasets and support app-based tools and cloud infrastructure through standardised protocols, user feedback, and training, enabling scalable, high-quality data acquisition.
- ***FIT* model evaluation:** Assess classical *FIT* models for *species*, *sex* and *individual identification* using rigorous validation and external confirmation through genetic *DNA*-referenced data.
- **Methodological innovation:** Develop a complementary open-source back-end to existing *FIT* applications, embedding it within automated *ML* pipelines and testing approaches to reduce data requirements and enhance accessibility.
- **Deep learning and interdisciplinary integration:** Explore *AI*-driven pipelines and integrate insights from ecology, artificial intelligence, statistics, and *TEK* to guide scalable, real-world footprint monitoring.

## Dissertation Structure

This dissertation centres on morphometric footprint analysis and the methodological advancement of the *FIT*, embedding *AI*-readiness as a guiding principle. Parallel exploratory work on deep learning—particularly regarding data collection, processing, and annotation—is briefly summarised, with detailed accounts planned for subsequent publications.

The thesis is organised into four parts:

**Part I: Background and Motivation** — Chapter 1 introduces the biodiversity crisis, advocates scalable, non-invasive monitoring with otters as a model species, and defines the research aims. Chapter 2 reviews footprint-based monitoring methods, outlines current limitations of the *FIT*, and discusses innovations such as *AI*-driven classification and data imputation. Key methodological gaps within the *FIT* framework are identified alongside solutions disseminated through peer-reviewed publications and conference contributions.

**Part II: Primary Data Collection, Field Campaigns, and Data Annotation** — Chapter 3 details the acquisition of high-quality footprint datasets from captive and wild otters, supplemented by benchmark data from other species and targeted field campaigns designed to refine *FIT* protocols. Chapter 4 introduces standardised reference drawings for consistent landmark placement, systematic annotation workflows, and app-based tools supported by cloud infrastructure to enable scalable, collaborative data management.

**Part III: Analytical Workflow and Results** — Chapter 5 presents the modelling framework for morphometric *FIT*-based classification, encompassing *species*-, *sex*- and *individual identification*-level models, supported by genetic *DNA*-validation and statistical imputation strategies for incomplete datasets. Chapter 6 reports model performance under rigorous validation schemes, implemented through automated and reproducible pipelines applicable to both research and applied conservation contexts.

**Part IV: Synthesis, Critical Discussion, and Future Perspectives** — Chapter 7 synthesises the main findings and methodological advances, evaluating the broader applicability of *FIT* for wildlife monitoring. Chapter 8 outlines prospective developments, including automated pipelines, integration with *AI*-based approaches, and community-driven implementation frameworks.



## State of the Art

### 2.1 State of the Art in Otter Monitoring

#### Current Challenges and the Need for Non-Invasive Approaches

Monitoring Eurasian otter populations is essential for conservation but remains difficult due to their elusive, semi-aquatic lifestyle [31, 33]. Standard protocols under the *Fauna–Flora–Habitat Directive of the European Union (FFH)* rely mainly on indirect signs such as spraints and footprints [26, 29], which confirm presence but do not allow population estimates or insights into territorial behaviour. Invasive techniques like radio telemetry yield detailed spatial data [34–36], yet are costly and raise welfare concerns [8, 26, 37].

This has led to growing demand for scalable, non-invasive methods. The *International Union for Conservation of Nature Otter Specialist Group (IUCN OSG)* highlights this need in its global strategy, particularly emphasising genetic approaches [31]. While *eDNA* and spraint-based *DNA* analyses can inform on population size, sex ratio, and genetic diversity [28, 30, 38], their success depends on sample quality and laboratory infrastructure. Camera traps offer another non-invasive option but face otter-specific challenges such as low infrared detectability after water contact [39].

Footprint-based approaches combine ecological validity with potential for both individual and population-level inference. Supported by modern *ML* and *AI*, they provide a scalable, cost-efficient complement—or, in some contexts, an alternative—to genetic monitoring [31]. Table 2.1 summarises current otter monitoring methods, highlighting non-invasive techniques and scalability.

**Tab. 2.1.:** Summary of current monitoring methods in wildlife monitoring with special focus on Eurasian otters and non-invasive approaches (*Lutra lutra*), ordered by increasing scalability.

Method	Key points (Strengths & Limitations)
<b>Telemetry</b>	<ul style="list-style-type: none"> <li>+ Detailed space use, movement ecology and habitat selection; long-term individual tracking. [35, 36]</li> <li>– Variable fix accuracy; signal impaired by water, topography, dense canopy or built environments. [26, 35]</li> <li>– Limited scalability and invasiveness: high costs/effort; few individuals; live-trapping required; implants carry welfare risks; harness units may affect behaviour/swimming and need retrieval logistics. [35, 36]</li> </ul>
<b>Camera traps</b>	<ul style="list-style-type: none"> <li>+ Non-invasive, relatively affordable; standardised, automated long-term data. [40, 41]</li> <li>+ Direct visual records enable behavioural/social insights (e.g., otter latrines, resting sites). [42, 43]</li> <li>+ Multi-species detections support community-level inference. [40]</li> <li>+ – Large data volumes; resource-intensive without efficient/AI-assisted workflows. [41]</li> <li>+ – Individual ID/density only conditionally feasible (natural marks; rigorous capture–recapture/SECR); sensitive to survey design and detectability. [41]</li> <li>– Point sampling: dense arrays needed for high spatial resolution; detection highly placement-sensitive. [40, 44]</li> <li>– False negatives common; in semi-aquatic otters PIR trigger probability drops with wet fur and distance. [39, 42, 43]</li> </ul>
<b>DNA / eDNA</b>	<ul style="list-style-type: none"> <li>+ Individual identification and sexing; supports capture–recapture population estimates. [30, 45]</li> <li>+ Effective for discriminating sympatric otter species when signs are scarce/ambiguous. [46]</li> <li>– Spraint DNA degrades quickly; success depends on freshness, storage and low temperatures. [30, 45]</li> <li>– Lab infrastructure/expertise and costs limit scalability; detection/genotyping can be sex-biased (males over-represented). [26, 46]</li> </ul>

*Continued on next page*

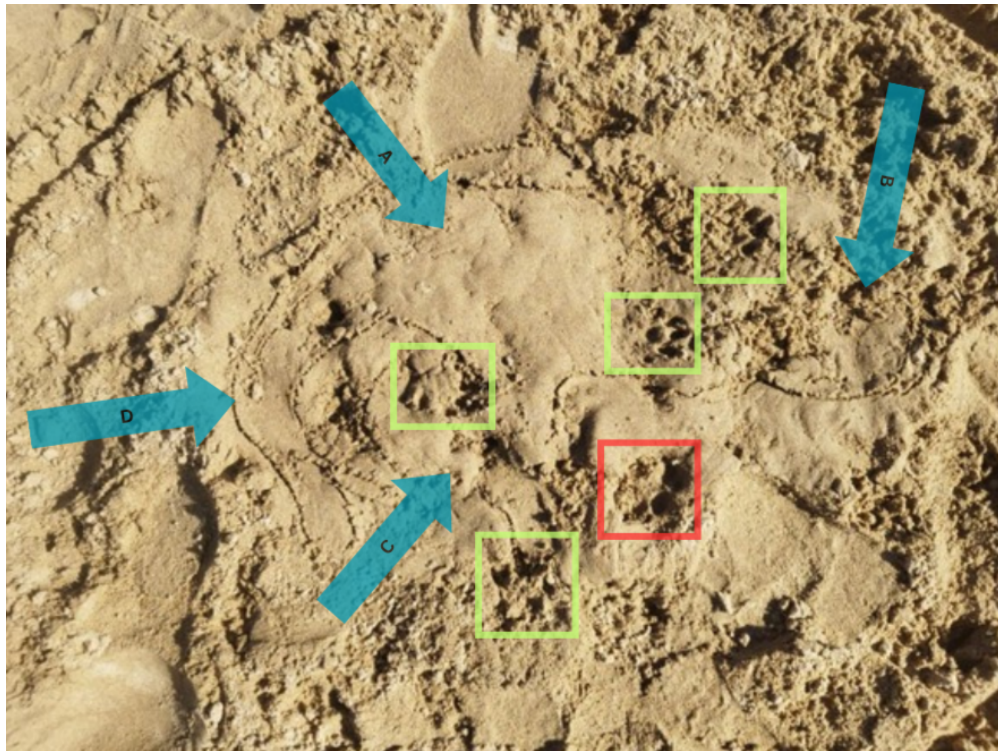
Method	Key points (Strengths & Limitations)
<b>Morphometric analyses including FIT</b>	<ul style="list-style-type: none"> <li>+ Applied to species, sex and individual identification across taxa; peer-reviewed; promising preliminary otter results. [47–51]</li> <li>+ Low equipment/operational cost; amenable to community science, enabling scalability. [22, 24]</li> <li>+ In suitable substrates, footprint sequences can yield movement information at a higher spatial resolution than collar signal localisation (radio-collar triangulation or GPS), particularly where conventional collar tracking performs poorly. [22, 36]</li> <li>– Success depends on tracking skills and substrate; strict print-quality criteria; anatomical expertise and time-consuming landmarking can be limiting. [22, 52]</li> </ul>
<b>Sprints and footprints (without morphometrics)</b>	<ul style="list-style-type: none"> <li>+ Historic use for sex/individual ID, population indices and territoriality/habitat-use insights. [26, 53–55]</li> <li>+ Standardised presence–absence protocols for Natura 2000; cost-effective and scalable; can combine with DNA, morphometrics and community science. [24, 26, 29]</li> <li>– Individual ID generally not possible without specialised methods; observer bias; population size hard to infer robustly. [26, 32]</li> </ul>
<b>AI-based methods</b>	<ul style="list-style-type: none"> <li>+ Scalable processing of large image/footprint datasets; automated species and (for some taxa) individual recognition; explainable AI (XAI) available. [49]</li> <li>+ With standardised protocols and QA/QC, deployable at scale; for otters, robust individual ID remains conditional on image quality and stable natural marks. [41]</li> <li>– Need annotated training data and higher compute; performance affected by substrate/lighting/viewpoint; interpretability lower without XAI. [41, 49]</li> </ul>
<b>Community science</b>	<ul style="list-style-type: none"> <li>+ App-based/community surveys increase scalability compared to expert-only surveys; well suited to footprint/sign surveys when clear protocols are provided. [12, 13, 41, 56]</li> <li>+ Combining community-sourced data collection with AI-assisted processing and analytics offers high scalability potential. [41]</li> <li>– Data quality is variable; requires training, feedback, and QA/QC; sampling bias must be modelled/controlled. [41]</li> <li>– Standardised protocols are needed to enable integration into official large-scale monitoring schemes. [41]</li> </ul>

## Traditional Tracking and Field-Based Methods

Animal tracking is as old as humanity and has been described as the origin of empirical science [9]. While once essential for survival, traditional tracking skills have largely faded in modern contexts. In ecological research, footprints are often reduced to presence–absence indicators, despite containing much richer behavioural and demographic information. Yet in remote or technologically underdeveloped regions, this traditional method remains indispensable, providing valuable data for conservation, wildlife management, and ecological research [41, 57].

Footprint-based monitoring has long provided a cost-efficient, non-invasive method for studying otters. Seminal work by Sam Erlinge demonstrated that individual otters could be identified from plaster casts of their tracks in multi-year studies in southern Sweden, yielding valuable insights into home ranges, territorial boundaries, and sex composition [53, 54]. Subsequent studies explored complementary indicators such as urine traces for sex determination [58], sexual size differences in tracks [59], and snow tracking, which in northern regions often produced population estimates comparable to *DNA* analyses [45, 55]. Being able to read and interpret tracks and signs can therefore provide valuable insights into animal behaviour that are otherwise difficult to observe or capture with other monitoring methods. An example is shown in Figure 2.1, which depicts the impression left by a resting otter on a riverbank.

A comprehensive overview of the current and historical use of tracks and signs beyond otters as a research method is, for instance, provided by Dorfman et al. [57], who argue that tracking is not an outdated method and outline how it can be combined with modern technologies. These points are further elaborated in the following sections.



**Fig. 2.1.:** Contour imprint of a Eurasian otter (*Lutra lutra*) left after the still-wet animal briefly rested—likely at night or in the early morning. The outer outline was carefully traced in the sand to aid documentation without altering the original structure. Blue arrows (clockwise): **A** body impression; **B** head impression; **C** right hind-leg impression; **D** tail impression. **Green boxes:** footprints of all four otter feet. **Red box:** dog footprint (not from the otter). Recorded during a CyberTracker course (spring 2023) led by Senior Tracker José María Galán; discovered and photographed by the author. This figure underscores how skilled tracking can reconstruct behaviour from subtle sign using non-invasive observation. After rising, the otter moved toward the upper-right edge of the frame.

## Digital Morphometrics and Classification Techniques

Building on the field-based foundations of traditional tracking, digital morphometric techniques—quantitative analyses of shape and structure derived from standardised footprint landmarks—have emerged that translate footprint evidence into robust data suitable for advanced statistical and **ML** analyses. Across taxa, footprint-based morphometrics have demonstrated their versatility and reliability.

Morphometric methods have been applied for sex determination in pine martens (*Martes martes*) [60], age-class modelling in giant pandas (*Ailuropoda melanoleuca*) [22], and individual identification in black rhinoceroses (*Diceros bicornis*), white rhinoceroses (*Ceratotherium simum*), cheetahs (*Acinonyx jubatus*), and tigers (*Panthera tigris*) [19, 21, 61]. Several of these studies further demonstrated that footprints can also serve for estimating minimum population sizes [7, 23, 61].

Track plate methods have extended this approach to small mammals and even insects [25, 62, 63], highlighting its potential for broad-scale biodiversity monitoring. Reported cross-validated accuracies often exceed 90%, underscoring the potential of footprint-based approaches as robust, non-invasive tools for conservation [19, 21–23, 25, 48, 52, 61, 64, 65]. In recent years, the integration of automated landmark extraction and **AI**-assisted pattern recognition has further expanded the applicability of footprint morphometrics, enabling large-scale analyses with improved reproducibility.

For otters, multivariate approaches based on **Linear Discriminant Analysis (LDA)** have shown considerable promise. Analyses of more than 600 tracks from 14 trails achieved nearly 100% accuracy [66], and subsequent studies confirmed the feasibility of sex classification and individual identification based on footprint shape [47, 50]. However, as the number of classes increased, the limitations of classical **LDA** became apparent, particularly regarding robustness and generalisability. Vetrovcova [50] explicitly highlighted these issues and pointed to the **FIT** as a potential solution. The **FIT** was first piloted for Eurasian otters in 2016 [51]<sup>1</sup>, marking the initial step towards an otter-specific **FIT** model.

---

<sup>1</sup>This pilot study was conducted as the author's diploma thesis and represented the first application of **FIT** to Eurasian otters.

## 2.2 Footprint Identification Technology (FIT)

Building on these foundations, the following section introduces the *FIT*, the most widely applied and methodologically advanced framework for footprint-based morphometrics, highlighting its conceptual innovations and relevance for addressing the specific challenges of otter monitoring. Originally introduced in 2001 as an objective track identification approach for black rhinoceroses (*Diceros bicornis*) [18], the method was formally established as the Footprint Identification Technique (*FIT*) in 2008 [19]. Subsequent refinements improved its classification framework and accessibility through standardised workflows and a graphical user interface (*Graphical User Interface (GUI)*) [21]. The *FIT* extracts a set of morphometric features from anatomical landmarks on footprints to distinguish between individuals. The approach has been successfully applied across multiple endangered species, including felids and rhinoceroses, and later extended to population size estimation through hierarchical clustering [21, 22]. The detailed methodology—covering image annotation and feature creation (Section 4.2), classification procedures, and validation strategies (Section 5.2)—is described in the respective sections.

Building on this framework, *FIT* models have been successfully developed for a wide range of species, achieving cross-validated accuracies above 90 % and supporting estimates of minimum population sizes that have been independently validated [19, 21, 23, 25, 52, 61, 65]. These results established the *FIT* as the current gold standard in footprint-based morphometrics.

At the same time, recent advancements in the field of footprint analysis—including the integration of track plate data [25, 63], the introduction of prediction confidence measures and *DNA*-based validation [25], and applications of advanced *ML* methods such as gradient boosting (*eXtreme Gradient Boosting (XGBoost)*) [48] and convolutional neural networks (*CNN*) [49, 67]—illustrate that the field continues to evolve.

## 2.3 Current Challenges in the Field of Footprint Monitoring: Particular Focus on FIT

The dissertation focuses on enhancing the *FIT* by integrating methods from different fields such as ecology, statistics, *AI*, and community science, rather than advancing each discipline separately. Although Eurasian otters were used as a model species, the methodological improvements aim to strengthen the *FIT* framework more broadly. The research sought to evaluate and adapt external advances to overcome initial limitations in *FIT*'s scalability, robustness, and interpretability across various species. The dissertation outlines the initial research gaps and the methodological solutions developed to address these challenges.

### Requirement of Advanced Tracking Skills

The success and applicability of the *FIT* strongly depend on accurate tracking skills, making this one of the method's key constraints [65]. The proficiency of a data collector or annotator directly affects the likelihood of locating sufficient high-quality footprints [68]. Field experience during this research showed that refined tracking skills can reveal otherwise overlooked ecological information. For example, solitary otter trails were identified among thousands of dog prints on busy Algarve beaches—sites previously considered unsuitable for otters. Similar observations, such as otter activity on beaches near Barcelona, were reported by other trained trackers during exchanges at an international *IUCN OSG* conference.

#### **Problem: High Tracking Skill Requirements and Risk of Misinterpretation**

When first piloted for otters, available footprint guides—though existing—often provided illustrations that differed from tracks encountered in zoological settings, with incomplete descriptions of inter-foot variation [51]. This made consistent landmark placement challenging and introduced additional uncertainty into early analyses. More recently, improved reference books with high-quality otter footprint illustrations have become available [26, 69, 70], providing a stronger foundation for training, annotation, and practical application within the *FIT* framework.



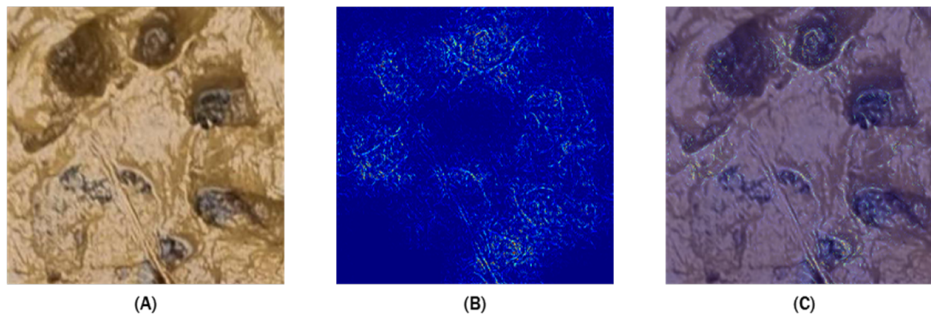
**Fig. 2.2.:** Technical reference drawings of Eurasian otter (*Lutra lutra*) tracks, shown clockwise from the top left: left fore (Left Front (foot), LF), right fore (Right Front (foot), RF), right hind (Right Hind (foot), RH), and left hind (Left Hind (foot), LH). Frequently under-registered forefoot elements—the thumb and the carpal “heel” pad—are rendered with lighter shading. When all five toes register, front tracks are typically broader and appear more symmetrical than hind tracks. If the thumb does not register, hind tracks may appear as the more symmetrical pair and can be misidentified as dog prints. Claws often print and are directly connected to the toe imprint, typically teardrop-shaped; interdigital webbing prints mainly in soft substrates and are best observed between toes 4 and 5. Illustrations by Asaf Ben-David; adapted and modified from [48].

## Proposed Solutions

This dissertation addresses these challenges through two complementary strategies:

First, the scientific community requires more effective training resources. To this end, high-quality schematic footprint diagrams were developed in cooperation with professional trackers, providing anatomically precise and scalable reference material. For example, Figure 2.2 presents detailed illustrations of all four feet of a Eurasian otter and is adapted from the drawings published in [48].

Second, *FIT* pipelines can be adapted to become more transparent and engaging for field practitioners. Trackers frequently ask what the model actually “sees” in a footprint and which measurements carry the greatest weight. This dissertation therefore proposes the development of *Explainable Artificial Intelligence (XAI)* models and visual outputs that highlight essential classification features and distributions within footprints, as illustrated in Figure 2.3 and detailed in [49]. Such outputs can bridge the gap between data-driven classification and traditional interpretive skills, enhancing understanding and trust in *AI*-assisted monitoring.



**Fig. 2.3.:** Model insights supporting explainable FIT applications (adapted from [49]). (A) Right front footprint of an Asian small-clawed otter (*Aonyx cinereus*); (B) visualisation of a saliency heatmap indicating the regions emphasised by a convolutional neural network (CNN) during species classification; (C) overlay of (A) and (B).

Together, these strategies are intended to build trust in *FIT* predictions and to foster dialogue between trackers, conservationists, and data scientists. Beyond technical improvements, the approach emphasises the value of interdisciplinary collaboration, where expertise from ecology, statistics, *AI*, and traditional tracking converge to strengthen both methodology and application. Platforms such as the *WildTrack Specialist Group (WSG)* [71] provide promising arenas for such collaborative exchange and knowledge transfer.

## Data Quality Requirements

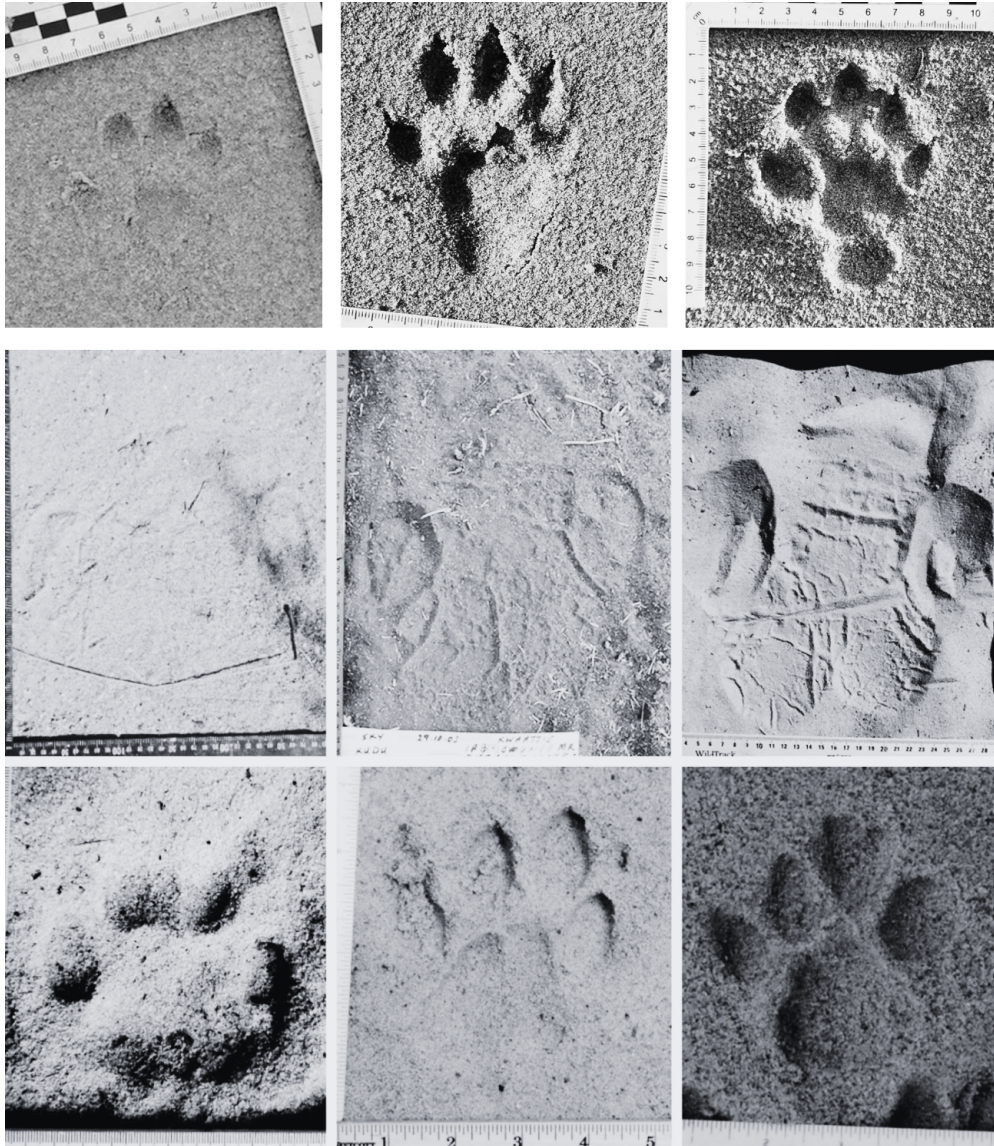
A key challenge for the *FIT* and similar footprint-based monitoring approaches is their reliance on high-quality, continuous trails. For reliable individual identification, six to eight clear footprints from the same foot are typically required [19]. Under field conditions—especially for elusive species such as the Eurasian otter—such trails are often difficult to obtain, as substrate, moisture, gait, and debris frequently reduce print clarity and completeness [69, 72]. Consequently, partial or degraded prints are often excluded, limiting data availability, scalability, and the broader applicability of *FIT*-based monitoring.

### Problem

Adhering to strict quality standards presents a methodological challenge: while excluding flawed data preserves reliability, it may also lead to the loss of valuable ecological information, particularly when high-quality data are scarce. In cheetah monitoring, combining high-quality sub-tracks improved precision, whereas the inclusion of low-quality prints reduced it [72]. In a comparable study, Kistner et al. [68] demonstrated that treating low-quality tracks as missing data and applying imputation techniques such as *MICE* or *PMM* can retain ecological value without compromising model accuracy. Conversely, overly stringent criteria may result in underestimating population sizes [68].

### Proposed Solutions

To address these constraints, three strategies have proven effective: (1) *Refined landmark definitions* that focus on consistently visible anatomical regions while omitting unreliable areas [25]. (2) *Mixed-quality training data*, as demonstrated in this dissertation's *CNN*-based species classification research [49], where expert-labelled degraded prints expanded the applicability of *FIT* without major accuracy loss. (3) *Missing-value imputation* [68], using approaches such as *MissForest* [73] and multivariate imputation by chained equations (*MICE*) [74, 75] to integrate incomplete prints into analyses.



**Fig. 2.4.:** Examples of footprints of varying quality, increasing from left to right. Top row: Eurasian otter (*Lutra lutra*); middle row: white rhinoceros (*Ceratotherium simum*); bottom row: mountain lion (*Puma concolor*).

## Infrastructure Limitations in Terms of Data Transfer, Metadata Capture, and Cloud Storage

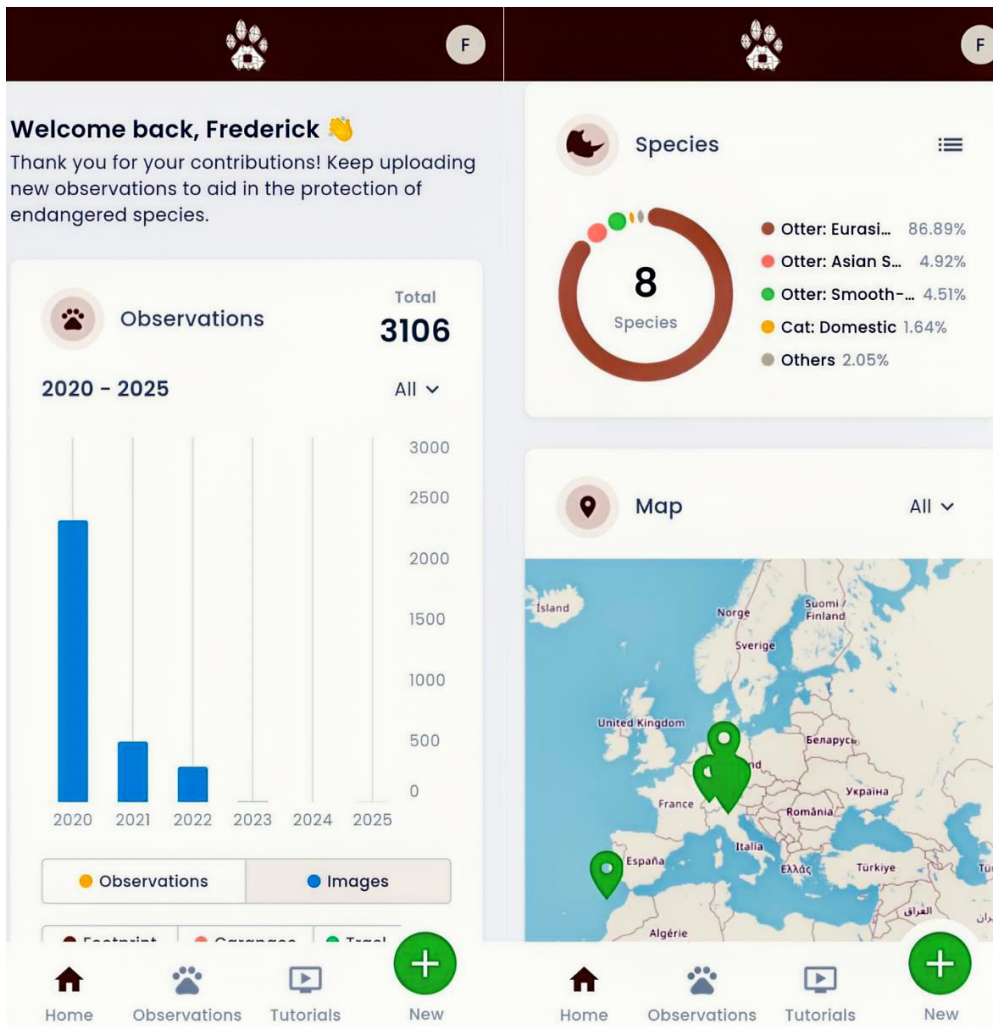
Reliable data transfer and secure storage are essential for maintaining quality and scalability in ecological monitoring. In early *FIT* workflows, however, metadata were recorded manually and images stored locally, increasing the risk of data loss, annotation errors, and barriers to cross-institutional collaboration. Similar issues are common across other ecological monitoring frameworks, where the absence of standardised, cloud-enabled solutions has limited real-time access, *Quality Control (QC)*, and large-scale data integration [76, 77]. The following section outlines *FIT*-specific challenges in metadata capture and data sharing, and proposes a transition towards cloud-based infrastructures.

### **Problem: Lack of Scalable Infrastructure**

Early *FIT* protocols relied on paper-based metadata sheets placed beside footprints before image capture [19, 21], followed by manual transcription into *FIT* data tables. While suitable for small-scale projects, this workflow was inefficient, error-prone, and unsuitable for large or multi-institutional studies. Local storage further increased the risk of unrecoverable data loss and hindered data exchange with external partners. Similar vulnerabilities in other ecological monitoring projects underscore the risks of decentralised, manual data management [25, 61, 68]. At the beginning of this dissertation, no cloud-based infrastructure for large-scale footprint data collection existed.

### **Proposed Solution: Cloud-Based Sharing and Collaborative Annotation**

This dissertation contributed to the transition towards cloud-enabled workflows. Initial trials employed the Epicollect5 platform [76–78], which enabled georeferenced uploads with structured metadata and facilitated cross-institutional data access. The author’s metadata schema—first presented at the *WildTrack* 2020 workshop—was later adopted and refined within the *AI*-driven *WildTrackAI* initiative [14]. The *WildTrackAI* platform now provides standardised input forms, real-time metadata capture, and seamless backend integration for *AI*-assisted *FIT* analyses, addressing the demand for transparent, scalable infrastructures in biodiversity monitoring [24]. By October 2025, the platform had processed 23,077 footprint images from 652 contributors representing 60 species [14], demonstrating the feasibility of cloud-based, community science-driven footprint monitoring.



**Fig. 2.5.:** WildTrackAI mobile application dashboard (2020–2025) showing total footprint uploads by the author, species-level distribution, and geographic coverage. The interface exemplifies the integration of community science with real-time data capture, structured metadata recording, and centralised cloud storage for Footprint Identification Technology (FIT) analyses.

## Artificial Intelligence for Scalable Footprint Monitoring

The rapid expansion of digital wildlife monitoring since the early 2020s has generated increasingly large datasets [24, 79]. While this offers opportunities for more robust ecological insights, it also presents major challenges for data processing [80]. In specialised domains such as footprint-based monitoring, the number of trained experts capable of reliably annotating and analysing large datasets remains limited, creating bottlenecks that constrain scalability and the timely application of monitoring results [81]. At the same time, the effort required to sort, filter, and verify low-quality or incomplete data further increases the workload for human experts [81].

### **Problem: Limited Expert Capacity and High Workload**

The annotation and processing of footprint data require both advanced ecological knowledge and specialised tracking expertise. With the sharp rise in available digital data, a relatively small pool of experts is tasked with ensuring **QC**, annotation consistency, and preparation for **ML** model training. This reliance on scarce expertise leads to delays in analysis pipelines and increases the risk of excluding large volumes of potentially valuable but imperfect data. In many cases, manual pre-screening to filter low-quality prints is itself time-consuming and may still overlook usable information.

### **Proposed Solution: Deep Learning for Scalable Data Processing**

Artificial intelligence (**AI**), particularly deep learning methods, offers promising tools to overcome these scalability barriers. Convolutional neural networks (**CNN**) have already demonstrated strong performance in ecological applications ranging from species classification [82–84] to acoustic monitoring [85, 86]. In the context of footprint research, early **CNN**-based approaches showed that robust classification was achievable even when data quality was heterogeneous, provided that expert-driven annotation ensured reliable labels [49]. Building on these results, this dissertation integrates **FIT**'s morphometric rigour with scalable **CNN** pipelines, lowering thresholds for data inclusion while maintaining analytical reliability. This dual approach expands the applicability of footprint-based monitoring and enables more effective use of community science contributions.

## The Role of Open-Source Tools in Expanding FIT's Accessibility

Open-source tooling is fundamental to participatory science, as it enables transparency, reproducibility, and broad access across institutions. *WildTrack* has long shared *FIT* with conservation partners, and code fragments have been published in several studies [23, 68], ensuring that the method remains accessible rather than proprietary. As ecological datasets continue to grow, open and parallelisable workflows are increasingly needed to enable distributed, auditable, and collaborative analyses across institutions.

### **Problem: Constraints of the Current JMP FIT Add-in**

The core *FIT* implementation currently operates in the *JMP Scripting Language (JSL)* within the proprietary *JMP* statistical environment. While this provides a stable and user-friendly platform, it limits reproducibility, automation, and integration with modern data-science workflows, as *JSL* is less widely adopted than *Python*. Moreover, internal geometric transformations applied during image annotation are not preserved at the pixel level, preventing direct use in deep-learning pipelines [68]. Previous studies have successfully implemented this workflow in *JMP* for a range of species [21, 61], demonstrating its utility but also its dependency on a closed analytical environment. As a result, operational knowledge has become concentrated within a small group of experts, thereby reducing scalability and inhibiting methodological innovation.

### **Solution: Complementary Transition to Open-Source Modular Pipelines**

Building on the established *JMP* foundation rather than replacing it, this dissertation introduces a complementary transition towards an open-source *Python* stack aligned with standardised, FAIR-compliant metadata principles [87]. Using `scikit-learn` and related libraries [88], *FIT* modules were re-implemented as modular, configurable pipelines supporting headless execution, parallelisation, and seamless integration with version control and continuous testing. This design strengthens provenance and reproducibility, streamlines batch processing, and mitigates data leakage that can inflate model performance [89]. Although most morphometric annotations and early analyses were still conducted in *JMP*, recurring automation hurdles motivated this open-source expansion.

## 2.4 Positioning within Citizen Science and Biodiversity Monitoring

Integrating *FIT* into citizen-science and *AI*-based platforms increases its scalability for biodiversity monitoring [24]. Mobile applications such as the *WildTrackAI* app—available for iOS and Android [12, 13]—allow volunteers to collect and upload images directly to cloud systems. These frameworks expand data availability and support inclusive, evidence-based conservation [41].

## 2.5 Chapter Summary

### Methodological Context of Non-Invasive Monitoring and FIT in Otters

This chapter reviewed the current state of otter monitoring, positioning footprint-based approaches within the wider spectrum of non-invasive methods. Conventional techniques such as spraint surveys, camera traps, telemetry, and *DNA*-based analyses were considered in light of the elusive and semi-aquatic behaviour of the Eurasian otter.

The chapter traced the methodological development of footprint monitoring from traditional tracking to digital morphometrics and the *FIT*, highlighting recent *AI*-based advances across taxa. Key challenges in footprint identification were summarised together with emerging strategies to address them. Transitioning *FIT* workflows into modular, open-source pipelines was identified as essential for reproducibility, scalability, and integration within modern ecological data infrastructures.

Embedding these approaches in community-science and *AI*-enabled frameworks offers strong potential for real-time data sharing, stakeholder participation, and evidence-based conservation. While large-scale implementation lies beyond this dissertation's scope, the advances reviewed here provide a foundation for subsequent applications. Building on this groundwork, the next part introduces the chapters on data collection and annotation.



# Part II

---

Primary Data Collection, Field Campaigns,  
and Data Annotation



# Data Collection in Zoos and Field Campaigns

## 3.1 Ethics Statement

### Ethical Compliance and Non-Invasive Data Collection

All data in this study were collected non-invasively, following best-practice guidance for *FIT*-based monitoring in animal enclosures and free-ranging wildlife [21, 52]. Footprints of captive Eurasian otters were obtained with institutional permission and under animal-care supervision; no direct handling, behavioural manipulation, or physical contact occurred at any time. Field data were acquired exclusively by indirect means—i.e., footprint impressions on natural substrates—without direct observation, capture, or interaction with wild animals.

## 3.2 Overview and Objectives

This chapter outlines the creation and analysis of datasets developed to classify *species*, *sex*, and *individual identity* of otters based on their footprints. It details the acquisition of high-quality reference prints from known individuals through collaborations with zoological institutions and experts, as well as the collection of field data across several campaigns. The author undertook extensive tracker training, achieving a CYBERTRACKER Level III *Track & Sign* qualification [70]. The chapter emphasises the variability and challenges in footprint quality arising from environmental conditions and anatomical differences, underlining the need for rigorous methodology and domain expertise. It further presents benchmarking analyses using both existing and proprietary datasets, demonstrating how data quality and refinement enhance the robustness of *FIT*-based monitoring. A summary table of the datasets used concludes the chapter, highlighting their relevance to the overall investigation.

### 3.3 Development of Data Collection Protocols and Tools

Building on previous *FIT* methodologies established for various animal species [21], the following protocol was specifically adapted and formalised for otters in Kistner et al. [48] (see Appendix therein). The protocol components reproduced below follow the original structure and wording of Kistner et al. [48] to ensure methodological fidelity.

#### Standardised Data Collection Protocol for Otters (adapted from Kistner et al. [48])

- **Substrate preparation:** Prepare the enclosure floor to yield clear, unambiguous footprints suitable for measurement and analysis.
- **Individual identification:** Record the correct animal identity for every print—particularly where multiple individuals share an enclosure—to prevent mislabelling.
- **Camera positioning:** Align the camera sensor plane perpendicular to the substrate and position it directly above the print to avoid parallax distortion.
- **Scale reference:** Place an L-shaped measuring scale adjacent to the footprint to enable precise geometric calibration and assess any horizontal or vertical displacement.
- **Close-up imaging:** Capture high-resolution close-ups preserving ridge, claw, and pad details relevant for morphometric analysis.
- **Metadata annotation:** Record essential metadata on a slate within the frame or digitally, following the schema specified in Kistner et al. [48].

## 3.4 Data Collection

This section outlines the data-collection methodology. The first subsection covers data from captive individuals used to train *FIT* models. The second describes field collection under standard monitoring protocols, including *DNA*-referenced validation data. The final subsection summarises campaigns in south-west Portugal focusing on advanced tracking to follow otters *in situ*.

### Collection of Reference Footprints from Captive, Individually Known Otters

Standardised sand-beds were installed in 18 European zoos to collect reference footprints, mainly of the Eurasian otter. Fine builder's sand ensured clear, consistent impressions, and mild attractants promoted regular movement. Continuous observation allowed prints to be linked to known individuals, though not always in social groups. Confirmed-*ID* data were gathered in 2015, 2019, and 2020, supplemented by natural mud prints from Vetrovcova [50]. Additional material from smooth-coated and Asian small-clawed otters was collected in 2021 and labelled at species level.



**Fig. 3.1.:** Standardised sand-beds in zoological settings. **Top:** outdoor placements along fences and transitions; **bottom:** indoor setups near dens and mixed habitats. These enabled controlled, repeatable footprint collection from known individuals.

## Collection of Field Data with a Focus on Bridges Following Natura 2000 Monitoring Protocols

Field campaigns between 2019 and 2023 refined footprint acquisition under natural conditions. Early work focused on bridge underpasses to meet *FFH* monitoring requirements for otters [29]. While many sites provided suitable substrates, some were flooded or showed initial avoidance behaviour, and early efforts were further limited by COVID-19 restrictions. Figure 3.2 shows one of the prepared bridge sites with sand patches designed to enhance track clarity.



**Fig. 3.2.:** Prepared bridge site with sand patches used for the collection of wild otter footprints under controlled field conditions.

A pilot study in Lower Saxony, Germany (2020) paired otter footprints with *DNA*-analysed spraints, providing ground-truth data for subsequent *FIT* validation. From this effort, 42 *DNA*-validated prints from five trails were obtained. Genetic analyses confirmed sex and individual identity, enabling internal consistency testing. The *DNA* analyses were conducted by Aktion Fischotterschutz e. V., which provided the results to the author. Figure 3.3 shows examples from this campaign.



Prepared bridge site with otter tracks prior to DNA sampling.



Documentation of a DNA-referenced footprint before upload and metadata recording.

**Fig. 3.3.:** Examples from the Lower Saxony, Germany, campaign. Despite the limited dataset (42 DNA-validated footprints), the pilot provided key insights that informed a methodological transition towards adaptive trail-following approaches.

While initial efforts focused exclusively on otters, feedback from the 15th International Otter Congress in Sospel (2022) [90] underscored the need to include sympatric species frequently mistaken for otters. Sampling was therefore expanded to encompass several semi-aquatic and terrestrial mammals, including domestic dogs (*Canis lupus familiaris*), badgers (*Meles meles*), stone martens (*Martes foina*), and Egyptian mongooses (*Herpestes ichneumon*).

A dedicated dataset comprising several hundred images per species was compiled for eleven co-occurring species. Although preliminary models for multispecies classification yielded promising results, final quality control was not yet complete at the time of submission and is therefore discussed only in the outlook, with plans for subsequent publication. Figure 3.4 illustrates the challenge of distinguishing otter tracks from those of morphologically similar species.



**Fig. 3.4.:** Representative footprints collected during field campaigns. Row 1: 1. **Eurasian otter** (*Lutra lutra*), 2. stone marten (*Martes foina*), 3. Eurasian badger (*Meles meles*), 4. Egyptian mongoose (*Herpestes ichneumon*). Row 2: 1. Eurasian badger (*Meles meles*), 2. Eurasian beaver (*Castor fiber*), 3. **Eurasian otter** (*Lutra lutra*), 4. muskrat (*Ondatra zibethicus*). Row 3: 1. **Eurasian otter** (*Lutra lutra*), 2. smooth-coated otter (*Lutrogale perspicillata*), 3. Asian small-clawed otter (*Aonyx cinereus*), 4. domestic dog (*Canis lupus familiaris*).

## Field Data Collection and Protocol Refinement in Southwest Portugal

Between 2019 and 2023, field campaigns in south-west Portugal were conducted with the *Conservation and Environmental Research & Education – International Project (field centre, SW Portugal) (CERES)* field centre in Aljezur [91], targeting Eurasian otter footprints in coastal dune and estuarine habitats. Rapid wind erosion and heavy dog traffic complicated track retention, so advanced tracking was used: trailing (continuous following) and speculative tracking (projecting the likely path). Dog–otter discrimination relied on the characteristic 3×4 lope gait [69]. Complementary signs (spraints, rub marks) were recorded, yielding a substantial image archive for subsequent *ML* training and protocol refinement.



(a) Otter trail at sunrise.

(b) Close-up of the same gait pattern.

**Fig. 3.5.:** Examples of the characteristic 3×4 lope gait of the Eurasian otter (*Lutra lutra*), recorded on wet dune sand in south-west Portugal. Panel (a) shows an otter trail across a coastal dune at dawn, illustrating the typical habitat conditions where tracks were documented. Panel (b) provides a close-up of a similar gait sequence with a scale reference. The LF print lacks a visible hypertarsal pad; the LH oversteps the RF, obscuring its first digit; and the RH shows a reduced first toe impression, partially covered by overlying bird tracks. Such incomplete impressions are typical for *Lutra lutra* and can easily be mistaken for those of domestic dogs (*Canis lupus familiaris*) or other mustelids.

On several occasions, individual otter trails were successfully followed for several kilometres, providing valuable insights into intraspecific, interspecific, and individual behaviour. Complementary signs such as *sprints* and rub marks were also documented, offering a broader ecological context and supporting reliable species and individual identification. Figure 3.5 illustrates a typical **3×4 lope gait** sequence captured at sunrise in a coastal dune system, together with a close-up of the same gait pattern. The latter highlights missing or occluded print elements that can cause single impressions to resemble tracks of the domestic dog (*Canis lupus familiaris*) to untrained observers.

Overall, these campaigns refined field tracking protocols and generated an extensive image archive encompassing otter footprints and signs of diverse co-occurring fauna. This dataset forms an essential foundation for training future *ML* models and was complemented by targeted workshops and seminars designed to enhance methodological consistency and practitioner skill development.

### 3.5 Morphometric Benchmark Datasets for Evaluating FIT Pipeline Improvements

The data-collection campaigns laid the empirical foundation for this dissertation, yielding a broad and diverse archive of otter footprints and associated ecological signs across multiple substrates and taxa. While these efforts primarily targeted continuous otter trails, only part of the material is fully annotated and *QC* at this stage. For systematic evaluation and pipeline development, therefore, a structured and comparable dataset design was required.

Table 3.1 provides a detailed breakdown of the Eurasian otter dataset by data origin. These subsets were annotated in *JMP* as described in the following chapter and represent only a small fraction of the total archive. They were selected because they constitute the most directly comparable *FIT* datasets across species and therefore form the core material on which the analytical part of this dissertation is based. For the trails from the Portuguese field data, the focus was placed on those found at locations separated by several average home ranges or identified in the field as originating from different individuals. We evaluated pipeline improvements using seven additional WildTrack benchmark datasets (validated in prior *FIT* studies [19–22, 52, 61]) that share a common structure and standardised annotation (Sec. 4.2); see Table 3.2.

**Tab. 3.1.:** Detailed breakdown of the Eurasian otter (*Lutra lutra*) dataset by data origin.

Data origin	No. of unique trails	No. of unique individuals
Own data collection (captive individuals)	54	26
Vetrovcova [50] (captive individuals) (tripod-mounted, downward-facing camera)	12	11
Fieldprints Lower Saxony	5	4
Fieldprints Portugal	8	—
<b>Total</b>	<b>79</b>	<b>41</b>

**Tab. 3.2.:** Eurasian otters and seven additional species; captive and field sources (incl. *DNA*-referenced trails) support *FIT* morphometrics and downstream *ML* analyses.

Species (Latin name)	No. of known individuals	No. of footprints / trails	No. of landmarks	No. of measurements
Eurasian Otter ( <i>Lutra lutra</i> ) (cf. this study)	41	628 / 79	17	208
Amur Tiger ( <i>Panthera tigris altaica</i> ) (cf. [20, 61])	44	605 / 86	25	128
Bengal Tiger ( <i>Panthera tigris tigris</i> ) (cf. unpublished dataset )	40	586 / 87	25	128
Cheetah ( <i>Acinonyx jubatus</i> ) (cf. [21])	38	781 / 109	25	136
Giant Panda ( <i>Ailuropoda melanoleuca</i> ) (cf. [22])	42	523 / 77	25	124
Mountain Lion ( <i>Puma concolor</i> ) (cf. [52])	35	535 / 77	25	133
White Rhino ( <i>Ceratotherium simum</i> ) (cf. [19])	41	1273 / 159	13	112
<b>Total</b>	<b>299</b>	<b>5131 / 705</b>	-	<b>1086</b>

### 3.6 Additional Datasets from Publications Originating in this Dissertation

Beyond the primary benchmarks, three complementary datasets were developed in peer-reviewed studies arising from this dissertation. In Kistner et al. [48], approximately 100 footprint impressions per species—Eurasian otter, Asian small-clawed otter, and smooth-coated otter—were collected at eight European zoological institutions and analysed using landmark-based morphometrics within the *FIT* framework.



**Fig. 3.6.:** Representative footprints from Kistner et al. [48]: Eurasian otter, Asian small-clawed otter, and smooth-coated otter. Images were used for landmark-based morphometric analysis in the *FIT* context.

Kistner et al. [49] expanded this repository to 2562 high-resolution images spanning a wide range of capture qualities, including verified field prints from Portugal and Malaysia. Each image was annotated with bounding boxes to train and validate a *CNN* for footprint-based *species* classification within the *FIT* framework.

Kistner et al. [68] then extended this work by augmenting six of the benchmark datasets summarised in Table 3.2 with controlled artificial missingness to systematically assess the robustness of the *FIT* pipeline. In addition, a newly compiled cheetah (*Acinonyx jubatus*) dataset—characterised by naturally missing landmarks and variable print quality—was analysed to evaluate the performance of imputation strategies under real-world data incompleteness, full details are provided in the publication.

## 3.7 Chapter Summary

### Data Collection and FIT Dataset Generation in Otter Research

This chapter documents the development of a comprehensive, multimodal dataset for *FIT* analyses at the *species*-, *sex*- and *individual identification*-level. All material was collected non-invasively from captive and free-ranging contexts to maximise ecological validity and ensure ethical compliance.

**Captive reference data.** In collaboration with European zoological institutions, standardised sand-bed installations yielded high-resolution reference prints from individually known otters. Protocols covered substrate preparation, controlled imaging with defined *Ground Sampling Distance (GSD)*, placement of metric scale references, and consistent metadata under *Quality Assurance / Quality Control (QA/QC)* guidelines. These data provide the core training material for morphometric *FIT* models.

**Field campaigns in Germany.** Bridge-based surveys aligned with *FFH*-compliant otter monitoring produced *DNA*-referenced trails via paired footprint–*spraint* sampling, enabling internal validation of *FIT* outputs under routine monitoring conditions.

**Field campaigns in south-west Portugal.** Working with the *CERES* field centre, advanced trailing and speculative tracking were applied in coastal dune and estuarine systems to collect continuous gait sequences and complementary signs. These campaigns refined field protocols across diverse substrates and disturbance regimes and expanded the archive with behaviourally contextualised trails.

**Benchmarking and readiness.** Complementary footprints from sympatric species commonly confused with the Eurasian otter were added for specificity checks. Together, the captive and field components form a harmonised, reproducible basis for subsequent morphometric workflows and *AI/ML*-assisted analyses presented in later chapters.



## Methods A: Data Annotation and Exploration

### 4.1 Generation of Reference Track Drawings

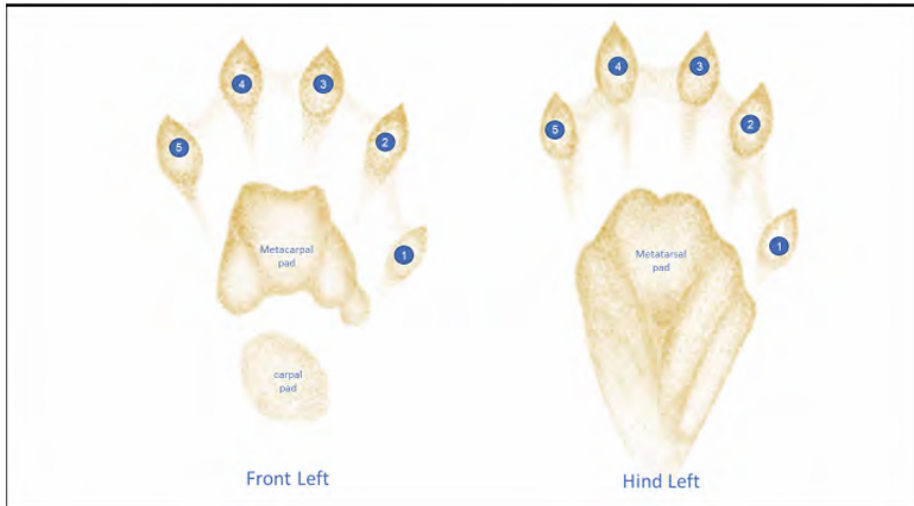
To support anatomically consistent landmark placement and improve the interpretability of species-classification models, technical reference drawings were produced for all three otter species analysed in this study: the Eurasian otter, the Asian small-clawed otter, and the smooth-coated otter . These drawings build on prior work and were adapted and refined specifically to guide annotation workflows used in both morphometric analyses and *CNN*-based image modelling.

Beyond providing guidance for consistent landmark placement, the reference drawings also established a standardised visual framework for footprint nomenclature. This facilitated clear communication across annotators and ensured comparability between different analytical environments, including both the morphometric *FIT* workflow and image-based pipelines.

The illustrations were produced by a trained wildlife tracker and scientific illustrator, based on high-resolution photographs of footprints collected during controlled zoo studies and images of the underside of otter feet. These reference drawings represent idealised impressions of the left front (*Left Front (foot) (LF)*) and left hind (*Left Hind (foot) (LH)*) feet.

These visual templates fulfilled three key roles:

- They illustrated the scientific nomenclature of the footprint and its originating foot.
- They supported standardised landmark placement.
- They enabled visual comparison of field-collected footprints with idealised anatomical models, helping to assess track quality, completeness, and suitability for inclusion in training data sets.



**Fig. 4.1.:** Anatomical **labelling** of otter footprints (left front and left hind). Numbering of toes and labelling of anatomical structures. Adapted from Kistner et al. [48]. Used for **standardising** landmark placement across species and annotation environments, and to support the morphometric **FIT** workflow.

## 4.2 Annotation

The digitisation process converted footprint images into structured datasets via a standardised **FIT** annotation workflow. Photographs were preprocessed by resizing, rotation to a common orientation, and scaling to a reference ruler to ensure consistent measurement. Analyses used left-front (**LF**) footprints. To increase sample size for the field **DNA** validation, right-front (**Right Front (foot) (RFF)**) footprints were horizontally mirrored. For geometric alignment, two reference points were placed at the lower edges of toes 2 and 4. Annotation comprised 11 anatomical landmarks capturing footprint morphology. Landmarks were placed manually in **JMP** using a custom module with cross-fade support. Each footprint was annotated three times; the median landmark configuration across replicates was retained for analysis.

In Kistner et al. [49], footprints were annotated with bounding boxes; that image-based procedure is not detailed here to maintain a morphometrics focus.

## 4.3 Feature Generation

An automated *JSL* script generated six derived landmarks, yielding 17 points per footprint. From these, 208 geometric variables (linear distances, angles, triangle areas) were computed for downstream modelling; full variable definitions are listed in Kistner et al. [48]. Later versions of the *FIT* tool added landmark-uncertainty flags and fold-safe missing-value handling [68]; these metadata were not available for the present dataset.



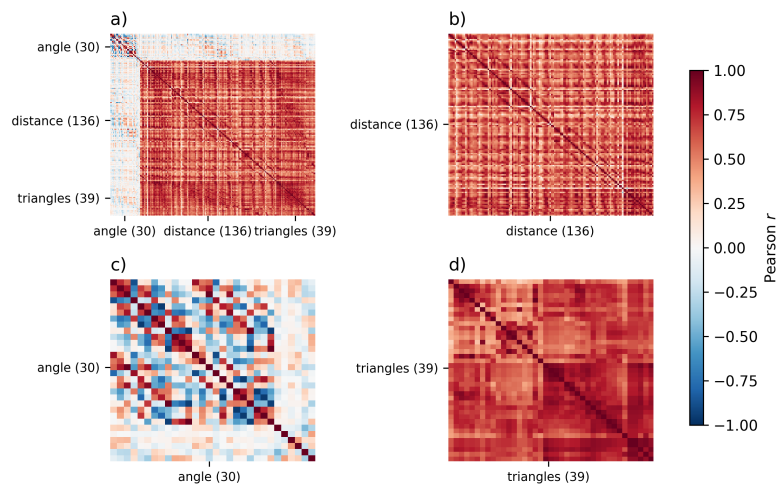
**Fig. 4.2.:** Geometric profile from *FIT* annotation in *JMP*: landmark layout with example distances, angles, and triangle areas used as features. Variable definitions follow Kistner et al. [48].

## 4.4 Data Exploration

This section presents exploratory data analyses for the morphometric *FIT* dataset of otter footprints, with a focus on the target variables *sex* and *individual id*. Comparable automated routines were applied to the other benchmark species but are not described further here for brevity. From this point onward, references to functions in the accompanying GitHub repository<sup>1</sup> are provided for transparency and reproducibility.

### Feature Interdependencies

To assess relationships among variables, correlation heatmaps were generated using the function `plot_feature_correlations` (Figure 4.3). The results show a strong positive correlation among distance- and triangle-based features, reflecting their shared scaling with overall footprint size (see Figure 4.2 for landmark definitions). By contrast, angular variables exhibited a more heterogeneous pattern, including both strong positive and negative associations as well as weaker dependencies. This suggests that angle-based features capture more subtle aspects of footprint shape beyond overall size. The pronounced multicollinearity observed in these measurements was an important consideration for the feature-selection strategies described in the following chapter.



**Fig. 4.3.:** Correlation heatmaps of morphometric variables grouped into distances (cm), angles (degrees), and triangle areas (cm<sup>2</sup>). Distances and triangles are strongly positively correlated, whereas angles show both positive and negative associations.

<sup>1</sup>[https://github.com/FKistner85/FIT\\_package/tree/A\\_new\\_development\\_branch](https://github.com/FKistner85/FIT_package/tree/A_new_development_branch)

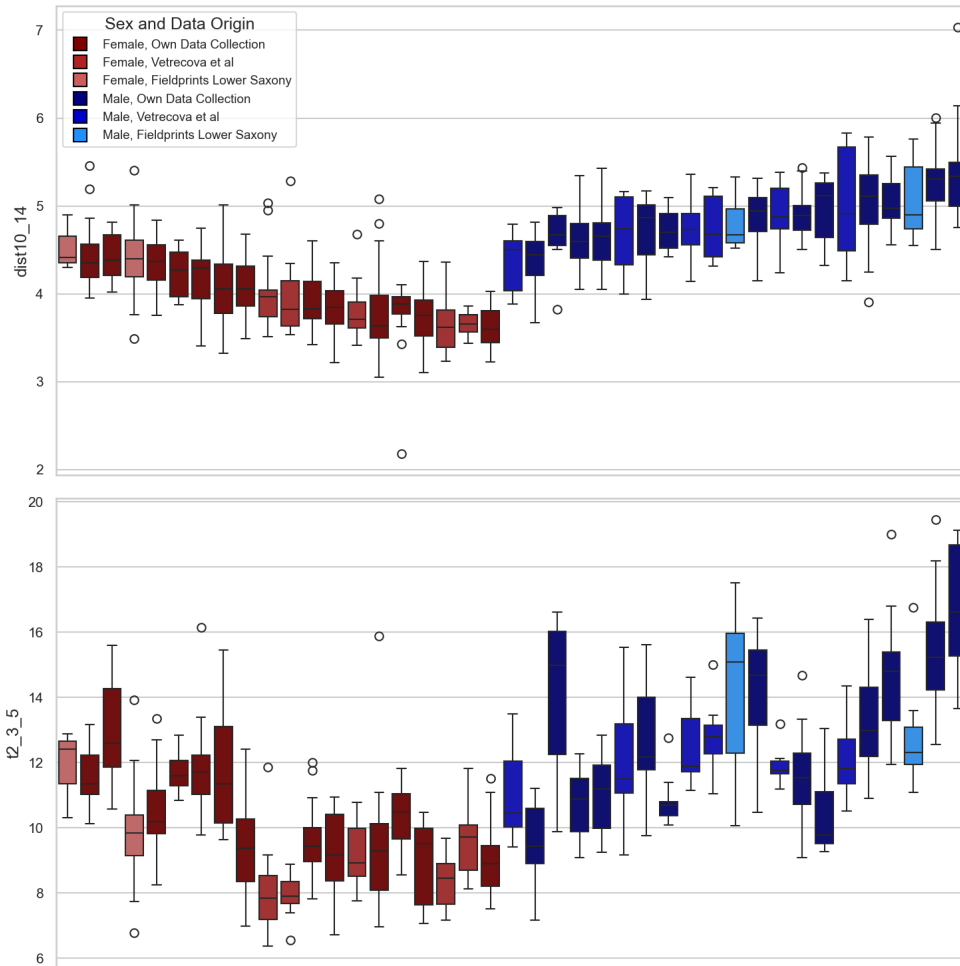
## Discriminative Features.

To identify the most informative features for each classification task, we applied an *Analysis of Variance (ANOVA)* F-test using the function `select_top_features`, implemented via `sklearn.feature_selection.SelectKBest`. Feature importance was evaluated separately for *sex* and *individual ID* classification, allowing a targeted variable subset for each objective.

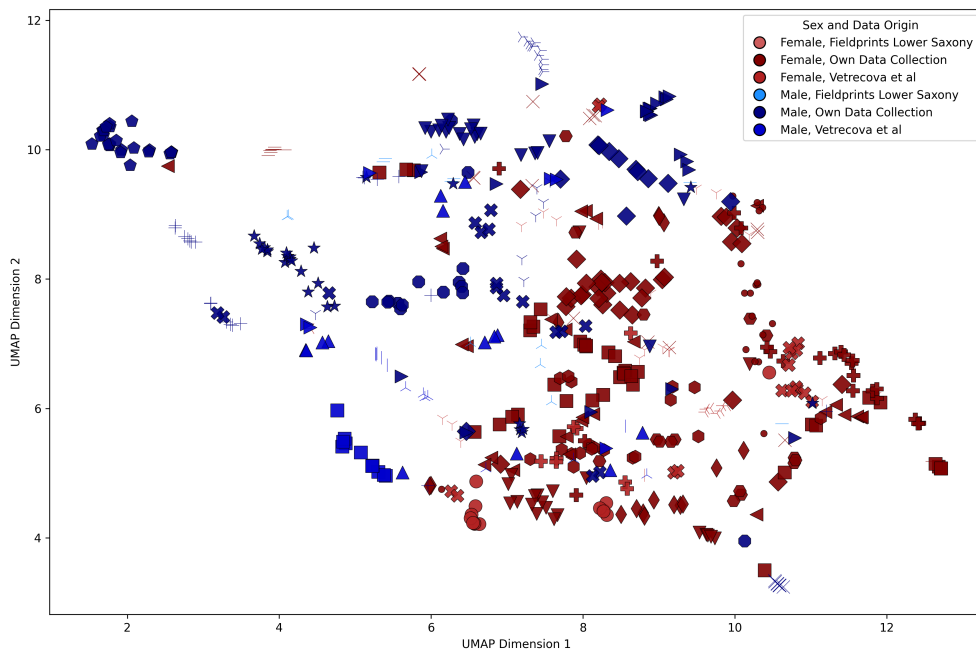
For sex classification, the most informative feature was the inter-landmark distance between points 10 and 14. Males (blue) showed higher values than females (red), though classes still overlapped. For individual identification, the top feature was the triangle area defined by landmarks 2, 3, and 5. This shape-based descriptor captures angular and positional structure and showed broader value ranges in males. The distributions of both variables are shown in Figure 4.4.

## Low-Dimensional Embeddings.

To visualise clustering structure, a two-dimensional supervised *Uniform Manifold Approximation and Projection (UMAP)* embedding was computed on the training data and then applied to the test set. *UMAP* [92] was chosen over *Principal Component Analysis (PCA)* because it better preserves both local and global structure in high-dimensional, non-linear data and is effective at revealing latent clusters that do not necessarily align with axes of maximum variance. The resulting projection is shown in Figure 4.5. Points are coloured by *sex* (red = female, blue = male), with shading indicating data origin (Own Data Collection, Vetrovcova et al. 2006, Field-prints Lower Saxony). Individual otters are distinguished by unique marker symbols, while the legend shows only sex and data-origin combinations for readability.



**Fig. 4.4.:** Boxplots of the most discriminative features for *sex* und *individual ID* (bottom). Distances are in centimetres, triangle areas in  $\text{cm}^2$ . Colours indicate sex and data origin. Although males tend to exhibit higher values, substantial intra- and inter-class overlap remains.



**Fig. 4.5.:** Two-dimensional supervised UMAP embedding of Eurasian otter footprints. Points are coloured by sex (red = female, blue = male), with shading indicating data origin (Own Data Collection, Vetrecova et al. 2006, Fieldprints Lower Saxony). Unique marker symbols denote individual otters; the legend displays only sex–origin combinations for readability.

## 4.5 Chapter Summary

### Data Preparation and Exploratory Overview

This chapter described the methodological foundation for preparing *FIT* datasets. First, technical reference drawings were developed for all focal otter species and used to define consistent anatomical nomenclature and landmark placement. These drawings represent idealised impressions of the left front (*LF*) and left hind (*LH*) feet and were designed to support reproducible annotation across both morphometric and image-based workflows.

Second, a standardised annotation procedure in *JMP* was outlined. High-quality footprint images were preprocessed (rotation, scaling to a reference ruler, and alignment) before manual landmarking. Each footprint was annotated at 11 primary anatomical landmarks, placed mainly at structurally stable regions of the print, and subsequently re-annotated to obtain three replicates. The median configuration was retained to reduce operator-related noise. An automated *JSL* routine then generated six additional derived landmarks, resulting in 17 landmarks per footprint. From these, 208 geometric measurements (distances, angles, and triangle areas) were computed for downstream modelling. Later extensions of the *FIT* workflow introduced landmark uncertainty flags and missing-value handling, although these metadata were not yet available for the otter dataset.

Finally, exploratory analyses were presented to characterise the resulting dataset. Correlation heatmaps revealed strong collinearity among distance- and area-based measurements, whereas angular measurements showed more heterogeneous relationships. Discriminative feature screening using an *ANOVA* F-test (*SelectKBest*) highlighted key variables for *sex* and *individual ID*. In addition, a supervised *UMAP* embedding was used to visualise the data structure: samples were displayed by *sex*, data origin, and individual identity. Together, these steps establish the core data representation on which subsequent modelling and evaluation are based.

# Part III

---

Analytical Workflow and Results



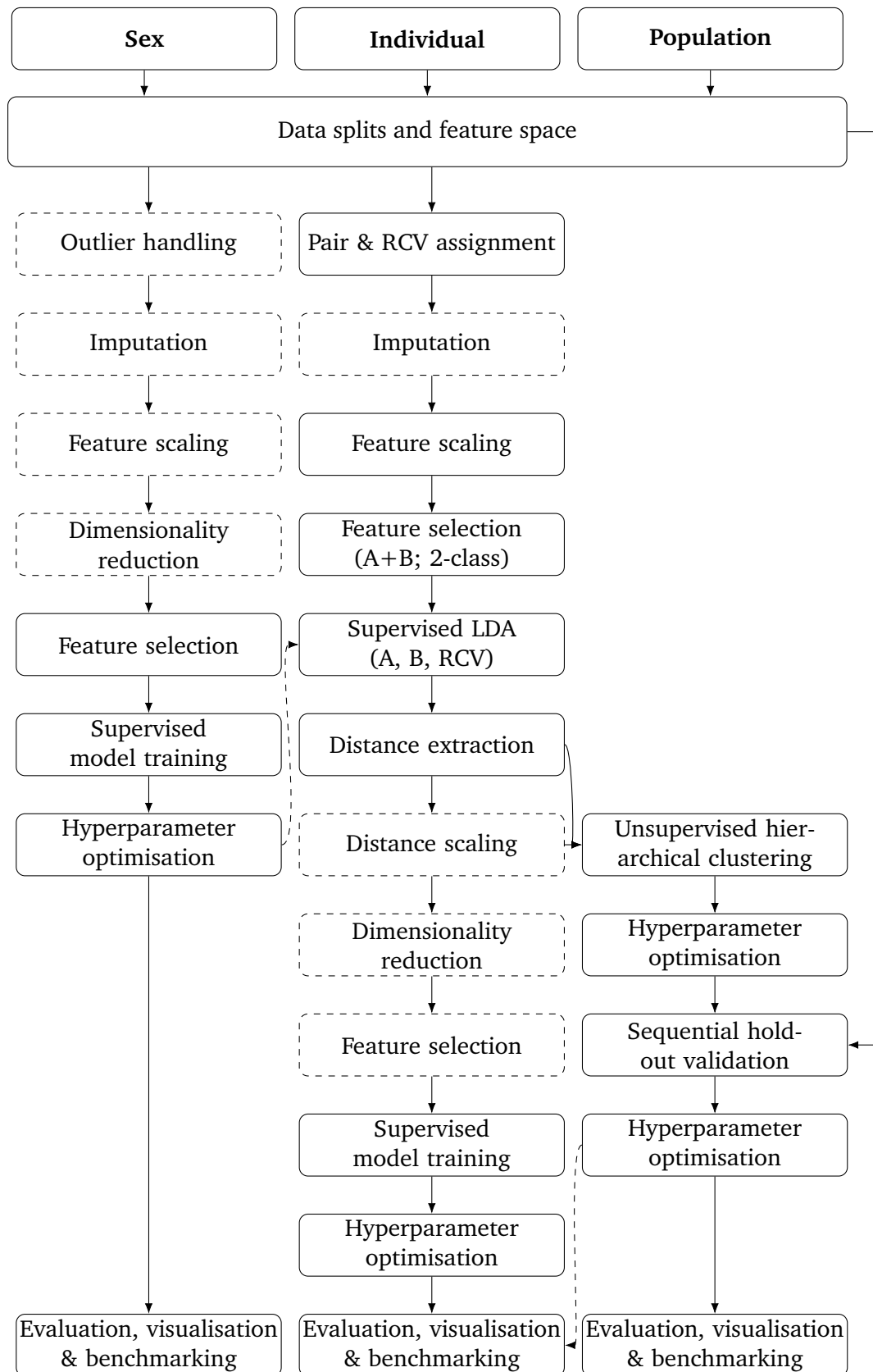
## Methods B: Analytics Pipeline and Experimental Design

### 5.1 Original Pairwise Method in JMP for FIT

The current pairwise protocol implemented in *JMP* originates from the updated *FIT* package introduced by Jewell et al. [21]. This protocol was later applied in our missing-values study [68]. *Individual identification* follows a strict *pairwise* procedure: for each trail pair ( $A, B$ ), a cross-validated *LDA* model with forward stepwise feature selection (based on the F-ratio) is trained jointly on  $A, B$ , and a reference centroid class (**Reference Centroid Value (RCV)**) [18, 19]. Trails are projected into canonical space, centroids with 95 % circular confidence ellipses are plotted, and Euclidean distances between centroids are extracted [22]. The resulting distance matrix is then used in Ward's hierarchical clustering to infer population structure, supported by graphical tools for visual interpretation of cluster patterns [21].

### 5.2 Modular Modelling Framework for Classification and Population Estimation

This dissertation retains the pairwise analytical backbone (Section 5.1) but re-implements it within a modular *Python* environment to enable automation and scalability in model development. **Key innovations** include **group-aware, stratified data splits; integrated imputation and alternative feature selection;** the *definition of cluster-evaluation metrics and automated optimal cut-off selection;* and **staged modelling workflows.** These extensions enhance reproducibility and flexibility while preserving the interpretability of the legacy *FIT* framework. Figure 5.1 summarises the workflow.



**Fig. 5.1.:** Pipelines for the three targets—*Sex* (left), *Individual* (middle), and *Population* (right)—constructed as stacked matrices: a header row, shared data preparation (grouped/stratified; grouped CV), and a subsequent  $3 \times N$  block of steps. Optional steps are shown with dashed borders.

## Data Splits and Feature Space

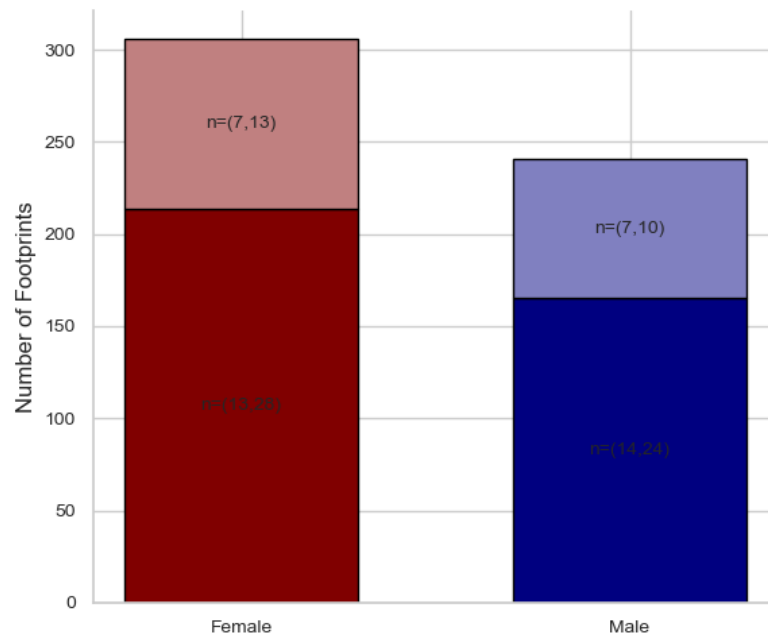
All dataset partitions used fixed random seeds, with a single global seed in the *FIT* package applied consistently across all analyses to ensure full reproducibility. Figure 5.2 summarises the Eurasian Otter dataset composition used in this study, including both label balance and relative proportions for the training, validation, and test sets. All *DNA*-confirmed wild trails were allocated to the test set to enable realistic evaluation on true field data, alongside randomly sampled prints from captive individuals.

### Data Splits

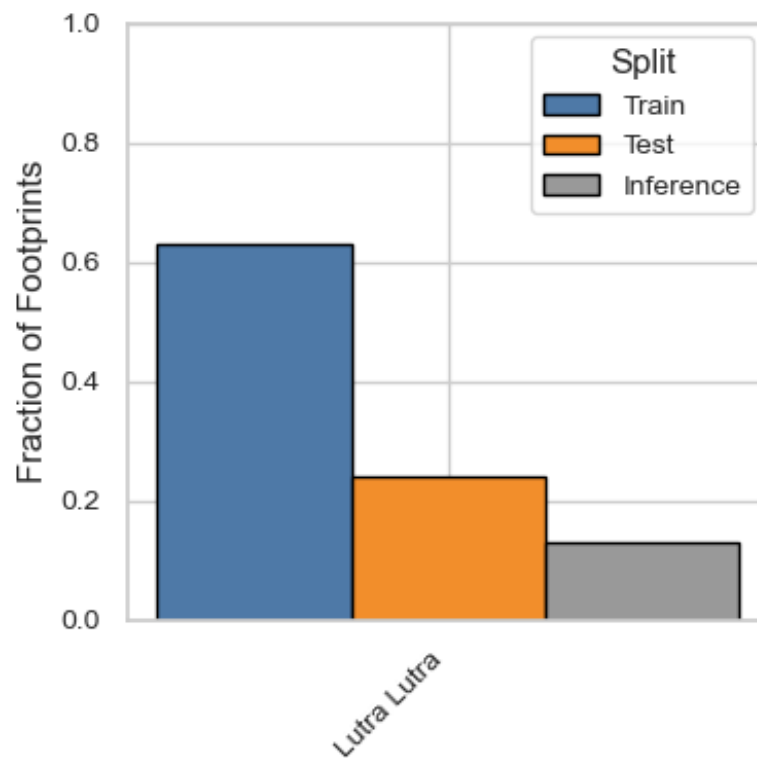
1. **Morphometric species classification** (after [48]): 300 footprints (100 per species) were split 75 % / 25 % into training and test sets; the training split was further divided into five *Cross-Validation (CV)* folds.
2. **Sex and individual-ID benchmarks with group-aware, stratified data splits and grouped CV**: 70 % / 30 % training/test, grouped by individual *ID* and stratified by sex. For Eurasian otters, splits were additionally balanced by substrate (mud/sand).
3. **Population estimation** (sequential hold-out; after [22]): For each hold-out size, validation sets were sampled without replacement (10 iterations per size). Within each, all held-out individuals were compared pairwise, the remainder formed the *RCV*, and metrics were aggregated across iterations.

### Feature Space

1. **Morphometric features**: see morphometric feature definitions in 4.3.
2. **Sex-model predictions**: see optional stacked sex probabilities.
3. **Pairwise distances (A–B)**: see pairwise protocol and distance extraction.



(a) Sex distribution across dataset splits



(b) Relative fraction of footprints per split

**Fig. 5.2.:** Summary of the Eurasian otter (*Lutra lutra*) dataset composition used in this study. **(a)** Distribution of footprint samples by sex across the Train/Test splits. Test set: 7 females (13 trails) and 7 males (10 trails); overall: 20 females (41 trails) and 21 males (34 trails). The slight female over-representation was not adjusted for downstream analyses. **(b)** Relative fraction of footprints per split, highlighting the dominance of training data with smaller test and inference partitions.

## Pipeline steps following data splits

The following subsections describe the pipeline steps, with an emphasis on unpublished modifications relative to prior publications. Unless stated otherwise, implementations rely primarily on functions from *Python*'s `scikit-learn` package [88] and follow standard *ML* practices within the *CV* framework.

### Pair & RCV Assignment

Pairwise comparisons are constructed for all trail pairs within each *CV* fold. Duplicates—where *Trail A* and *Trail B* appear in reversed order (*Trail B* and *Trail A*)—are removed to ensure that each unique combination is represented only once. Each resulting pair is subsequently processed through the same pipeline steps, and all pairs are generated exclusively from the validation splits.

The *RCV* set is always derived from the training data and represents a copy of that set in which all grouping variables (individual *ID* and trail) are assigned the label *RCV* [19]. Unlike the legacy *JMP* implementation, which constructs the *RCV* from the complete dataset (excluding only the focal pair) [21], our approach enforces a novel and stricter separation: *RCVs* are generated *solely from the training set*, while comparison pairs are drawn *exclusively from the current validation fold*.

### Outlier handling

In [48, 68], no explicit outlier handling was applied. In the present study, the *sex* model exposes 99<sup>th</sup>-percentile clipping as an optional hyperparameter within the search grid. For *individual identification* analyses, the default remains `None` to prevent the introduction of additional variance or bias, particularly given the heterogeneity in trail sizes. The step is implemented within the pipeline to facilitate future experiments and ablation studies.

## Missing-value handling and imputation

As a core dissertation publication, [68] benchmarked imputation strategies across seven species. Tree-based imputers (e.g. `missForest` and `missRanger`) achieved the highest overall performance; a concise methodological overview is provided in [68]. The experimental design of the missing-value study is summarised at the end of this chapter, and key outcomes are revisited in the *Results*.

For the *sex* and *individual identification* benchmarks of the new pipeline, all morphometric inputs were complete; accordingly, the imputation hyperparameter was fixed to `None`. If missing values are present in future datasets, a tree-based imputer is recommended as the default option, following [68].

During pairwise distance extraction, a small subset of pairs yielded non-computable distances (e.g. due to insufficient prints or ill-conditioned covariance matrices). Given the rarity of such cases, we excluded the affected pairs from downstream *individual identification* and clustering analyses rather than applying distance imputation, thereby avoiding unnecessary complexity in the evaluation of the new pipeline modifications.

## Feature scaling

In [48, 68], feature scaling was not applied, as it has not typically been reported as an explicit preprocessing step in prior *FIT* publications [22, 61]. In the present study, scaling is treated as a tunable component and evaluated at several stages of the analytical pipeline.

For the *sex* model, `StandardScaler` is included as an optional hyperparameter in the search space. In our implementation of the pairwise analyses, feature scaling with the same scaler serves as the default step to standardise inputs for *LDA*. For distance features extracted from the pairwise models (see below), `MinMaxScaler` is used, with the inclusion of scaling exposed as a parameter in the search grid.

In all cases, scaling is fitted *within* each training fold and, at test time, refitted on the full training partition before being applied to the held-out set.

## Feature selection and feature importance

Feature selection is an essential component for improving the efficiency and accuracy of *FIT*. Performance is evaluated by classification accuracy and wall-clock runtime per pair and per *CV* fold. At the species level, Kistner et al. [48] applied stepwise forward *LDA*. In pairwise *FIT* analyses, a separate selector is trained for each trail pair to identify discriminative variables, with previous studies typically selecting  $k \approx 12-18$  features for *individual identification* [19, 21–23, 52].

**Selector choices.** We employ interpretable selectors designed to handle multicollinearity: (i) stepwise forward *LDA* replicating the *FIT* implementation in *JMP* (custom implementation), (ii) **Random Forest (model context) (RF)**-based feature ranking with top- $k$  selection, and (iii)  $\ell_1$ -regularised **Least Absolute Shrinkage and Selection Operator (LASSO)** with cross-validated penalty optimisation. *RFs* compute feature importance from impurity-based reductions [93], while the *LASSO* applies coefficient regularisation to address multicollinearity in high-dimensional feature spaces [94]. Aggregated rankings across folds are combined to obtain consolidated feature-importance scores.

### Optional sex information (stacking)

After tuning and fitting the *sex* classifier within each outer *CV* split, per-trail sex probabilities (out-of-fold for training folds; full-model for test and inference) are concatenated with selected morphometric variables immediately before the pairwise *LDA* or embedding step. This stacking step adds auxiliary probability features while maintaining independence between training and validation data [89, 95].

## Dimensionality reduction

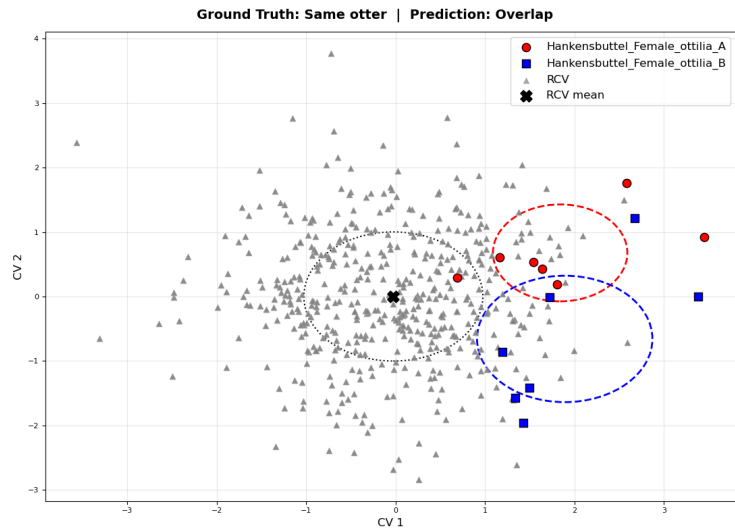
**PCA in the sex classification pipeline.** Principal Component Analysis (*PCA*) is optionally applied as a preprocessing step to scaled inputs. In ablation settings, *PCA*—reducing to  $d$  components tuned via grid search—substitutes forward feature selection. *PCA* is fitted within each *CV* fold using only the training data and, during final evaluation, refitted on the complete training split. Transformation of validation and test data is subsequently performed with the fitted scaler to prevent data leakage during dimensionality reduction [89].

**Supervised embedding with *RCV* (*CV* and test).** Within each *CV* fold, all unique unordered trail pairs ( $A, B$ ) are enumerated. For each pair, the (*RCV*) is defined as the complement of  $A \cup B$  within the training subset. A supervised two-dimensional *LDA* is then fitted on the three groups  $\{A, B, RCV\}$ , with the *RCV* providing a fixed centroid for the canonical axes to ensure consistent geometry across pairwise comparisons and avoid data leakage.

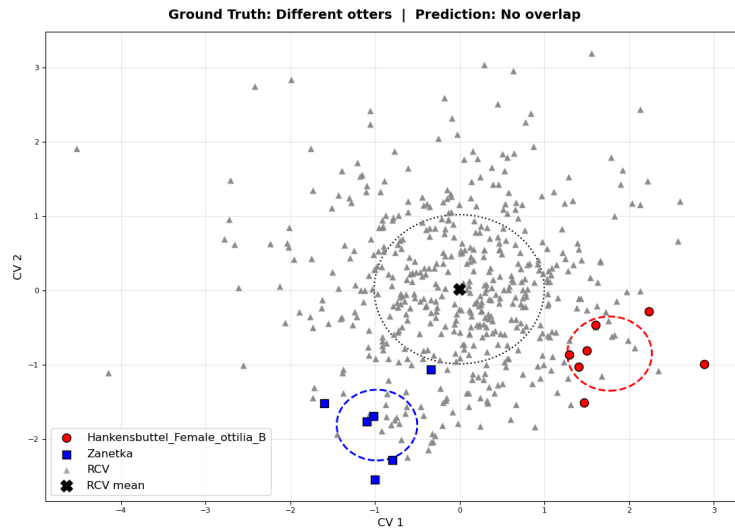
At test time (and during inference), all held-out pairs are evaluated analogously but reference a *single RCV* computed from the complete training set. This common reference frame is reused across pairs to maintain alignment of canonical axes. Figure 5.3 provides an example of this configuration, where each trail mean ( $\mu$ ) is represented by a multivariate Normal 95% confidence circle, and the relative spacing between circles corresponds to the computed pairwise distances between trails.

## Distance extraction

In line with *FIT*, trail geometry in canonical space is summarised by the *Euclidean centroid distance* between trail means. Historically, decisions on *same* versus *different* individuals were based on the overlap of their 95% multivariate Normal confidence circles [19, 21], as shown in Figure 5.3. In this study, the *JMP* overlap calculations are not reproduced. Instead, Euclidean centroid distances are computed in the same canonical space, together with auxiliary metrics (Manhattan, Chebyshev, cosine, Canberra, and Bray–Curtis) and descriptive summaries (mean and median between-trail distances, within-trail dispersions). These distance measures serve as input features for downstream supervised modelling [96].



(a) Example 1: Overlapping 95 % multivariate Normal confidence circles from trails of the same individual



(b) Example 2: Non-overlapping 95 % multivariate Normal confidence circles from trails of two different individuals

**Fig. 5.3.:** Illustration of the historical Footprint Identification Technology decision rule based on 95% multivariate Normal confidence circles (equal radius in all directions, centred on trail means  $\mu$ ) in canonical space [19, 21]. Each plot shows two trails ( $A, B$ ) and the Reference Centroid Value group used to anchor canonical axes across comparisons. In the legacy JMP statistical software package implementation, overlap indicated “same” and non-overlap “different”. In this study, only distances (Euclidean and auxiliary metrics) are computed in the same Reference Centroid Value-anchored canonical space and used for supervised modelling. This visualisation demonstrates the core principle of the stepwise pairwise analyses of Footprint Identification Technology. Aspect ratios and relative scales are preserved.

## Supervised classification models (sex, species, individual)

- **Sex and species:** Linear Discriminant Analysis (*LDA*) serves as the baseline classifier, with *XGBoost* included as a tree-based alternative. *LDA* has been widely used in footprint-based studies [25, 47, 48, 63], w*XGBoost* represent a state-of-the-art approach for tabular data [97]. Hyperparameters (e.g. *LDA* solver and shrinkage; *XGBoost* tree depth, learning rate, subsampling ratio, and column sampling) are tuned via predefined grid searches [88].
- **Individual identification (same/different):** Instead of reproducing the *JMP* overlap logic, a discriminant classifier is trained on the distance-feature vector obtained from the pairwise stage (Euclidean and auxiliary metrics). An *LDA* classifier is used, with compact grids covering solver, shrinkage, and, in reduced pipelines, feature-selection or *PCA* settings.

## Unsupervised hierarchical clustering

Population size is estimated using Ward's hierarchical clustering applied to the trail-by-trail distance matrix  $D$  (one matrix per species, pipeline, and split), consistent with previous *FIT* studies [7, 22, 49]. Distances derived from the pairwise stage are represented with  $D_{ii}=0$ , and non-computable pairs are excluded. We compare raw distances against min-max-scaled values in the range  $[0, 1]$  and evaluate scalar cut-off thresholds  $t$  for each distance type to determine cluster partitions.

## Hyperparameter optimisation

Automated optimisation was conducted using (i) *Bayesian optimisation* [98] and (ii) *grid search* [88]. All optimisation routines were nested within outer *CV* splits to prevent information leakage.

For *species* models [48], tuning parameters from the original *JMP* implementation were retained. *Sex* models used Bayesian searches covering preprocessing, feature-selection, dimensionality-reduction, and classifier parameters. For *individual identification*, grid searches were applied to `SelectKBest+LDA` and `PCA+LDA` pipelines. *Population estimation* employed grid searches across distance metric, optional min-max scaling, and dendrogram cut-off thresholds (Ward linkage fixed).

## Evaluation and benchmarking

Model performance was evaluated following Kistner et al. [68] using two primary metrics: **BA** for classification and **DER** for population estimation. For binary classification, **BA** equals the mean of the **true positive rate (TPR)** and **true negative rate (TNR)**; for  $K$  classes, it corresponds to macro-averaged recall. The term “BCR” used by Kistner et al. [68] is equivalent to **BA**. For clustering, homogeneity ( $H$ ), completeness ( $C$ ), and their harmonic mean—the **V-measure**—were also reported [99]. Cluster cut-offs were optimised by maximising the combined selection score  $S_\alpha$ , balancing count fidelity and label agreement. Ward dendrograms for representative folds and the test set display this final cut-off with annotations for **DER**, homogeneity, completeness, and the **V-measure**. Key evaluation formulas are summarised in Box 5.2, and confidence categories for classification outputs are defined in Box 5.2.

### Evaluation metrics

*Balanced accuracy (binary):*

$$BA = \frac{1}{2} (TPR + TNR)$$

*Balanced accuracy (multi-class):*

$$BA = \frac{1}{K} \sum_{k=1}^K \text{recall}_k$$

*Dendrogram error rate (cluster-count error):*

$$DER = \frac{|\widehat{K} - K_{\text{ref}}|}{K_{\text{ref}}}$$

*V-measure (harmonic mean of homogeneity and completeness):*

$$V = \frac{2HC}{H + C}$$

*Combined selection score:*

$$S_\alpha = \alpha (1 - DER) + (1 - \alpha) V, \quad \alpha = 0.5$$

### Confidence categories

- **High confidence:** confidence score  $> 0.9$
- **Moderate confidence:** confidence score  $> 0.7$  and  $\leq 0.9$
- **Low confidence:** confidence score  $\leq 0.7$

Confidence categories were defined from model probability outputs of the *sex* classifier and applied both to single-print predictions and to majority-class predictions when grouping by trail or individual identity. Additional metrics (accuracy, log loss, precision, recall, F1, and **Receiver Operating Characteristic – Area Under the Curve (ROC–AUC)**) were logged using standard `scikit-learn` implementations [88].

## Validation and Inference on Field Prints

Final *ML* models were applied to *trails* collected during the Portugal field campaign and the *DNA*-validated survey in Lower Saxony, Germany. For the latter, known identities were included in the test pool but are reported here primarily to demonstrate real-world applicability under field conditions. Predictions from the *sex classification* model were assessed for consistency across *trails* belonging to the same individual, providing an additional **QC** layer for classifier reliability. Results from *individual identification* were interpreted in relation to *trail* location and origin, allowing for the assessment of plausible matches given spatial distribution patterns.

## 5.3 Evaluated Methodological Improvements to Morphometric FIT

This section summarises methodological extensions evaluated relative to the classical *FIT* procedure implemented in *JMP*. The analytical framework was re-engineered in *Python* to enhance automation, reproducibility, and modularity while preserving conceptual comparability with the original pairwise design. The evaluated extensions comprise three components: pipeline modifications, missing-value imputation, and complementary deep-learning developments.

The updated pipeline introduces group-aware, stratified *CV* splits with fixed random seeds; strict separation of training, validation, and *Out-of-Fold (validation setting) (OOF)* partitions to prevent leakage; and optional modules for outlier handling, feature scaling, and fold-safe imputation. Feature selection was implemented as a tunable component supporting stepwise *LDA*, *LASSO*, and *RF*-based ranking, with *PCA* as an alternative dimensionality-reduction step. Automated hyperparameter optimisation combined nested Bayesian and grid searches, and Ward's hierarchical clustering was equipped with automated cut-off optimisation for population estimation. All components operate within a unified *Python* environment to ensure consistent preprocessing, training, and evaluation across *sex*, *individual*, and *population* workflows.

To improve robustness under real-world conditions, the *FIT* framework was extended to handle incomplete morphometric datasets without excluding partially degraded prints. Enhancements include refined landmark definitions, incorporation of mixed-quality training data, and model-compatible imputation using the MissForest algorithm [73], which preserves multivariate relationships among observed variables. These methods are detailed in Kistner et al. [68], which showed that imputation substantially improves model stability and data utilisation.

Accordingly, all analytical pipelines include imputation within the preprocessing stage of their respective *ML* workflows, executed inside each *CV* fold to prevent information leakage.

Complementary work on footprint annotation and *Deep Learning (DL)*-based *species* classification is reported in Kistner et al. [49]. As this dissertation retains a morphometrics-centred methodological scope, these deep-learning components are not elaborated here but revisited in the Discussion to contextualise their integration within the broader evolution of *FIT*.

## 5.4 Chapter Summary

### Summary of Methods and Key Enhancements

This chapter outlined the methodological framework for morphometric footprint identification using *FIT*, covering *species* and *sex* classification, *individual identification*, and *population estimation*. The core pairwise protocol from the legacy *JMP* implementation was retained for comparability but re-engineered in a modular *Python* environment for automation, reproducibility, and scalability. Fixed random seeds ensured reproducible dataset splits, with *DNA*-validated field prints reserved for testing.

Enhancements included group-aware, stratified *CV* splits, staged workflows, and optional modules for outlier handling, feature scaling, and fold-safe imputation. Feature selection became a tunable core component, evaluated for accuracy, runtime, and stability across folds, with optional stacking of *sex*-model predictions before *individual-ID* analysis. *PCA* was implemented as an exclusive alternative to forward selection in controlled ablations.

Supervised modelling employed *LDA* as the interpretable baseline, with *XGBoost* evaluated for *sex* and *species* classification. Population estimation applied Ward's hierarchical clustering to the pairwise distance matrix, using refined cut-offs from sequential hold-out validation. Hyperparameters were optimised via Bayesian and grid search approaches, nested within outer *CV* folds to prevent information leakage.

Evaluation relied on task-specific primary metrics—*BA*, *DER*, and *V-measure*—and secondary measures for robustness. Benchmarking extended the best *Lutra lutra* pipelines to seven additional species under identical protocols. Finally, field-print validation on Portugal and Lower Saxony datasets assessed *sex*-prediction consistency and *individual-ID* plausibility relative to trail locations. Improvements addressing incomplete data—landmark refinement, mixed-quality training, and *MissForest* imputation—expanded usable sample size without loss of accuracy, strengthening the *FIT*'s applicability in real-world monitoring contexts while maintaining methodological transparency.

## Results

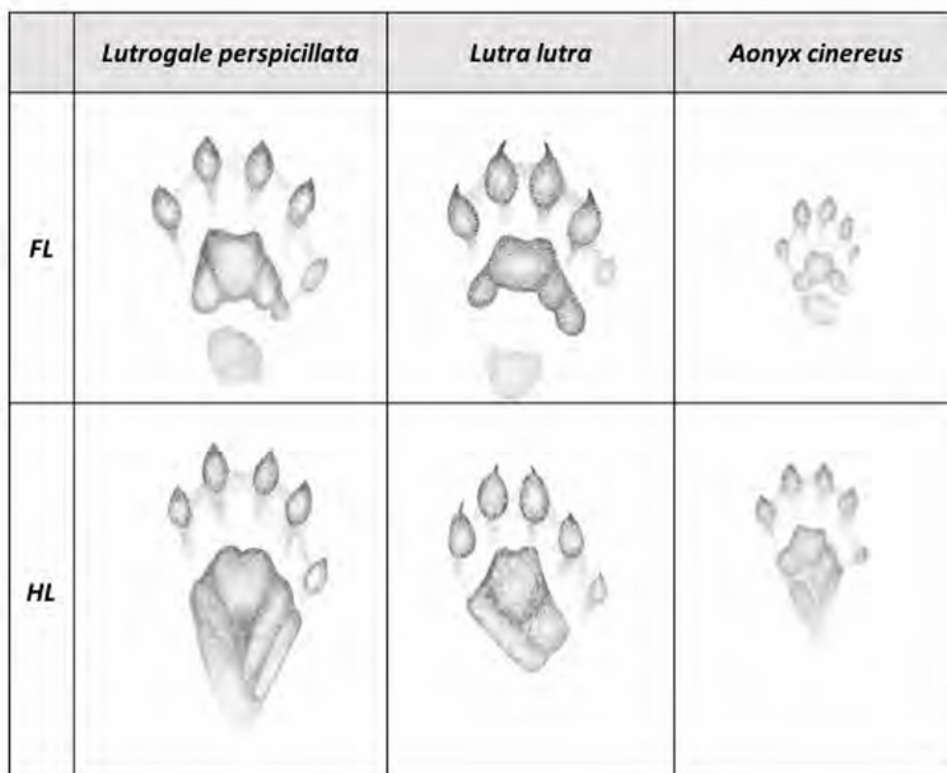
This chapter presents the results by analytical component. We begin with a traditional tracker-based description of species-level footprint characteristics, then turn to *FIT* morphometrics—covering *species*, *sex*, *individual identification*, and *population estimation*. Species identification builds on earlier work [48]; results from Kistner et al. [49] are not reproduced here to maintain a morphometrics-only scope, although their implications are discussed later. The chapter closes with targeted methodological improvements, including missing-value imputation implemented in the legacy *JMP* environment [68] and further advances in the new *Python* implementation. Most content is conveyed through stand-alone figures with detailed captions; the running text serves as a guide to these visual results.

We applied *FIT* morphometrics to four analytical tasks and refer directly to the plots containing the details. For *species* classification among three Asian otters, see the held-out confusion matrix and dominant geometric cues in Figs. 6.2–6.3. For *sex classification* and *individual identification*, we provide **stacked CV/test confusion matrices** in Figs. 6.6 and 6.8, respectively. Sex-classification results are further detailed by probability-based confidence stratifications (Figs. 6.6–6.7) and feature importances (Fig. 6.5). Population estimation is derived from hierarchical clustering with cross-validated cutoff selection and sequential-holdout evaluation (Figs. 6.9–6.10). Numerical performance metrics are reported in the corresponding figure captions.

After optimising the models on the Eurasian otter, we benchmarked the *FIT* pipeline across additional species. Table 6.1 summarises cross-validated and held-out results (mean  $\pm$  **Standard Deviation (SD)**; ranges) for sex classification, individual identification, and population estimation. Methodological improvements and their empirical effects are compiled in Tables 6.2–6.3.

## 6.1 Professional Tracker-Based Analysis and Description of Footprints

To generate visual reference material and define species-specific footprint characteristics, we compiled tracker-based technical drawings of front- and hind-left impressions for three Asian otter species. These illustrations summarise the diagnostic traits that inform the qualitative interpretation in this section (Fig. 6.1).

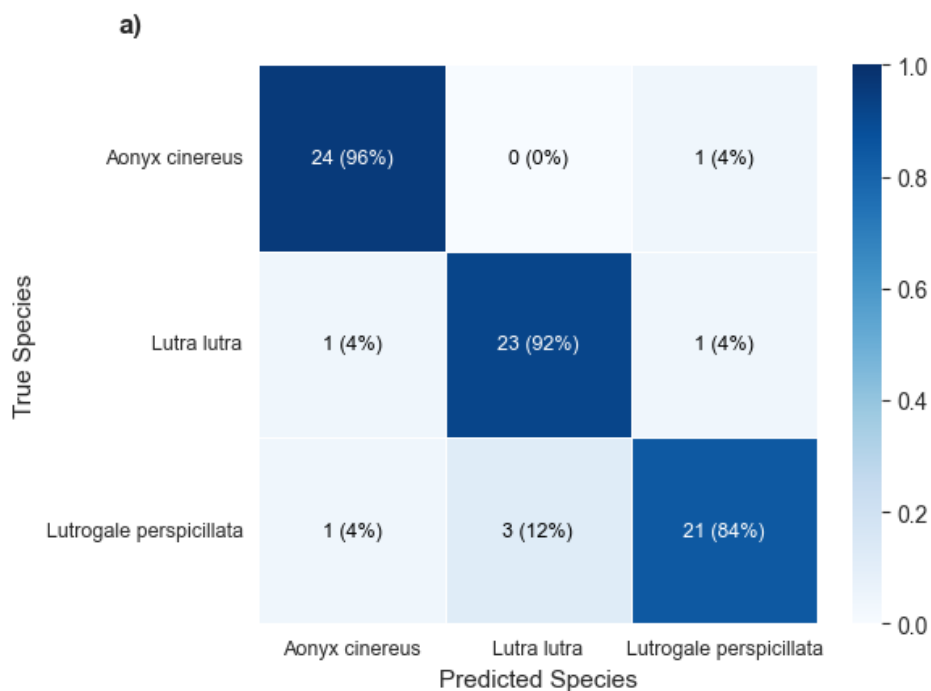


**Fig. 6.1.:** Tracker-based technical drawings of front-left (FL) and hind-left (HL) footprints for three Asian otter species (illustrations and copyright: Assaf Ben-David). **Smooth-coated otter** (*Lutrogale perspicillata*): large, elongated impression; claws short and blunt; toes oval with long proximal phalanges that register clearly; toe 1 (thumb) aligned almost straight beneath toe 2; negative space between toes and metacarpal/metatarsal pad approximately 1.5–2 × toe-pad size; metacarpal pad broad with a central saddle; metatarsal pad about twice the size of the metacarpal. **Eurasian otter** (*Lutra lutra*): front track less asymmetrical than the hind; overall shorter and less elongated than the other two species; claw marks small, sharp, and well defined; toes distinctly teardrop-shaped; front carpal pad rounded with the greatest negative space relative to the metacarpal. **Asian small-clawed otter** (*Aonyx cinereus*): smallest footprint; claws very short or absent; toes oval; elongated front foot with the carpal pad angled towards toe 5.

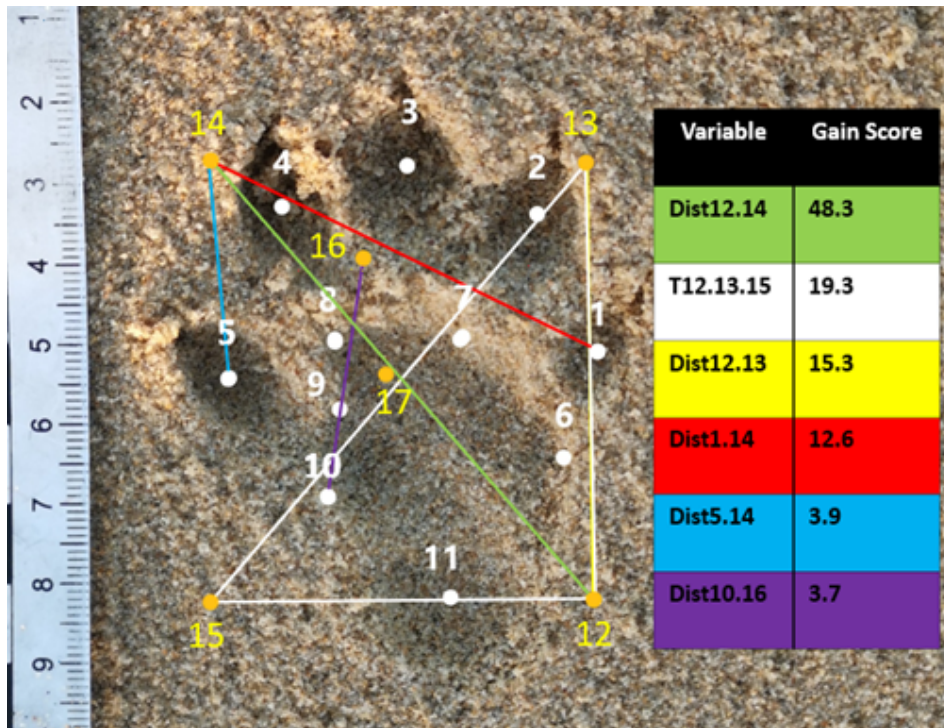
## 6.2 Morphometric Analyses

This section summarises the *FIT* morphometric results across four analytical targets, each directly linked to its figures: (i) *Species discrimination* among three Asian otters (Figs. 6.2–6.3); (ii) *Sex classification* in Eurasian otters, showing the best *CV* and test confusion matrices, prediction-confidence stratifications, and feature importances (Figs. 6.4a–6.5); (iii) *Individual identification*, summarised by stacked “Same/Different” *CV*/test confusion matrices (Fig. 6.8); and (iv) *Population estimation*, derived from hierarchical clustering with *CV*-based cutoff optimisation and sequential-holdout validation (Figs. 6.9–6.10).

### Species Classification of Asian Otters Using FIT

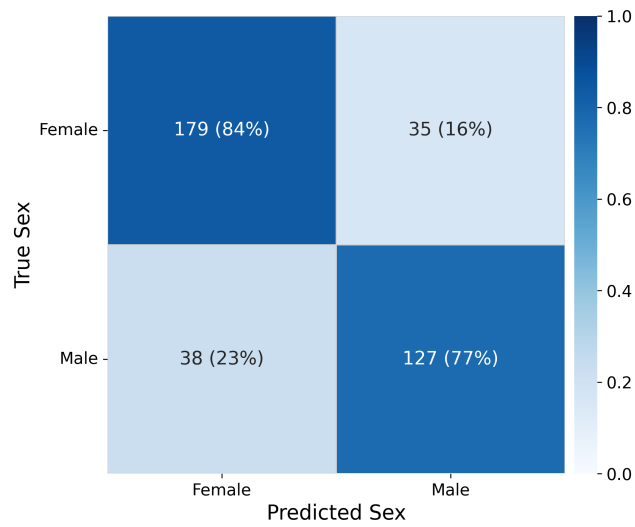


**Fig. 6.2.:** Test-set confusion matrix for three Asian otters using FIT morphometrics. Classes: *Aonyx cinereus*, *Lutra lutra*, *Lutrogale perspicillata*; **25 prints/species** ( $n = 75$ ), stratified train/test protocol. Cells report counts and row percentages of the *true* class; shading indicates row-normalised rates. A gradient-boosted tree model (XGBoost) trained on FIT features attains **91 %** accuracy on the held-out test set, whereas LDA yields **85 %** ([48]). Off-diagonal entries are rare, indicating strong separability at species level. Landmark definitions and feature construction follow the standard FIT scheme (Methods). See [48] for the original experiment design and dataset description.

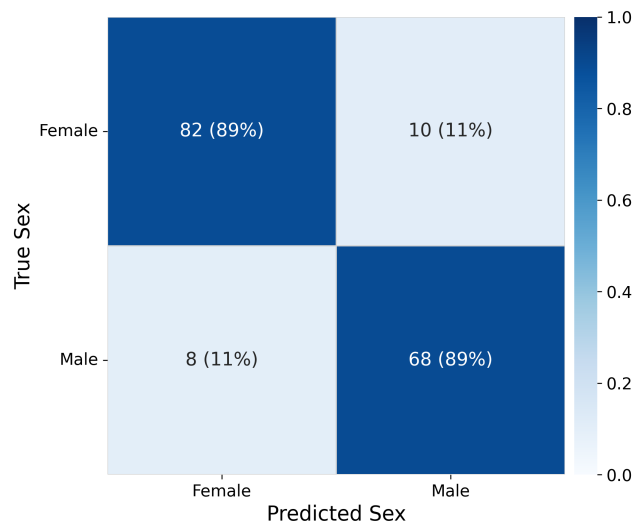


**Fig. 6.3.: Top features by XGBoost gain for species discrimination.** The gain profile shows that inter-landmark distance **Dist12–14** contributes nearly half of the predictive signal (~48 %), followed by **T12.13.15** and **Dist12.13**. Inter-distances dominate the ranking, indicating that species-level differences are captured mainly through proportional geometry of toe pads and the metacarpal region. Landmark numbering follows the standard FIT set (Methods); the background image illustrates the landmark layout for orientation. These patterns align with XGBoost outperforming LDA on the same task (91 % vs. 85 %; [48]).

## Sex Classification of Eurasian Otters Using FIT

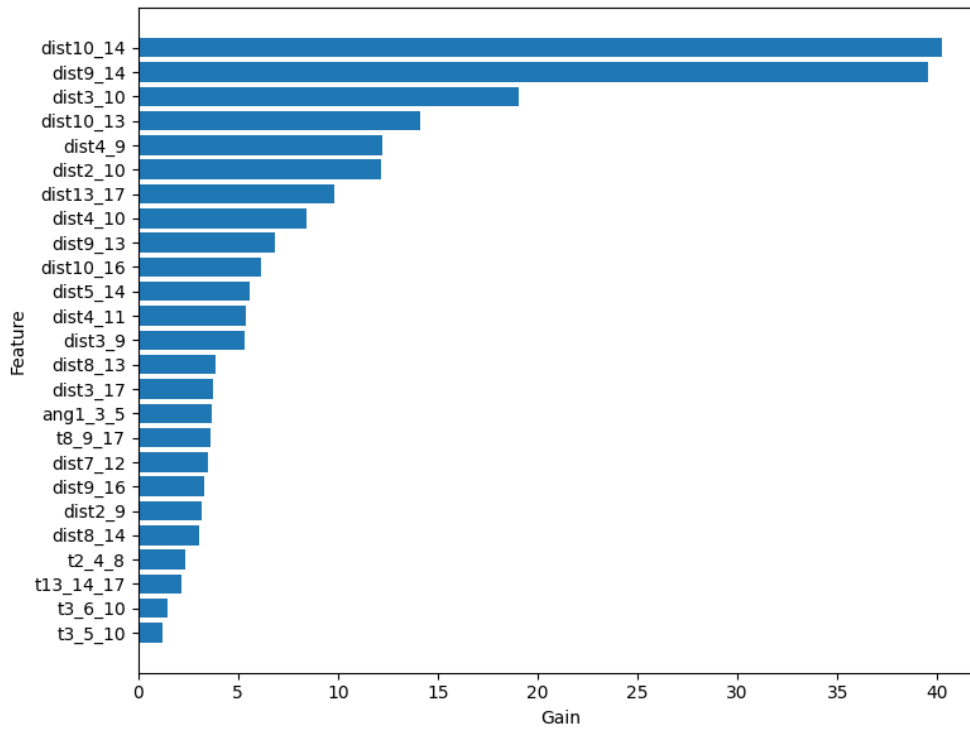


(a) Cross-validation

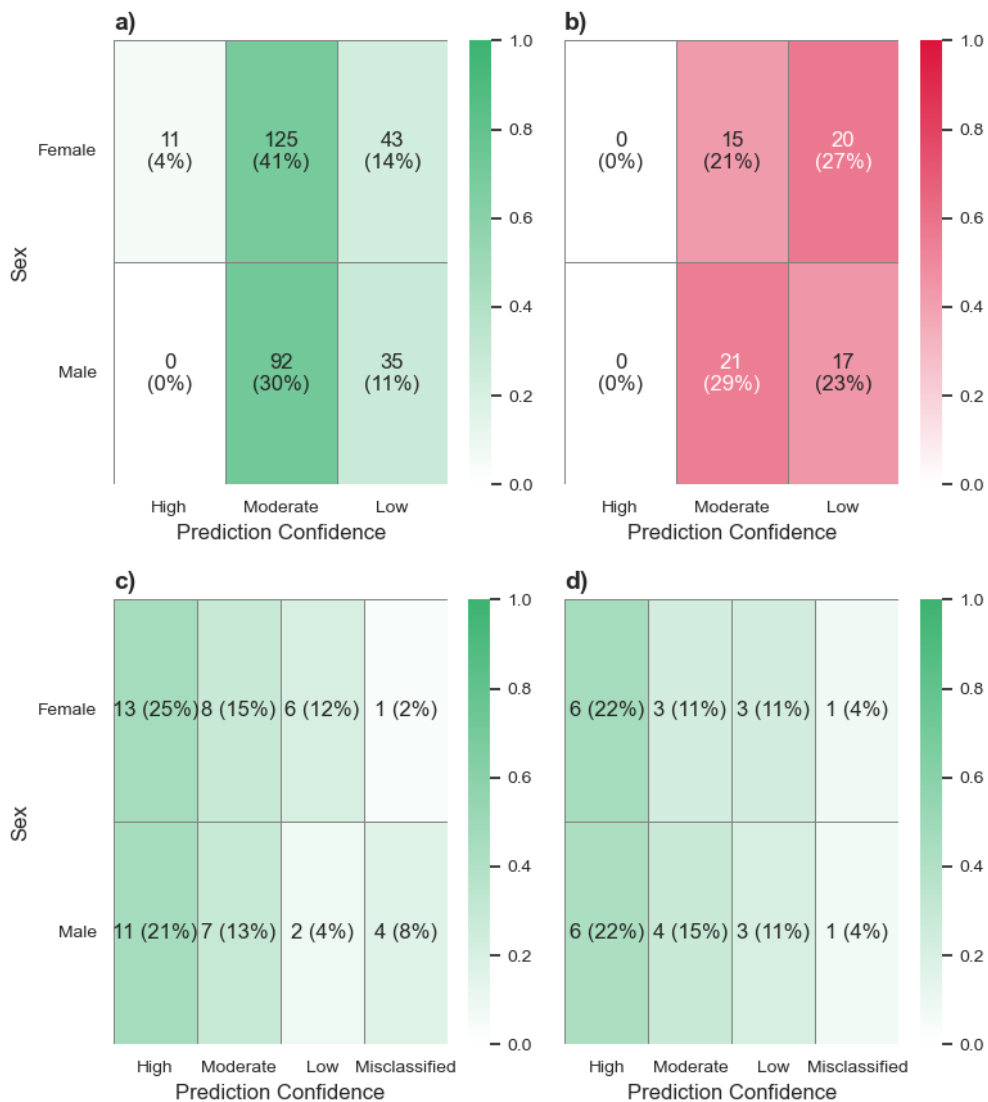


(b) Independent test set

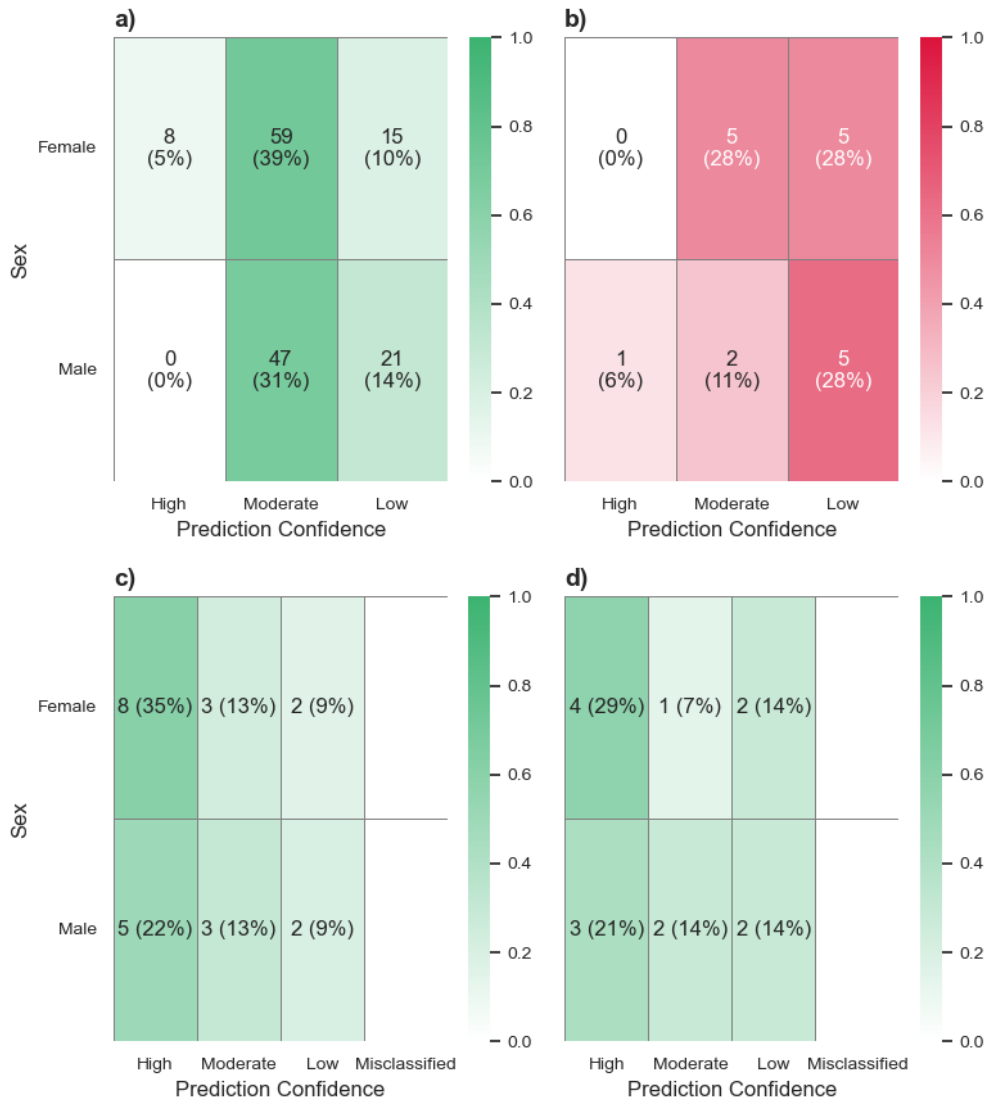
**Fig. 6.4.:** Eurasian otter sex classification: confusion matrices for *CV* and independent test. XGBoost on *FIT* morphometrics; hyperparameters via Bayesian optimisation with stratified grouped five-fold *CV* and a hold-out test set. Cells show counts and row-normalised percentages by true sex (colour encodes percentages). *CV* **BA**=81 % (F 84 %, M 77 %); test **BA**=89 % (F 89 %, M 89 %).



**Fig. 6.5.:** Top-25 feature importances for the Eurasian otter sex classifier (XGBoost; gain metric). Bars report each feature’s *gain*, i.e., the expected reduction in the model’s loss attributable to splits using that feature; values are relative and sum to 100 %, reflecting *model usage* rather than causal effect sizes. The ranking is **distance-dominated**: inter-landmark measures such as `dist10_14`, `dist9_14`, `dist3_10`, and `dist10_13` lead the profile, while angular or area-like composites appear lower. This pattern indicates that footprint **scale and proportions**—e.g., pad spacing and toe spread within the FIT landmark system—carry most of the sex signal.

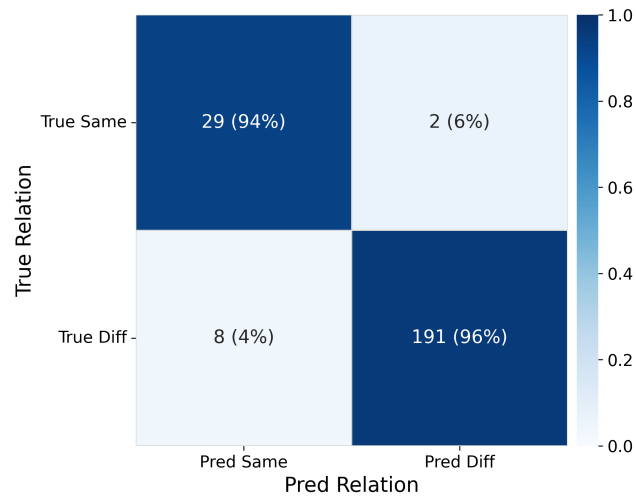


**Fig. 6.6.: Confidence heatmaps for cross-validation folds.** Predictions are partitioned into (a) correct single-print, (b) misclassified single-print, (c) trail-level aggregated, and (d) individual-level aggregated outcomes. Confidence categories (*high/moderate/low*) are derived by thresholding the classifier’s probability outputs (see Methods). Percentages are computed by *true sex within each panel* to enable like-for-like comparison across strata. Misclassifications concentrate in the *low–moderate* bins, whereas aggregation shifts mass into the *high* bin.

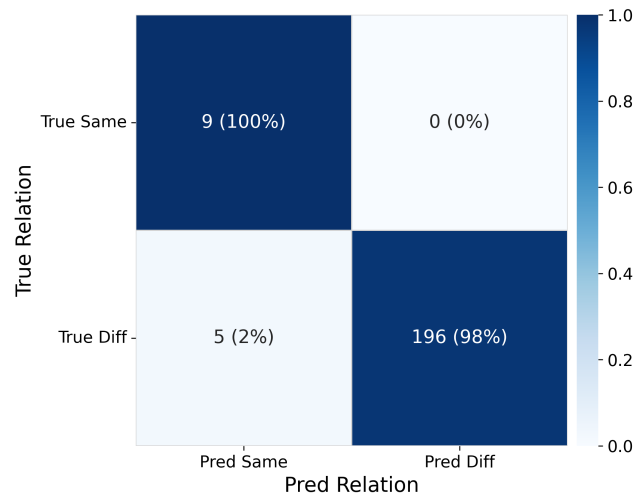


**Fig. 6.7.:** Confidence heatmaps for the independent test set. Panels mirror Fig. 6.6—(a) correct single-print, (b) misclassified single-print, (c) trail-level aggregation, (d) individual-level aggregation—and exhibit the same qualitative pattern: most single-print errors occur at *low–moderate* confidence, while aggregation shifts predictions into the *high* band.

## Individual Identification



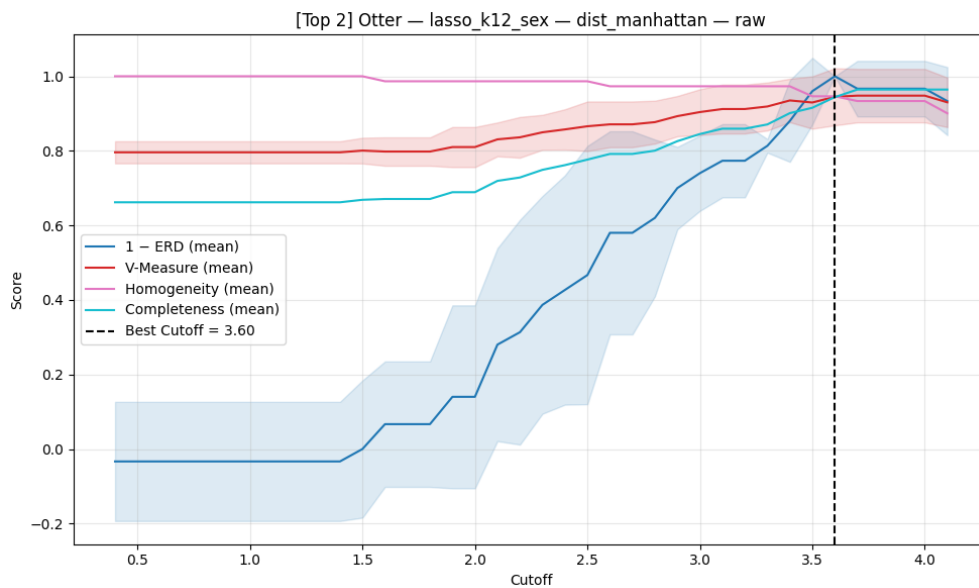
(a) Cross-validation



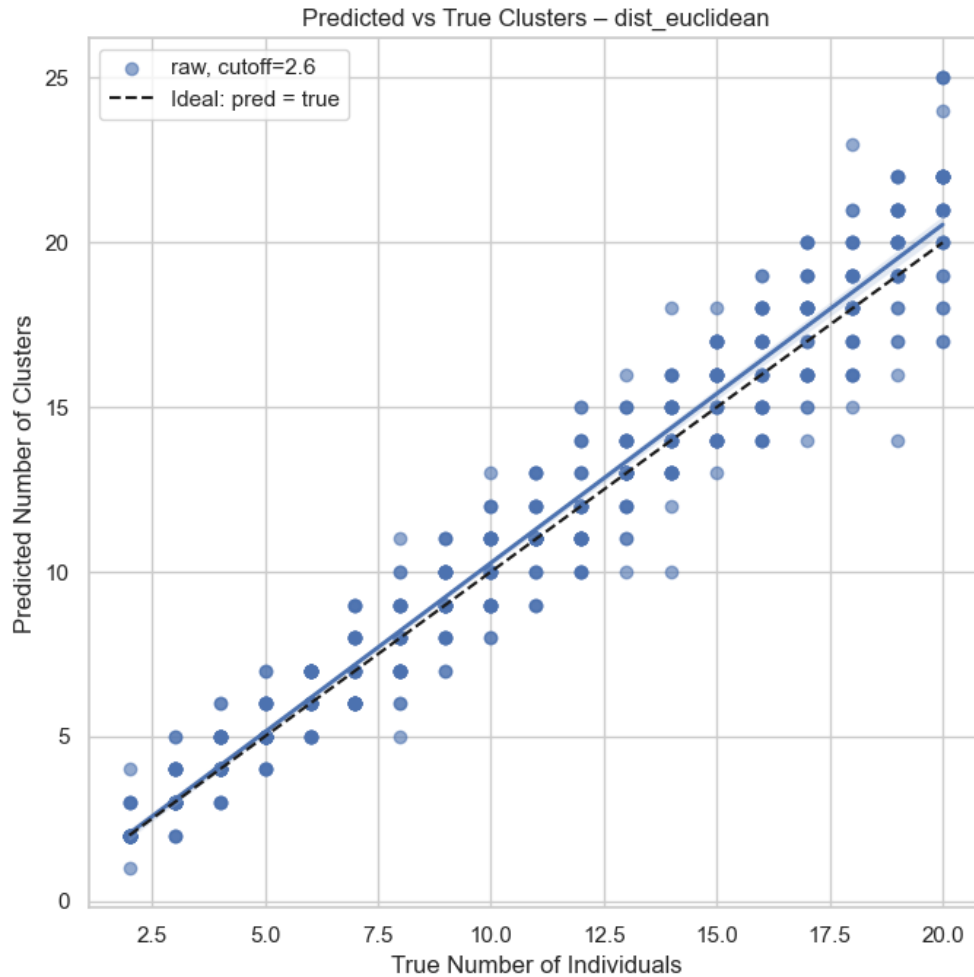
(b) Independent test set

**Fig. 6.8.: Eurasian otter individual identification (Same/Diff): confusion matrices for CV and independent test.** Pipeline: LDA in canonical space with `SelectK=12` (lasso-guided); pairwise setup and stratified grouped folds as in Methods; matching train/test split. Cells: counts and row-normalised percentages by true label. **CV BA=94.8 %**; **test BA=98.8 %**, with no within-class confusions (strong generalisation).

## Unsupervised Individual Identification and Population Estimation



**Fig. 6.9.: Population estimation for Eurasian otter via hierarchical clustering in canonical space: cross-validated cutoff selection.** Cross-validation curves for the Top-2 configuration (lasso\_k12\_sex, Manhattan distance, raw features). We plot mean $\pm$ SD across folds for 1 - ERD, V-measure, homogeneity, and completeness; the **dashed vertical line** marks the selected cutoff ( $\approx 3.60$ ) according to the combined score  $S_\alpha = \alpha(1 - \text{ERD}) + (1 - \alpha)V$  (see Methods for DER/V definitions and choice of  $\alpha$ ). The chosen cutoff is used for external evaluation in Fig. 6.10.



**Fig. 6.10.: Population estimation for Eurasian otter via hierarchical clustering: sequential-holdout evaluation.** Evaluation for a second configuration (dist\_euclidean, raw features; cutoff **2.6**). Each point contrasts the *predicted* with the *true* number of individuals; proximity to the **diagonal** indicates accurate recovery across varying ground-truth sizes. Together with the CV cutoff analysis (Fig. 6.9), these results show that the CV-chosen decision threshold generalises to held-out data and that FIT morphometrics contain usable population signal.

## 6.3 Benchmarking with Other Species

After optimising the models for Eurasian otters, the pipeline was benchmarked across additional species. Performance for *sex classification*, *individual identification*, and *population estimation* is summarised in Table 6.1. Metrics are reported as mean  $\pm$  *SD* and range, derived from *CV* on the training data and inference on held-out test sets.

**Tab. 6.1.:** Benchmarking across species: (a) Sex classification, (b) Pairwise individual identification, (c) Population clustering. Values: mean  $\pm$  SD (min–max). Bold in brackets = best model (*DER*: Minimum; *BA (CV)/V*: Maximum).

(a) Sex classification (*BA*)

Species	<i>BA (CV)</i>	<i>BA (Test)</i>
Amur tiger	0.94 $\pm$ 0.01 (0.88–0.95)	0.89 $\pm$ 0.02 (0.85–0.96)
Mountain lion	0.92 $\pm$ 0.01 (0.89–0.95)	0.74 $\pm$ 0.04 (0.66–0.83)
Bengal tiger	0.80 $\pm$ 0.03 (0.72–0.86)	0.70 $\pm$ 0.05 (0.58–0.83)
Giant panda	0.79 $\pm$ 0.01 (0.73–0.84)	0.39 $\pm$ 0.02 (0.35–0.47)
Eurasian otter	0.78 $\pm$ 0.02 (0.73–0.81)	0.85 $\pm$ 0.03 (0.80–0.91)
Lowland tapir	0.55 $\pm$ 0.03 (0.46–0.59)	0.48 $\pm$ 0.03 (0.38–0.56)
Cheetah	0.53 $\pm$ 0.03 (0.43–0.58)	0.51 $\pm$ 0.04 (0.37–0.64)
White rhino	0.53 $\pm$ 0.03 (0.44–0.60)	0.37 $\pm$ 0.06 (0.26–0.60)

(b) Pairwise individual identification (*BA*)

Species	<i>BA (CV)</i>	<i>BA (Test)</i>
Amur tiger	0.96 $\pm$ 0.02 (0.89–0.99)	0.93 $\pm$ 0.05 (0.63–0.99)
Bengal tiger	0.93 $\pm$ 0.02 (0.84–0.97)	0.98 $\pm$ 0.02 (0.78–1.00)
Cheetah	0.98 $\pm$ 0.01 (0.95–0.99)	0.98 $\pm$ 0.01 (0.94–1.00)
Eurasian otter	0.90 $\pm$ 0.03 (0.81–0.96)	0.90 $\pm$ 0.07 (0.60–0.99)
Giant panda	0.95 $\pm$ 0.02 (0.87–0.98)	0.92 $\pm$ 0.04 (0.70–0.99)
Lowland tapir	0.92 $\pm$ 0.03 (0.85–0.98)	0.91 $\pm$ 0.04 (0.84–0.97)
Mountain lion	0.97 $\pm$ 0.01 (0.91–1.00)	0.94 $\pm$ 0.04 (0.85–1.00)
White rhino	0.94 $\pm$ 0.01 (0.90–0.97)	0.96 $\pm$ 0.02 (0.91–0.99)

(c) Population clustering (*DER* & V-Measure)

Species	<i>DER</i>	V-Measure
Amur tiger	0.07 $\pm$ 0.05 (0.01–0.51)	0.92 $\pm$ 0.05 (0.65–0.98)
Bengal tiger	0.10 $\pm$ 0.07 (0.01–0.70)	0.92 $\pm$ 0.05 (0.59–1.00)
Cheetah	0.08 $\pm$ 0.08 (0.00–0.72)	0.96 $\pm$ 0.06 (0.57–1.00)
Eurasian otter	0.15 $\pm$ 0.16 (0.00–1.00)	0.84 $\pm$ 0.06 (0.59–0.96)
Giant panda	0.13 $\pm$ 0.09 (0.00–0.33)	0.90 $\pm$ 0.06 (0.63–0.96)
Lowland tapir	0.07 $\pm$ 0.07 (0.00–0.55)	0.95 $\pm$ 0.05 (0.66–0.99)
Mountain lion	0.06 $\pm$ 0.07 (0.00–0.55)	0.95 $\pm$ 0.06 (0.64–0.99)
White rhino	0.15 $\pm$ 0.16 (0.00–0.98)	0.87 $\pm$ 0.06 (0.51–0.99)

## 6.4 Improvements to the *FIT* Workflow

This section introduces four targeted modifications to the *FIT* workflow aimed at improving feature selection, model stacking, and missing-value handling. Their comparative performance is summarised in Tables 6.2 and 6.3, which present the empirical evaluation of different selector algorithms, the integration of predicted sex information, and multiple imputation approaches under both simulated *Missing Completely At Random (MCAR)* and real-data conditions. A comprehensive analysis of missing-value mechanisms and imputation outcomes is provided in Kistner et al. [68].

**Tab. 6.2.: Improvements to the *FIT*.** (a) Feature selection: forward selection vs. *LASSO* and *RF*; *LASSO/RF* are 2–3× faster, with *LASSO* giving the highest mean *BA* on *CV*/test. (b) Sex stacking: adding predicted sex (*Morph\_Sex*) as a covariate slightly increases *CV BA* without affecting test performance. Values: mean ± *SD* (ranges).

(a) Feature selection. *LASSO* and *RF* are ~2–3× faster than forward; *LASSO* yields the highest mean *BA*.

Selector	Runtime (s)	<i>BA</i> ( <i>CV</i> )	<i>BA</i> (Test)
forward	34.48 ± 21.77	0.94 ± 0.03 (0.81–1.00)	0.94 ± 0.04 (0.63–1.00)
lasso	11.83 ± 6.73	0.95 ± 0.03 (0.87–1.00)	0.95 ± 0.04 (0.81–1.00)
random_forest	12.63 ± 7.12	0.94 ± 0.03 (0.84–0.99)	0.93 ± 0.06 (0.60–1.00)

(b) Sex stacking. Adding the sex signal (*Morph\_Sex*) slightly increases *CV BA* and maintains test performance vs. *Morph*.

Feature space	<i>BA</i> ( <i>CV</i> )	<i>BA</i> (Test)
<i>Morph</i>	0.94 ± 0.03 (0.81–1.00)	0.94 ± 0.05 (0.60–1.00)
<i>Morph_Sex</i>	0.95 ± 0.03 (0.86–1.00)	0.94 ± 0.04 (0.63–1.00)

**Tab. 6.3.: Imputation performance.** (c) Simulated *MCAR*: all strategies outperform case deletion; **missRanger** attains the best average rank, while **missForest** yields the highest mean *BA* near the complete-data control. (d) Real cheetah data: under *Missing Delete*, **missForest/missRanger** perform best; under *Missing & Unclear Delete*, *MICE – PMM* achieves the highest *BA* and lowest *DER*. For full details, see Kistner et al. [68].

(a) *MCAR* imputation. All methods outperform Delete Case; **missRanger** has the best average rank, **missForest** the highest *BA* near control.

Imputation Method	No. of Datasets	<i>BA</i> (SD)	<i>DER</i> (SD)	Avg. Rank
Control	5	0.93 (0.02)	0.08 (0.06)	–
Delete Case	30	0.51 (0.34)	0.51 (0.38)	5.63
Mean Imputation	30	0.82 (0.09)	0.07 (0.07)	3.51
MICE PMM	30	0.84 (0.09)	0.09 (0.08)	2.92
MICE RF	30	0.80 (0.10)	0.11 (0.10)	3.08
missForest	30	0.90 (0.06)	0.14 (0.10)	3.15
missRanger	30	0.88 (0.07)	0.08 (0.06)	2.72

(b) Real-data (cheetah). Left: only missing landmarks removed; right: missing & unclear removed. **missForest/missRanger** lead in the first; *MICE PMM* leads in the second.

Imputation Method	Missing Delete			Missing & Unclear Delete		
	<i>BA</i>	<i>DER</i>	Avg. Rank	<i>BA</i>	<i>DER</i>	Avg. Rank
Control	0.80	0.00	–	0.80	0.00	–
Delete Case	0.81	0.12	5.75	0.60	0.18	5.75
Mean Imp.	0.90	0.06	3.00	0.86	0.12	2.75
MICE PMM	0.84	0.12	5.25	0.88	0.06	1.50
MICE RF	0.87	0.00	2.75	0.75	0.00	2.50
missForest	0.90	0.00	1.25	0.81	0.12	3.25
missRanger	0.90	0.06	3.00	0.65	0.18	5.25

## 6.5 Chapter Summary

### Results Overview: FIT and Field Outputs

This chapter established a comprehensive foundation for footprint-based monitoring through both visual and analytical advances. Professionally rendered tracker drawings provided the first consistent field reference across Asian otter species, forming the basis for future *FIT* training and field-guide materials. *FIT* morphometrics achieved high accuracy across all analytical targets: species classification among Asian otters reached 91 % with *XGBoost*, outperforming *LDA* (85 %); the optimised Eurasian-otter sex model achieved 81 % *BA* in *CV* and 89 % on the test set, rising to 90–100 % when aggregated at trail or individual level. Pairwise individual identification using *LDA* with *LASSO*-guided features reached 94.8 % (*CV*) and 98.8 % (test). Population estimation via hierarchical clustering yielded low *DER* and high *V-measure*, accurately recovering true group sizes. Benchmarking across eight species confirmed strong generalisability, demonstrating that *FIT* can reliably support standardised, non-invasive wildlife identification and population assessment across taxa.



# Part IV

---

Synthesis, Critical Discussion, and Future  
Perspectives



## 7.1 Summary of Main Achievements

### Summary of Objectives, Achievements, and Outlook

**Development and field validation.** The *FIT* framework was adapted and validated for Eurasian otters, establishing a practical monitoring pipeline based on expert-annotated footprints. Model evaluation across species, sex, individual, and population levels—supported by *DNA*-referenced data—confirmed high accuracy and robustness under field conditions.

**Methodological expansion.** Originally developed as the proprietary *JMP FIT add-in*, the framework was re-engineered in open-source *Python (FIT-Py)* for automated, reproducible, and scalable analysis. Deep-learning components (*CNN*) extended *FIT* beyond a single species and integrated it with modern *ML* workflows for transferable, data-driven monitoring.

**Ongoing developments.** Core elements of *FIT-Py* are operational but not yet fully deployed across datasets. Current work focuses on systematic annotator calibration, automated detection and quality control, and user-friendly interfaces via API links to the WildTrack AI platform or local desktop tools (e.g., *JMP*). Parallel efforts expand *DNA*-referenced in-situ datasets to strengthen validation and generalisation. Uncertainty quantification and model interpretability are being formalised into dedicated modules to ensure transparent, reproducible evaluation.

## Field Data and Collaborative Standardisation

The methodological development of the *FIT* for Eurasian otters followed a sequential process linking empirical field validation, structured annotation, and quantitative assessment of data quality. The field campaigns in Germany and Portugal provided the empirical foundation for this work, representing the first systematic validation of the *FIT* under contrasting environmental and logistical conditions. Despite limited standardisation and variable substrate quality, the campaigns demonstrated that reliable morphometric data can be obtained even under non-ideal circumstances—provided that trail integrity, *QC*, and expert verification are prioritised. This stage established the empirical constraints within which later modelling and imputation strategies were designed.

Crucially, zoo-based reference datasets played a central role in the methodological evolution of the *FIT*. Although these animals were not *DNA*-calibrated, they offered verified individual identities under controlled conditions, enabling the construction of consistent ground truth data. This pragmatic but reliable reference base made it possible to train and validate analytical models, benchmark intra- and inter-individual variation, and systematically evaluate sources of measurement noise. In contrast to opportunistic field data, zoo collections provided continuity, controlled substrates, and repeatable imaging conditions—conditions indispensable for testing algorithmic reproducibility and for calibrating feature stability across sessions. In this sense, zoo data were not ancillary but foundational to the transition of the *FIT* from a descriptive to an analytical, model-driven framework.

The integration of *FIT*-based identification with *DNA* profiling in the field later served as an external validation step, confirming that morphological and molecular approaches can complement one another to strengthen confidence in individual-level assessments. This hybrid validation also clarified the limits of substitution: *FIT* can replicate much of the information captured by genetic methods, but its reliability depends on consistent ground truth and environmental context.

Field performance patterns further underscored the sensitivity of footprint detection to habitat conditions. While inland sites in Germany yielded sparse or fragmented trails, Portugal's coastal systems consistently produced high-quality impressions. Substrate type, moisture, and microclimate jointly determined footprint persistence, highlighting the need for environmental metadata to be systematically captured and linked with remote-sensing and meteorological layers.

Community-science approaches offer clear advantages for scaling data collection and fostering long-term monitoring participation. Yet, as demonstrated by Finerty et al. [65], the full potential of the *FIT* can only be realised through collaboration with professional trackers. Their expertise in interpreting trail context, assessing substrates, and verifying print integrity directly governs data quality — the key determinant of whether *FIT* analyses yield scientifically meaningful results. Effective cooperation between citizen observers and trained trackers therefore represents the most promising strategy for combining scalability with methodological rigour, ensuring that community-driven data meet the standards required for reliable analytical inference.

Together, these experiences point to the need for unified, standardised field protocols when working with footprint data. Such protocols should harmonise substrate documentation, environmental metadata, and sampling criteria across regions and teams, while adhering to clear ethical principles. Beyond methodological consistency, they must ensure minimal disturbance to animals and habitats, transparent data governance, and equitable participation in line with community science ethics. Only through coordinated, ethically grounded field standards can community-driven and professional datasets be seamlessly integrated, enabling reproducible, large-scale analyses that fulfil *Findable, Accessible, Interoperable, Reusable (data principles) (FAIR)* data principles.

## Annotation, Data Handling, and Technical Standardisation

Annotation and metadata management formed the second methodological axis. Manual landmarking in *JMP* ensured geometric accuracy but limited traceability, as image transformations and uncertainty flags were inconsistently recorded. These limitations echo the missing-data issues analysed in Kistner et al. [68], where robust imputation performance was shown to depend on well-documented data provenance. Applying missing-value imputation as a preprocessing step further improved data retention by allowing incomplete footprints to be retained rather than discarded, thereby stabilising downstream classification and reducing bias in population-level inference. Algorithmic correction can only recover information that is explicitly structured; thus, metadata completeness and annotator calibration determine both imputation reliability and model interpretability.

Standardised, computer vision ready formats such as *Common Objects in Context (computer vision dataset / annotation format) (COCO)*, as implemented in the *WildTrackAI* database [14], should therefore become standard. Tools with low entry barriers, such as the WildTrack app or digital field sketch interfaces, can expand participation while maintaining reproducibility and *FAIR* compliant standards. Establishing common metadata templates for footprint datasets, covering image parameters, transformations, annotator identity, and substrate descriptors, will be key for interoperability across species and studies.

Although traditional morphometric workflows may appear outdated compared to modern deep learning pipelines based solely on bounding boxes, such as those demonstrated in the *OpenAnimalSet*, recent unpublished results from WildTrack (pers. comm. Zoe Jewell and Sky Alibhai) suggest that comparable accuracy can be achieved with purely *DL* based detectors for certain target metrics. Nevertheless, a landmark oriented approach retains critical advantages in interpretability and cross substrate robustness. In my view, well defined anatomical landmarks, such as the centroid points used for otter toes, offer a more stable geometric reference than bounding box representations, particularly under substrate specific distortion and partial print conditions that are inherent to spoor based monitoring. The combination of explicit morphometrics with convolutional feature learning thus represents a promising hybrid direction that merges biological interpretability with the efficiency of deep learning. As highlighted by Petso et al. [100], such integrative frameworks provide both improved transparency and stronger generalisation across heterogeneous datasets, underlining that morphological structure remains central to explainable *AI* in conservation biometrics.

Automation and active learning strategies further provide a pathway to reconcile efficiency with expert oversight. Human in the loop frameworks [101] enable *AI* models to prioritise the most informative samples while retaining tracker validation where interpretive skill adds greatest value. Rather than replacing human expertise, such hybrid systems strategically combine fieldcraft, *TEK*, and data centric *ML* to advance the *FIT* towards a fully operational, scalable monitoring system. Ultimately, the credibility of *FIT* based population inference depends as much on the fidelity and structure of its training data, grounded in controlled zoo references and rigorous metadata design, as on the sophistication of its analytical models.

## Critical appraisal of classification models

This section critically examines the classification models developed in this dissertation with respect to their methodological implications and comparability to related studies. The discussion focuses on interpretability, robustness, and the degree to which model performance can be generalised beyond the specific datasets used.

Across all targets, morphometric *FIT* models demonstrated strong discriminatory power, mostly consistent with earlier morphometric frameworks for species individual identification and sex classification in large mammals [19, 21, 22, 48, 52, 61].

Within the context of otter research, the present *FIT* framework builds on a progressive line of morphometric studies that first established the feasibility of footprint-based discrimination. **Hartwick1998** provided the earliest empirical evidence that individual Eurasian otters could be distinguished by quantitative footprint measurements, while Mercier and Fried [47] demonstrated consistent sexual dimorphism in footprint shape, establishing the foundation for sex classification. Vetrovcova [50] subsequently formalised these insights by developing a multiclass *LDA* approach for individual identification, thereby defining the conceptual basis that later evolved into the *FIT* framework.

The present work extends this lineage by reformulating the task as a pairwise classification problem with probabilistic decision boundaries, enabling systematic threshold optimisation and distance-based clustering for population-level inference. In doing so, the *FIT* workflow operationalises and expands Vetrovcova’s recommendation to refine *LDA* identification using *FIT*. The high concordance between pairwise *FIT* predictions and field-derived population patterns thus corroborates these earlier findings, confirming that footprint geometry carries consistent, quantifiable individual and sex-specific signatures.

Compared to end-to-end *DL* frameworks, such as those in Kistner et al. [49] and Shinoda and Shiohara [67], the morphometric approach achieved higher accuracy but within a narrower taxonomic scope. While Kistner et al. [49] reported substantially higher accuracies than Shinoda and Shiohara [67], this difference largely reflects dataset composition rather than algorithmic performance: the *OpenAnimalTracks* dataset spans 18 species under heterogeneous imaging and annotation conditions, whereas the present and earlier otter-specific studies addressed only three closely related species.

The *OpenAnimalTracks* dataset reported an average of roughly 69 % accuracy using a SwinTransformer baseline, compared to near-perfect discrimination achieved for the sympatric otter species here. This contrast highlights that controlled, species-specific models can outperform generalised architectures trained on heterogeneous data, yet they also risk limited transferability. A closer inspection of the *OpenAnimalTracks* dataset revealed frequent annotation inconsistencies, including misassigned species labels and unlabelled impressions, which likely contributed to its lower overall performance and further emphasise that data integrity can outweigh model complexity.

Moreover, morphometric *FIT* workflows require only standard computational resources and can be executed on conventional hardware without dedicated GPUs, making them accessible for field-based or institutional monitoring contexts where computational capacity is limited. In contrast, large-scale *DL* pipelines demand substantial processing power, large training volumes, and extended pre-training time, which often restricts their use in applied conservation settings. The stronger generalisation observed in Shinoda and Shiohara [67] thus reflects dataset scale rather than methodological superiority, as broader ecological coverage was achieved at the expense of data precision and interpretability.

In this light, the *FIT* framework offers a practically scalable and computationally lightweight alternative to deep-learning architectures—favouring interpretability, reproducibility, and cost-efficiency over universal scalability.

Hybrid models integrating geometric landmarks and *CNN*-based feature extraction, as explored by Petso et al. [100] and in unpublished results from WildTrack (*pers. comm. Zoe Jewell and Sky Alibhai*), point to a convergence between interpretable morphometrics and scalable computer vision. These mixed architectures parallel trends in conservation biometrics, where shape-informed *DL* has improved explainability and reduced dependence on purely data-driven features [100]. The present morphometric pipeline contributes to this line of work by demonstrating that feature interpretability—rather than model depth—can provide a stable bridge between expert knowledge and algorithmic classification. In this sense, *FIT*-based morphometrics remain relevant not as a legacy technique but as a structured baseline for hybrid explainable models.

Within the broader context of the morphometric *FIT* framework, this work has already introduced key methodological improvements that address several of the limitations identified here; these developments are discussed in detail in the following section.

## 7.2 General Improvements of *FIT* and Remaining Challenges When Working With Footprints

From its conception, the *FIT* framework has proven a robust analytical system for spoor-based monitoring. The methodological advances developed in this dissertation therefore represent refinements rather than replacements—aimed at increasing scalability, reproducibility, and accessibility across research and conservation contexts.

A major improvement was the transition from the proprietary *JMP*-based implementation to an open, modular *Python* pipeline. While the *JMP* environment remains highly valuable as an intuitive, explorative tool for non-coding practitioners, the new *Python* workflow enables automated benchmarking, cross-validated model evaluation, and reproducible experimentation. Together, the two systems form complementary layers: *JMP* for accessible field application and visual interpretation, and *Python* for large-scale, research-driven development. The comparative strengths and use cases of both are summarised in Table 7.1.

**Tab. 7.1.:** Comparison of FIT implementations for conservation monitoring: *JMP* vs. *Python* package.

Implementation	Key points
<b>FIT in <i>JMP</i> (application focus)</b>	<ul style="list-style-type: none"> <li>+ All-in-one GUI: image annotation, modelling, and visualisation within a single platform.</li> <li>+ Low barrier to entry; well suited to conservation practitioners and monitoring teams.</li> <li>+ Monitoring-oriented workflows and reporting tools.</li> <li>– Limited persistence of image-level annotations; no AI-assisted pre-labelling.</li> <li>– Sequential holdout schemes and batch evaluations not fully automated; each split requires manual initialisation.</li> <li>– Algorithmic extensions and species-specific adaptations only partially automated; substantial manual parameterisation required.</li> <li>– Developed in JSL (proprietary); commercial software with potential licensing costs.</li> </ul>
<b>FIT Python package (R&amp;D focus)</b>	<ul style="list-style-type: none"> <li>+ Modular, script-based design for experimentation within an open-source Python ecosystem.</li> <li>+ High automation and reproducibility via pipelines, staged workflows, and batch experiments.</li> <li>+ Group-aware, stratified splits to prevent data leakage and enable systematic benchmarking.</li> <li>– Higher entry barrier due to coding requirements.</li> <li>– Image annotation not yet integrated (external tools required).</li> </ul>

These developments illustrate that improving *FIT* is not primarily a matter of inventing new algorithms but of integrating existing components into reproducible, transparent, and user friendly workflows. While the *Python* version opens new possibilities for automation and comparative research, maintaining an interface with a low entry barrier remains essential for practitioners in the field. Future progress should therefore prioritise bridging these layers, linking the computational power of open source environments with the usability of graphical platforms, to ensure that methodological innovation continues to translate into practical conservation impact.

## Methodological Reflection and Limitations

In reflecting on these developments, it is important to acknowledge that the *FIT* framework, despite its growing analytical maturity, remains dependent on data quality, interpretative skill, and cross disciplinary cooperation. Its success therefore lies not in technological sophistication alone but in the continued integration of empirical field knowledge, statistical transparency, and ethical responsibility. Future progress will depend as much on humility in interpreting results as on innovation in producing them.

## 7.3 Synthesis and Final Discussion

This dissertation advances the *FIT* from a species specific framework for Eurasian otters to a reproducible, scalable system for noninvasive population monitoring. Starting from otter footprints, it demonstrates that morphological, behavioural, and environmental variability can be systematically modelled to yield reliable individual and population level inferences. Even under semiaquatic, substrate limited conditions, traditionally seen as unsuitable for track based monitoring, *FIT* produced biologically meaningful results when supported by standardised protocols and expert validation.

Building on this empirical foundation, *FIT* evolved from classical morphometrics to machine learning, from controlled to field validation, and from idealised to incomplete data. These developments transformed footprints from qualitative traces into reproducible biometric evidence, showing that transparent, data-driven analyses of non-invasive material can approach the reliability of laboratory based methods while remaining more ethical and field practical.

The new *Python*-based workflow consolidates these advances within an open, modular environment that automates preprocessing, feature selection, and validation while minimising data leakage. It enables reproducible benchmarking across taxa and supports integration of morphometric and image-based *AI* models. Yet accessibility remains central: the established *JMP* interface continues to serve practitioners through its intuitive, low-barrier design. Together, both systems illustrate that sustainable progress depends on coupling analytical rigour with operational usability rather than privileging one over the other.

Ultimately, *FIT*'s reliability rests more on data quality than algorithmic sophistication. Imputation methods, enhanced metadata capture, and structured annotation protocols have strengthened reproducibility, but credible inference still depends on traceable, well-curated datasets. Methodological progress in *FIT* is therefore best understood as a process of standardisation and integration—linking empirical fieldcraft, participatory science, and open analytical infrastructure.

In essence, this work demonstrates that the lasting value of the *FIT* lies in uniting domains once separate. By formalising tracker expertise within transparent computational workflows, it bridges field observation, quantitative analysis, and ethical conservation. The resulting framework represents a new generation of integrative biodiversity monitoring, combining reproducibility, accessibility, and respect for the species it studies.



## Conclusion and Outlook

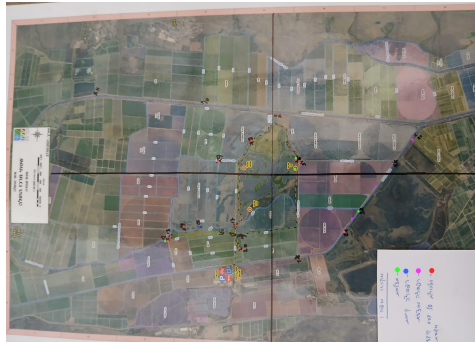
This dissertation establishes the *FIT* as a validated, scalable framework for reliable species-, sex-, and individual-level identification in Eurasian otters (*Lutra lutra*), providing a solid foundation for future large-scale applications.

### 8.1 Future Work and Outlook

#### Validation and Generalisation

Future work will focus on rigorous field validation of the *FIT* alongside established non-invasive methods—particularly microsatellite *DNA*—to enable systematic cross-validation under real-world conditions. Once validated, the *FIT* could partly replace genetic approaches due to its lower cost, reduced dependence on sample quality, and comparable accuracy.

Collaborations are underway with the *IUCN OSG* and *IOSF*. A joint study with Assaf Ben David and colleagues at Tel-Hai College exemplifies this approach: in the Hula Valley, the *FIT* is integrated with microsatellite genotyping using a floating sand-rack system for standardised, low-disturbance footprint collection (Fig. 8.1). The racks reliably attract otters and yield high-quality impressions, enabling site-matched comparisons with genetic assignments and providing a replicable model for multi-site validation.



(a) Study-area map with rack sites.



(b) Rack in wetland setting.



(c) Tracks on the rack surface, with blue arrows marking prints attributable to Eurasian otter (*Lutra lutra*).

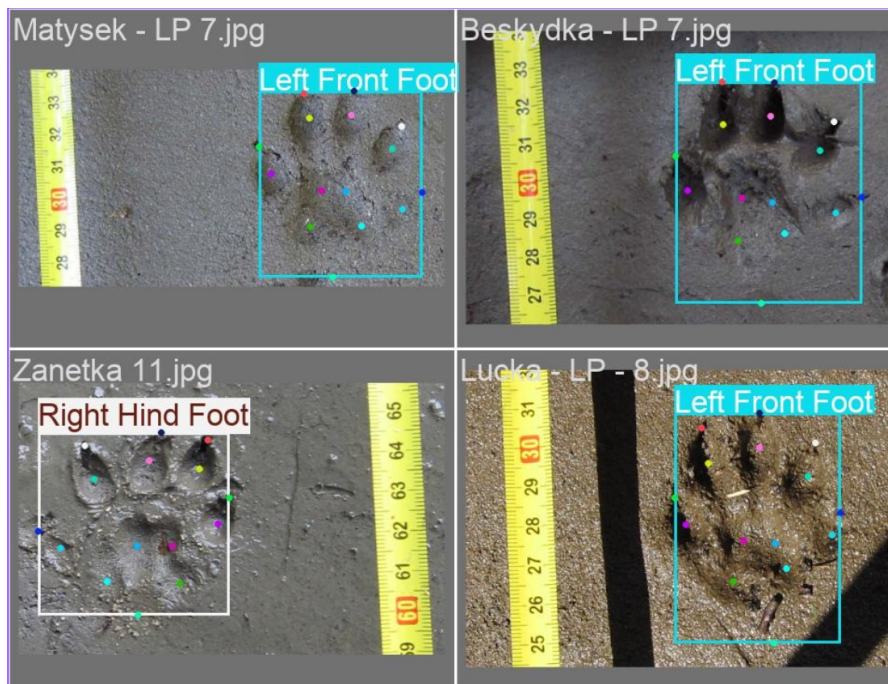
**Fig. 8.1.:** Floating sand rack system for standardised otter track collection in the Hula Valley: (a) map of deployment sites, (b) rack in wetland habitat, and (c) detailed view of track impressions.

## Automation and Deep Learning

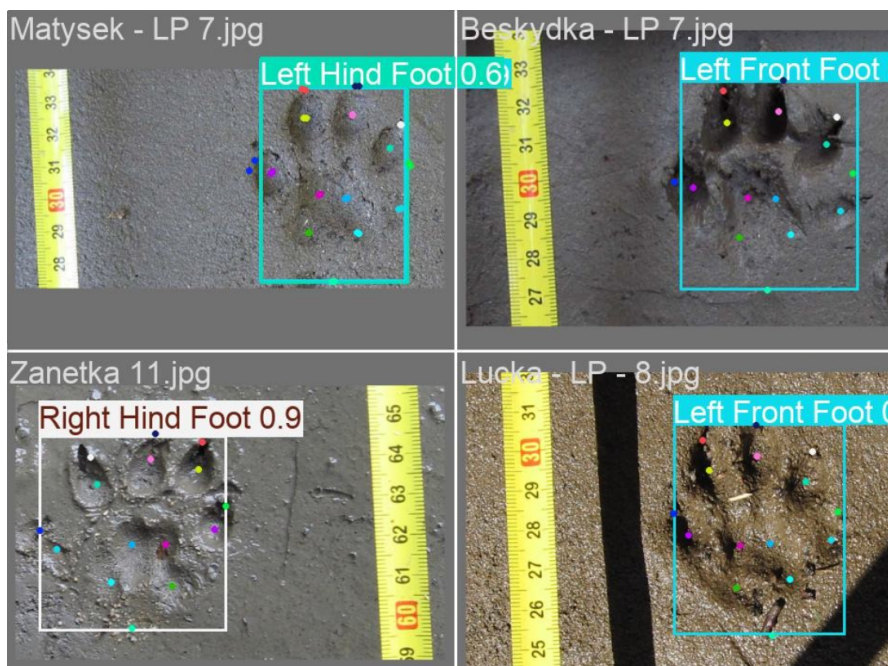
This work was among the first to apply *DL* to footprint-based analysis [49]. Building on these results, we piloted automated landmark detection using a *You Only Look Once (object detection architecture) (YOLO)v11*-based model (Fig. 8.2; [102]).

Recent findings by Petso et al. [100] independently support this concept: lightweight *AI* architectures that incorporate landmark geometry—such as *YOLOv8 nano* or *MobileNet convolutional neural network architecture (lightweight model optimised for mobile and embedded vision applications) (MobileNet)*—achieved near-perfect species classification of African ungulates and confirmed that geometric cues enhance detection precision and generalisability. These results validate the approach proposed here, highlighting landmark-informed *DL* as a scalable and cost-efficient pathway for non-invasive monitoring.

Future work will focus on production-grade annotation pipelines with assisted labelling, model-in-the-loop feedback, and active-learning strategies to prioritise informative samples. Cross-species pretraining and few-shot adaptation can further reduce data demands, while calibrated uncertainty measures will guide expert review and ensure consistent annotation quality.



(a) Ground-truth landmark annotations on otter footprints.



(b) Predicted landmark positions from the *YOLOv11*-based model.

**Fig. 8.2.:** Examples of (a) ground-truth and (b) predicted landmarks on otter footprints from the validation set, demonstrating the feasibility of automated annotation assistance.

## Scalability and Implementation

Scaling the *FIT* requires an open, end-to-end workflow that unites image ingestion, curation, modelling, and reporting within a reproducible *Python* stack. The proposed architecture emphasises standardised metadata capture, batchable pipelines, built-in *QA/QC*, and exportable audit trails to ensure transparency and re-analysis.

Many of these elements—such as versioned annotations and controlled transformation metadata—are already implemented in the WildTrack AI cloud platform [14], which offers a secure environment for large-scale coordination.

Complementary offline variants in *Python* or *JMP* remain essential for field use, provided that they follow the same harmonised data schema and *FAIR* principles. Such interoperability will allow conservation teams to collect and analyse data consistently, regardless of connectivity or platform.

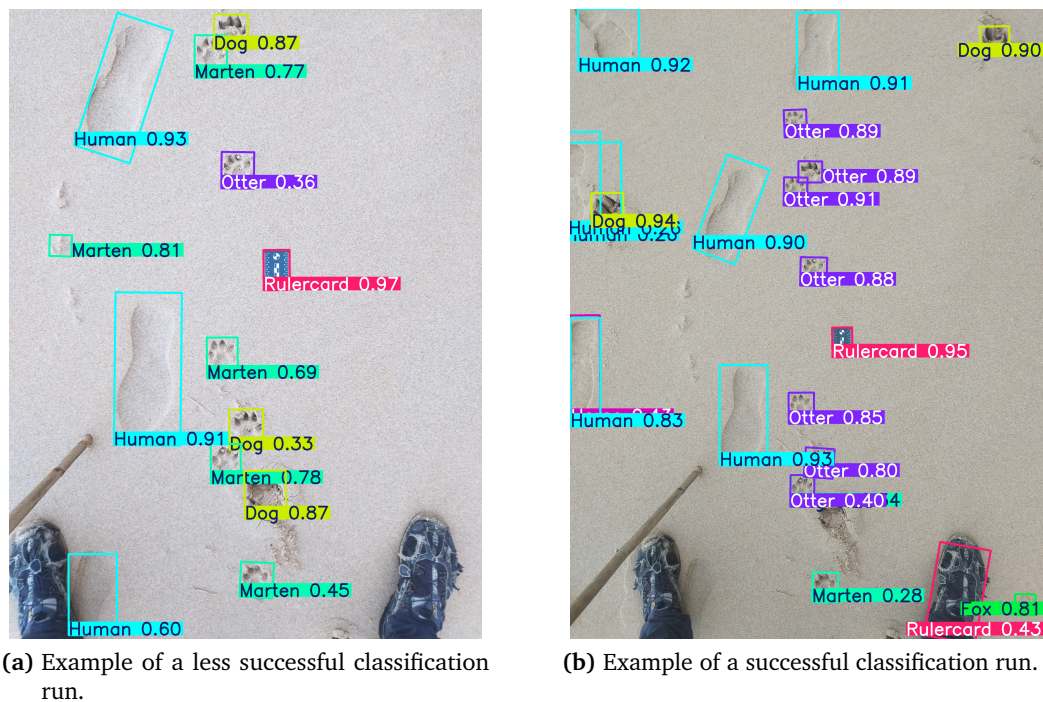
## Integrating Tracking, Remote Sensing, and AI for Conservation

Future biodiversity monitoring will increasingly rely on multimodal integration—linking footprint data, bioacoustic recordings, and remote-sensing imagery within unified, non-invasive workflows. Robust multiclass detectors must adapt to varying scales, backgrounds, and class balances, combining ground-truthable spoor and sound data with environmental context to achieve both spatial coverage and ecological depth.

Ongoing pilots in Portugal demonstrate how field imagery captured at multiple heights and *GSD* can feed into such frameworks. The WildTrack AI app [12, 13] already standardises image capture and metadata, while hands-free or drone-based acquisition promises to expand coverage and consistency. Recent studies using *Unoccupied Aerial Vehicle (drone)s (UAVs)* and very-high-resolution satellite imagery [103, 104] show that automated wildlife detection at landscape scales is feasible when supported by rigorous *QC* and site-specific retraining.

The next frontier lies in connecting these layers through data fusion: combining acoustic, visual, and spatial observations to calibrate and validate large-scale *AI* models that infer animal presence, habitat use, and behavioural context. Reliable ground-truth data—produced through efficient, expert-friendly labelling tools—will be central to this process. Empowering trackers and field biologists to annotate multimodal data directly in the field ensures that local ecological expertise remains embedded in model training and evaluation.

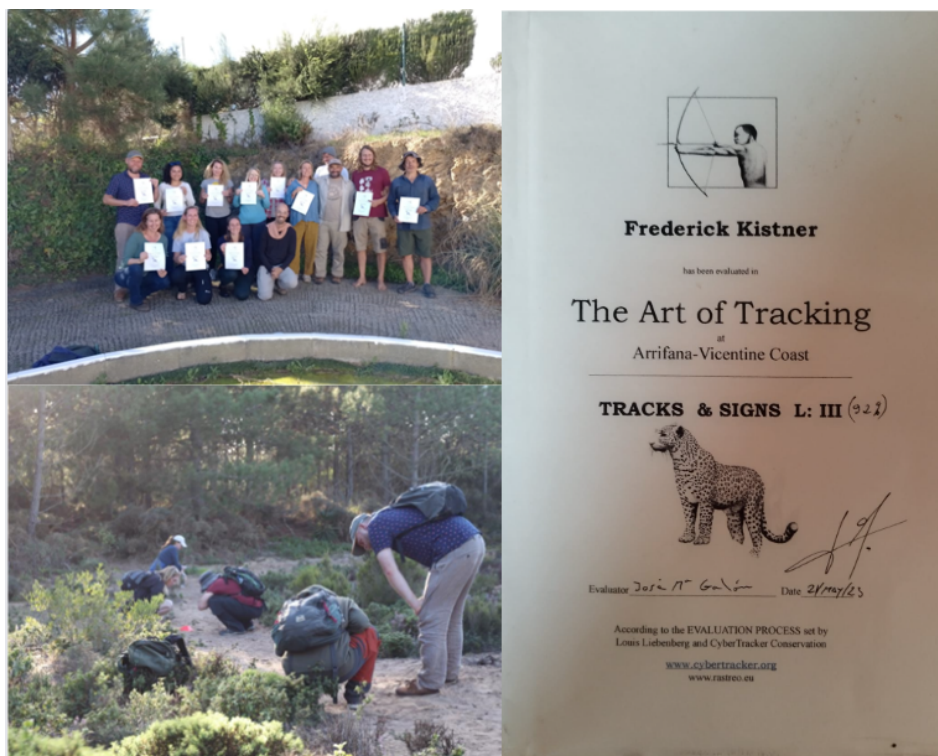
Such multimodal, non-invasive systems can transform biodiversity monitoring by merging field precision with continental-scale observability, enabling transparent, reproducible, and ethically grounded conservation analytics.



**Fig. 8.3.:** Predictions from a multiclass object detection model applied to coastal dune imagery, including human, otter, dog, marten, fox, and ruler card classes (bounding boxes colour-coded by class, with confidence scores shown). In both images, a clearly visible otter trail runs diagonally across the frame. While some prints were misclassified (e.g. as *Marten* or *Dog*), overall model performance remained high even under challenging field conditions.

## Knowledge Transfer and Capacity Building

Several workshops at the **CERES** Field Centre—developed as part of this dissertation—embodied the integration of field expertise with digital analytics (Fig. 8.4). Bringing together practitioners from multiple disciplines, they demonstrated how rigorous observation and inference can be combined through iterative cycles of hypothesis generation and verification. Under the guidance of José Galán, participants experienced how disciplined field practice meets scientific reasoning in addressing the tracker’s classical questions: whether a trace belongs to an animal, which species it represents, what event took place, when and where it occurred, and why the animal was present at that site. These structured enquiries turned observation into a systematic act of reasoning—illustrating how the **FIT** approach bridges intuitive field interpretation and analytical verification, and how even subtle traces can be transformed into evidence through disciplined inference.



**Fig. 8.4.:** Workshop at the CERES Field Centre. Participants from multiple disciplines worked with experienced trackers to derive ecological information from footprints, trails, and sign, iterating between observation, inference, and verification.

## 8.2 Final Reflections – On Tracking and Knowing the World

### Final Reflections – On Tracking and Knowing the World

For me, tracking has been a decade-long journey of attentive learning—guided by patient teachers and countless hours in the field, walking slowly, observing carefully, visualising patterns, and taking time to think.

Recognising change is the first step to counteracting it; as Louis Liebenberg observes in *The Origin of Science*, tracking is an epistemic method—disciplined attention to traces, guided by evidence and imagination—where the tracker infers hidden causes, predicts what should appear next, and tests those claims through the same hypothetico–deductive cycle that underlies modern science [9].

A single track is a compressed record of behaviour—a doorway into the immediate past. With trained attention, a sequence of impressions becomes a narrative: where an animal slowed, scanned, or committed to a line. This is pattern recognition in the wild. Our ancestors’ success at reading signs—inferring hidden causes from noisy evidence—likely selected for the cognitive capacities that underpin modern science and, by extension, today’s *ML*. The ability to imagine what is not directly seen bridges footprint and hypothesis, landscape and model. My closing plea is simple. In an age of satellites, drones, and *AI*, we should still walk slowly and attend with all senses. Tracking is a multisensory practice that grounds us in the living world. It should be taught alongside natural history as a core competency in conservation and as a foundation of public scientific literacy. Tracking sharpens perception, improves data quality, grounds models in reality, and reconnects decision-making with place.

If this dissertation has one parting claim, it is this: pairing the human craft of reading the land with open, well-tested, scalable models allows us to grow insight without losing truth—turning signs in sand into sound decisions, and carrying those decisions back to the ground where conservation ultimately lives. Experienced trackers—including those without formal academic credentials—should be recognised as professionals; wherever feasible, structured collaborations with clear roles, fair compensation, and co-authorship when trackers make substantial contributions should be sought. This modest but practical plea—teach tracking as both science and art—raises standards, strengthens inference, and builds the trust needed for actionable, place-based conservation.



*At the end of the trail, you might find what you seek.*



## References

- [1]S. L. Pimm, C. N. Jenkins, R. Abell, et al. *The biodiversity of species and their rates of extinction, distribution, and protection*. In: *Science* 344.6187 (May 2014) (cit. on p. 3).
- [2]Secretariat of the Convention on Biological Diversity. *Report of the Conference of the Parties to the Convention on Biological Diversity on the second part of its fifteenth meeting*. CBD/COP/15/17. Convention on Biological Diversity (CBD), 2023 (cit. on p. 3).
- [3]H. D. Veilleux, M. D. Misutka, and C. N. Glover. *Environmental DNA and environmental RNA: Current and prospective applications for biological monitoring*. In: *The Science of the Total Environment* 782 (Aug. 2021), p. 146891 (cit. on p. 3).
- [4]P. Banerjee, G. Dey, C. M. Antognazza, et al. *Reinforcement of Environmental DNA Based Methods (Sensu Stricto) in Biodiversity Monitoring and Conservation: A Review*. In: *Biology* 10.12 (Nov. 2021), p. 1223 (cit. on p. 3).
- [5]S. Arzberger, A. Fairbairn, M. Hemauer, et al. *The potential of soundscapes as an ecosystem monitoring tool for urban biodiversity*. In: *Journal of Urban Ecology* 11.1 (2025) (cit. on p. 3).
- [6]M. A. Zemanova. *Towards more compassionate wildlife research through the 3Rs principles: moving from invasive to non-invasive methods*. In: *Wildlife Biology* 2020.1 (Mar. 2020), pp. 1–17 (cit. on p. 3).
- [7]Z. Jewell. *Effect of Monitoring Technique on Quality of Conservation Science*. In: *Conservation Biology* 27.3 (May 2013), pp. 501–508 (cit. on pp. 3, 14, 60).
- [8]T. P. Moorhouse, M. Gelling, and D. W. Macdonald. *Effects of habitat quality upon reintroduction success in water voles: Evidence from a replicated experiment*. In: *Biological Conservation* 142.1 (Jan. 2009), pp. 53–60 (cit. on pp. 3, 9).
- [9]L. Liebenberg. *The Origin of Science. The Evolutionary Roots of Scientific Reasoning and its Implications for Tracking Science*. 2nd ed. Second eBook Edition, Version 2.2; first eBook edition 2013. Cape Town, South Africa: CyberTracker, 2021 (cit. on pp. 3, 12, 100).
- [10]D. Fraisl, G. Hager, B. Bedessem, et al. *Citizen science in environmental and ecological sciences*. In: *Nature Reviews Methods Primers* 2.1 (2022), pp. 1–20 (cit. on p. 3).
- [11]D. L. Soteropoulos, C. R. De Bellis, and T. Witsell. *Citizen Science Contributions to Address Biodiversity Loss and Conservation Planning in a Rapidly Developing Region*. In: *Diversity* 13.6 (June 2021), p. 255 (cit. on p. 3).
- [12]WildTrack, Inc. *WildTrackAI Android App*. Mobile app (Android). Version 2.1. Android Application. WildTrack, June 18, 2025 (cit. on pp. 3, 11, 25, 97).

- [13] WildTrack, Inc. *WildTrackAI*. Mobile app (iOS/iPadOS). Version 2.1.0. iOS Application. WildTrack, June 18, 2025 (cit. on pp. 3, 11, 25, 97).
- [15] F. Danielsen, P. M. Jensen, N. D. Burgess, et al. *A Multicountry Assessment of Tropical Resource Monitoring by Local Communities*. In: *BioScience* 64.3 (2014), pp. 236–251 (cit. on p. 3).
- [16] C. T. Callaghan, I. Ozeroff, C. Hitchcock, and M. Chandler. *Capitalizing on opportunistic citizen science data to monitor urban biodiversity: A multi-taxa framework*. In: *Biological Conservation* 251 (2020), p. 108753 (cit. on p. 3).
- [17] S. L. Voigt-Heucke, M. Müller, and J. Rostin. *How Citizen Science Projects Contribute to Urban Biodiversity Monitoring and Conservation Frameworks—A German Case Study*. In: *Citizen Science: Theory and Practice* 8.1 (2023), pp. 1–13 (cit. on p. 3).
- [18] Z. C. Jewell, S. K. Alibhai, and P. R. Law. *Censusing and monitoring black rhino (*Diceros bicornis*) using an objective spoor (footprint) identification technique*. In: *Journal of Zoology* 254.1 (May 2001), pp. 1–16 (cit. on pp. 4, 15, 51).
- [19] S. Alibhai, Z. C. Jewell, and P. Law. *A footprint technique to identify white rhino *Ceratotherium simum* at individual and species levels*. In: *Endangered Species Research* 4.1-2 (Jan. 2008), pp. 205–218 (cit. on pp. 4, 14, 15, 19, 21, 36, 37, 51, 55, 57–59, 87).
- [20] J. Gu, S. K. Alibhai, Z. C. Jewell, G. Jiang, and J. Ma. *Sex determination of Amur tigers (*Panthera tigris altaica*) from footprints in snow*. In: *Wildlife Society Bulletin* 38.3 (2014), pp. 495–502 (cit. on pp. 4, 36, 37).
- [21] Z. C. Jewell, S. K. Alibhai, F. Weise, et al. *Spotting Cheetahs: Identifying Individuals by Their Footprints*. In: *Journal of Visualized Experiments* 111 (2016) (cit. on pp. 4, 14, 15, 21, 24, 29, 30, 36, 37, 51, 55, 57–59, 87).
- [22] B. V. Li, S. Alibhai, Z. Jewell, D. Li, and H. Zhang. *Using footprints to identify and sex giant pandas*. In: *Biological Conservation* 218 (Feb. 2018), pp. 83–90 (cit. on pp. 4, 11, 14, 15, 36, 37, 51, 53, 56, 57, 60, 87).
- [23] D. O. Moreira, S. K. Alibhai, Z. C. Jewell, et al. *Determining the numbers of a landscape architect species (*Tapirus terrestris*), using footprints*. In: *PeerJ* 6 (2018), e4591 (cit. on pp. 4, 14, 15, 24, 57).
- [24] S. L. Pimm, S. Alibhai, R. Bergl, et al. *Emerging Technologies to Conserve Biodiversity*. In: *Trends in Ecology & Evolution* 30.11 (Nov. 2015), pp. 685–696 (cit. on pp. 4, 11, 21, 23, 25).
- [25] J. M. Tucker, C. King, R. Lekivetz, et al. *Development of a non-invasive method for species and sex identification of rare forest carnivores using footprint identification technology*. In: *Ecological Informatics* 79 (2024), p. 102431 (cit. on pp. 4, 14, 15, 19, 21, 60).
- [26] I. Weinberger. *Der Fischotter. Ein heimlicher Jäger kehrt zurück*. Ed. by H. Baumgartner. 1. Auflage. Haupt Natur. Literaturverzeichnis: Seite 243-253. Bern: Haupt Verlag, 2018, p. 256. 256 pp. (cit. on pp. 4, 5, 9–11, 16).

- [28]M. Sittenthaler, H. Bayerl, G. Unfer, R. Kuehn, and R. Parz-Gollner. *Impact of fish stocking on Eurasian otter (Lutra lutra) densities: A case study on two salmonid streams*. In: *Mammalian Biology* 80.2 (Mar. 2015), pp. 106–113 (cit. on pp. 4, 9).
- [29]C. Reuther, D. Dolch, R. Green, et al. *Surveying and monitoring distribution and population trends of the Eurasian otter (Lutra lutra): Guidelines and evaluation of the standard method for surveys as recommended by the European section of the IUCN/SSC Otter Specialist Group*. Ed. by C. Reuther. IUCN/SSC Otter Specialist Group guidelines; PDF available. Hankensbüttel, Germany, 2000 (cit. on pp. 4, 9, 11, 32).
- [30]S. Lampa, J.-B. Mihoub, B. Gruber, R. Klenke, and K. Henle. *Non-Invasive Genetic Mark-Recapture as a Means to Study Population Sizes and Marking Behaviour of the Elusive Eurasian Otter (Lutra lutra)*. In: *PLOS ONE* 10.5 (2015), e0125684 (cit. on pp. 4, 9, 10).
- [31]N. Duplaix and M. Savage. *The Global Otter Conservation Strategy*. Electronic and paperback versions. Four Corners Institute, Salem, Oregon, USA: IUCN/SSC Otter Specialist Group, 2018 (cit. on pp. 4, 5, 9).
- [32]P. Yoxon. *Otters of the world*. Ed. by G. M. Yoxon. Dunbeath: Whittles Publishing, 2014. 154 pp. (cit. on pp. 4, 11).
- [33]P. Chanin. *Monitoring the Otter Lutra lutra*. Technical report. Peterborough, UK, 2003 (cit. on p. 9).
- [34]H. Kruuk, D. N. Carss, J. W. H. Conroy, and L. Durbin. *Otter (Lutra lutra L.) numbers and fish productivity in rivers in north-east Scotland*. In: *Symposia of the Zoological Society of London* 67 (1995), pp. 171–191 (cit. on p. 9).
- [35]L. Quaglietta, B. H. Martins, A. de Jongh, A. Mira, and L. Boitani. *A Low-Cost GPS GSM/GPRS Telemetry System: Performance in Stationary Field Tests and Preliminary Data on Wild Otters (Lutra lutra)*. In: *PLoS ONE* 7.1 (Jan. 2012). Ed. by R. H. Clarke, e29235 (cit. on pp. 9, 10).
- [36]I. C. Weinberger, S. Muff, A. de Jongh, A. Kranz, and F. Bontadina. *Flexible habitat selection paves the way for a recovery of otter populations in the European Alps*. In: *Biological Conservation* 199 (2016), pp. 88–95 (cit. on pp. 9–11).
- [37]J. A. Shivik. *The other tools for wolf management*. In: *Biological Conservation* 111.2 (2001), pp. 159–169 (cit. on p. 9).
- [38]L. R. Harper, L. Lawson Handley, A. I. Carpenter, et al. *Environmental DNA (eDNA) metabarcoding of pond water as a tool to survey conservation and management priority mammals*. In: *Biological Conservation* 238 (Oct. 2019), p. 108225 (cit. on p. 9).
- [39]M. A. Findlay, R. A. Briers, and P. J. C. White. *Component processes of detection probability in camera-trap studies: understanding the occurrence of false-negatives*. In: *Mammal Research* 65.2 (Feb. 2020), pp. 167–180 (cit. on pp. 9, 10).
- [40]A. C. Burton, E. Neilson, D. Moreira, et al. *REVIEW: Wildlife camera trapping: a review and recommendations for linking surveys to ecological processes*. In: *Journal of Applied Ecology* 52.3 (2015), pp. 675–685. eprint: <https://besjournals.onlinelibrary.wiley.com/doi/pdf/10.1111/1365-2664.12432> (cit. on p. 10).

- [41] T. Petso, R. S. Jamisola, and D. Mpoeleng. *Review on methods used for wildlife species and individual identification*. In: *European Journal of Wildlife Research* 68.1 (Dec. 2021) (cit. on pp. 10–12, 25).
- [42] X. Han, Z. Dong, G. Zhao, et al. *Using surveillance cameras to analyze the activity patterns of Eurasian otters (*Lutra lutra*) and the efficiency of camera-trap monitoring*. In: *Biodiversity Science* 29.6 (2021), p. 770 (cit. on p. 10).
- [43] M. A. Findlay, R. A. Briers, R. P. Ingledew, and P. J. C. White. *An evidence-based approach to identifying resting sites of Eurasian otter *Lutra lutra* from camera-trap and field-sign data*. In: *Wildlife Biology* 2023.1 (Nov. 2022), e01036 (cit. on p. 10).
- [44] J. M. Kolowski and T. D. Forrester. *Camera trap placement and the potential for bias due to trails and other features*. In: *PLOS ONE* 12.10 (2017), e0186679 (cit. on p. 10).
- [45] P. Hájková, B. Zemanová, K. Roche, and B. Hájek. *An evaluation of field and noninvasive genetic methods for estimating Eurasian otter population size*. In: *Conservation Genetics* 10.6 (2009), pp. 1667–1681 (cit. on pp. 10, 12).
- [46] S. Sharma, W. Chee-Yoong, A. Kannan, et al. *Identification of three Asian otter species (*Aonyx cinereus*, *Lutra sumatrana*, and *Lutrogale perspicillata*) using a novel noninvasive PCR-RFLP analysis*. In: *Ecology and Evolution* 12.12 (Dec. 2022), pp. 1–17 (cit. on p. 10).
- [47] L. Mercier and G. Fried. *Preliminary study of the tracks of captive otters (*Lutra lutra*) as a tool for field research*. In: *IUCN Otter Specialist Group Bulletin* 21.2 (2004). Published online by IUCN Otter Specialist Group Bulletin, pp. 94–99 (cit. on pp. 11, 14, 60, 87).
- [48] F. Kistner, L. Slaney, Z. Jewell, A. Ben-David, and S. K. Alibhai. *It's Otterly Confusing – Distinguishing Between Footprints of Three of the Four Sympatric Asian Otter Species Using Morphometrics and Machine Learning*. In: *Otter: Journal of the International Otter Survival Fund* 8 (2022), pp. 108–125 (cit. on pp. 11, 14, 15, 17, 18, 30, 38, 42, 43, 53, 55–57, 60, 65, 67, 68, 87, 112).
- [49] F. Kistner, L. Slaney, and N. Morant. *Can You Tell the Species by a Footprint? Identifying Three of the Four Sympatric Southeast Asian Otter Species using Computer Vision and Deep Learning*. In: *IUCN Otter Specialist Group Bulletin* 40.4 (2023). PDF available via IUCN Otter Specialist Group; no DOI found, pp. 197–210 (cit. on pp. 11, 15, 18, 19, 23, 38, 42, 60, 63, 65, 88, 95).
- [50] J. Vetrovcova. *Identifying Individual Eurasian Otters (*Lutra lutra*) Based on Measurements of Their Footprints: Standardization of the Method and Its Potential for Censusing and Monitoring Wild Otter Populations*. 2006 (cit. on pp. 11, 14, 31, 37, 87).
- [52] S. Alibhai, Z. Jewell, and J. Evans. *The challenge of monitoring elusive large carnivores: An accurate and cost-effective tool to identify and sex pumas (*Puma concolor*) from footprints*. In: *PLOS ONE* 12.3 (Mar. 2017). Ed. by S. S. Romanach, e0172065 (cit. on pp. 11, 14, 15, 29, 36, 37, 57, 87).
- [53] S. Erlinge. *Home range of the otter *Lutra lutra* L. in Southern Sweden*. In: *Oikos* 18.2 (1967), pp. 186–209 (cit. on pp. 11, 12).

- [54]S. Erlinge. *Territoriality of the otter Lutra lutra L.* In: *Oikos* 19.1 (1968), pp. 81–98 (cit. on pp. 11, 12).
- [55]R. Sulkava. *Snow tracking: a relevant method for estimating otter Lutra lutra populations.* In: *Wildlife Biology* 13.2 (2007), pp. 208–218. eprint: <https://nsojournals.onlinelibrary.wiley.com/doi/pdf/10.2981/0909-6396%282007%2913%5B208%3ASTARMF%5D2.0.CO%3B2> (cit. on pp. 11, 12).
- [57]A. Dorfman, A. Subach, and I. Scharf. *Tracking spoor and field signs as a zoological research method.* In: *Animal Behaviour* 203 (2023), pp. 133–141 (cit. on p. 12).
- [58]V. Sidorovich and G. Lauzhel. *Territorial behaviour and spatial structure of the otter population in Byelorussia.* In: *Acta Theriologica* 37.1-2 (1992), pp. 153–162 (cit. on p. 12).
- [59]K. M. Slauson, R. L. Truex, and W. J. Zielinski. *Determining the Gender of American Martens and Fishers at Track Plate Stations.* In: *Northwest Science* 82.3 (2008), p. 185 (cit. on p. 12).
- [60]A. Zalewski. *Identifying sex and individuals of pine marten using snow track measurements.* In: *Wildlife Society Bulletin* 27.1 (1999). Accessed via JSTOR, pp. 28–31 (cit. on p. 14).
- [61]S. K. Alibhai, J. Gu, Z. C. Jewell, et al. *‘I know the tiger by his paw’: A non-invasive footprint identification technique for monitoring individual Amur tigers (Panthera tigris altaica) in snow.* In: *Ecological Informatics* 73. December 2022 (Mar. 2023), p. 101947 (cit. on pp. 14, 15, 21, 24, 36, 37, 56, 87).
- [62]*Footprint Recognition of Rodents and Insects.* Aotearoa / New Zealand, 2004, pp. 167–173 (cit. on p. 14).
- [63]S. W. Yiu, T. R. Etherington, and J. C. Russell. *Discriminating footprints to improve identification of congeneric invasive Rattus species.* In: *Pest Management Science* (2025). Early View (cit. on pp. 14, 15, 60).
- [64]Z. C. Jewell, S. Alibhai, P. R. Law, K. Uiseb, and S. Lee. *Monitoring rhinoceroses in Namibia’s private custodianship properties.* In: *PeerJ* 8 (Aug. 2020), e9670 (cit. on p. 14).
- [65]G. E. Finerty, N. Borrego, S. K. Alibhai, et al. *The crossroads of tradition and modern technology: integrative approaches to studying carnivores in low density ecosystems.* In: *Frontiers in Conservation Science* 5 (June 2024), p. 1402500 (cit. on pp. 14–16, 85).
- [66]K. Hertweck, R. Klenke, and K. Henle. *Estimating the density of otter Lutra lutra populations using individual analyses of tracks.* In: *IUCN Otter Specialist Group Bulletin* 19 (1998). PDF available via ResearchGate, pp. 118–122 (cit. on p. 14).
- [67]*OpenAnimalTracks: A Dataset for Animal Track Recognition.* IEEE, Oct. 2024, pp. 110–116 (cit. on pp. 15, 88).
- [68]F. Kistner, J. Tulowietzki, L. Slaney, et al. *Enhancing endangered species monitoring by lowering data entry requirements with imputation techniques as a preprocessing step for the footprint identification technology (FIT).* In: *Ecological Informatics* 82 (2024), p. 102676 (cit. on pp. 16, 19, 21, 24, 39, 43, 51, 55, 56, 61, 63, 65, 77, 78, 86).

- [69]J. Grolms. *Tierspuren Europas. Spuren und Zeichen bestimmen und interpretieren. Mit Spuren und Zeichen von Säugetieren, Vögeln, Reptilien, Amphibien und Wirbellosen.* Stuttgart: Verlag Eugen Ulmer, 2021 (cit. on pp. 16, 19, 35).
- [70]J. Rhyder. *Tracks and Signs: A Guide to the Field Signs of Mammals and Birds of the U.K.* The History Press Ltd, 2021 (cit. on pp. 16, 29).
- [72]K. Laity. *The Field Validation of Footprint Identification Technology on Free-roaming Cheetahs in Namibia.* Master's project. Accessed: 2024-05-14. 2015 (cit. on p. 19).
- [73]D. J. Stekhoven and P. Bühlmann. *MissForest—non-parametric missing value imputation for mixed-type data.* In: *Bioinformatics* 28.1 (Oct. 2011), pp. 112–118 (cit. on pp. 19, 63).
- [74]L. Doove, S. Van Buuren, and E. Dusseldorp. *Recursive partitioning for missing data imputation in the presence of interaction effects.* In: *Computational Statistics & Data Analysis* 72 (Apr. 2014), pp. 92–104 (cit. on p. 19).
- [75]B. Ramosaj, L. Amro, and M. Pauly. *A cautionary tale on using imputation methods for inference in matched-pairs design.* In: *Bioinformatics* 36.10 (2020), pp. 3099–3106 (cit. on p. 19).
- [76]D. M. Aanensen, D. M. Huntley, E. J. Feil, F. al-Own, and B. G. Spratt. *EpiCollect: Linking Smartphones to Web Applications for Epidemiology, Ecology and Community Data Collection.* In: *PLoS ONE* 4.9 (Sept. 2009). Ed. by S. I. Hay, e6968 (cit. on p. 21).
- [77]D. M. Aanensen, D. M. Huntley, M. Menegazzo, C. I. Powell, and B. G. Spratt. *EpiCollect+: linking smartphones to web applications for complex data collection projects.* In: *F1000Research* 3 (Aug. 2014), p. 199 (cit. on p. 21).
- [79]M. Chandler, L. See, K. Copas, et al. *Contribution of citizen science towards international biodiversity monitoring.* In: *Biological Conservation* 213 (2017), pp. 280–294 (cit. on p. 23).
- [80]B. G. Weinstein. *A computer vision for animal ecology.* In: *Journal of Animal Ecology* 87.3 (2018), pp. 533–545 (cit. on p. 23).
- [81]D. Tuia, B. Kellenberger, S. Beery, et al. *Perspectives in machine learning for wildlife conservation.* In: *Nature Communications* 13.1 (2022) (cit. on p. 23).
- [82]P. Chen, P. Swarup, W. M. Matkowski, et al. *A study on giant panda recognition based on images of a large proportion of captive pandas.* In: *Ecology and Evolution* 10.7 (Mar. 2020), pp. 3561–3573 (cit. on p. 23).
- [83]M. L. Borowiec, R. B. Dikow, P. B. Frandsen, et al. *Deep learning as a tool for ecology and evolution.* In: *Methods in Ecology and Evolution* 13.8 (2021), pp. 1640–1660 (cit. on p. 23).
- [84]C. Capinha, A. Ceia-Hasse, A. M. Kramer, and C. Meijer. *Deep learning for supervised classification of temporal data in ecology.* In: *Ecological Informatics* 61 (2021), p. 101252 (cit. on p. 23).

- [85]S. Brodie, S. Allen-Ankins, M. Towsey, P. Roe, and L. Schwarzkopf. *Automated species identification of frog choruses in environmental recordings using acoustic indices*. In: *Ecological Indicators* 119 (2020), p. 106852 (cit. on p. 23).
- [86]K. I. Peña Carrillo, M. C. Lorenzi, M. Brault, et al. *A new putative species in the *Ectatomma ruidum* complex (Formicidae: Ectatomminae) produces a species-specific distress call*. In: *Bioacoustics* 31.3 (June 2021), pp. 332–347 (cit. on p. 23).
- [87]M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al. *The FAIR Guiding Principles for scientific data management and stewardship*. In: *Scientific Data* 3 (2016), p. 160018 (cit. on p. 24).
- [88]F. Pedregosa, G. Varoquaux, A. Gramfort, et al. *Scikit-learn: Machine Learning in Python*. In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830 (cit. on pp. 24, 55, 60, 62).
- [89]L. Sasse, E. Nicolaisen-Sobesky, J. Dukart, et al. *On Leakage in Machine Learning Pipelines*. In: (2023). arXiv: 2311.04179 [cs.LG] (cit. on pp. 24, 57, 58).
- [90]15th International Otter Congress, Sospel, France, 19.-23.09.2022. *15th IUCN/SSC OSG International Otter Congress “Communities and Otters”: Book of abstracts*. en. 2022 (cit. on p. 33).
- [92]L. McInnes, J. Healy, and J. Melville. *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. In: *arXiv preprint arXiv:1802.03426* (2018) (cit. on p. 45).
- [94]R. Tibshirani. *Regression Shrinkage and Selection Via the Lasso*. In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 58.1 (Jan. 1996), pp. 267–288 (cit. on p. 57).
- [95]*Stacking With Auxiliary Features*. International Joint Conferences on Artificial Intelligence Organization, 2017, pp. 2634–2640 (cit. on p. 57).
- [96]M. M. Deza and E. Deza. *Encyclopedia of Distances*. 4th ed. Hardcover; 778 pages. Berlin, Heidelberg: Springer Berlin Heidelberg, 2016 (cit. on p. 58).
- [97]*XGBoost: A Scalable Tree Boosting System*. KDD '16. ACM. ACM, Aug. 2016, pp. 785–794 (cit. on p. 60).
- [98]J. Snoek, H. Larochelle, and R. P. Adams. *Practical Bayesian Optimization of Machine Learning Algorithms*. In: *arXiv* (2012). eprint: 1206.2944 (cit. on p. 60).
- [99]J. Eisner, ed. *V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure*. Prague, Czech Republic: Association for Computational Linguistics, June 2007, pp. 410–420 (cit. on p. 61).
- [100]T. Petso, R. S. Jamisola Jr., S. Alibhai, et al. *An investigation of lightweight AI models to classify African ungulate species from tracks*. In: *Ecological Informatics* 92 (2025). Released September 2025, confirming the feasibility of deep-learning models for footprint classification., p. 103393 (cit. on pp. 86, 89, 95).
- [101]A. Tharwat and W. Schenck. *A Survey on Active Learning: State-of-the-Art, Practical Challenges and Research Directions*. In: *Mathematics* 11.4 (2023), p. 820 (cit. on p. 87).

- [102]R. Khanam and M. Hussain. *YOLOv11: An Overview of the Key Architectural Enhancements*. 2024. arXiv: 2410.17725 [cs.CV] (cit. on p. 95).
- [103]Z. Wu, C. Zhang, X. Gu, et al. *Deep learning enables satellite-based monitoring of large populations of terrestrial mammals across heterogeneous landscape*. In: *Nature Communications* 14.1 (May 2023), p. 3072 (cit. on p. 97).
- [104]P. T. Fretwell, M. A. LaRue, P. Morin, et al. *An Emperor Penguin Population Estimate: The First Global, Synoptic Survey of a Species from Space*. In: *PLOS ONE* 7.4 (2012), e33751 (cit. on p. 97).

## Online

- [14] *WildTrack AI Platform*. Cloud-based footprint collection, curation, review, and monitoring interface for non-invasive wildlife conservation. WildTrack. 2025. URL: <https://ai.wildtrack.org/app/dashboard> (visited on Oct. 30, 2025) (cit. on pp. 3, 21, 86, 97).
- [27] F. u. T. Bayerisches Staatsministerium für Ernährung Landwirtschaft. *Verordnung zur leichteren Fischotter-Entnahme tritt in Kraft*. July 2023. URL: <https://www.stmelf.bayern.de/service/presse/pm/2023/verordnung-zur-leichteren-fischotter-entnahme/index.html> (visited on Oct. 9, 2025) (cit. on p. 4).
- [51] F. Kistner. *Otterly great work! – FIT for the European Fish Otter (Lutra lutra)*. Blogpost describing the author’s early adaptation of FIT for European otters. Mar. 2016. URL: <https://www.wildtrack.org/blog/Otterly-great-work%21> (visited on Oct. 9, 2025) (cit. on pp. 11, 14, 16).
- [56] *Otter Spotter App*. Aktion Fischotterschutz e. V.; OTTER-ZENTRUM Hankensbüttel; Deutsche Fischotter Stiftung. URL: <https://www.otterspotter.de/otter-spotter-app> (visited on Oct. 26, 2025) (cit. on p. 11).
- [71] WildTrack. *WildTrack Specialist Group (WSG)*. Accessed 28 July 2025. 2023. URL: <https://www.wildtrack.org/about/wsg> (visited on July 28, 2025) (cit. on p. 18).
- [78] C. for Genomic Pathogen Surveillance. *EpiCollect5: Mobile & Web Application for free and easy data collection*. Ed. by C. for Genomic Pathogen Surveillance. Free and easy-to-use mobile data-gathering platform. 2025. URL: <https://five.epicollect.net> (visited on Oct. 10, 2025) (cit. on p. 21).
- [91] CERES International Project. *CERES International Project – Field Study Centre Aljezur, Portugal*. Accessed: 2025-11-01. 2025. URL: <https://www.ceresinternational.org/> (cit. on p. 35).
- [93] scikit-learn developers. *Feature importances with a forest of trees*. en. 2025. URL: [https://scikit-learn.org/stable/auto\\_examples/ensemble/plot\\_forest\\_importances.html](https://scikit-learn.org/stable/auto_examples/ensemble/plot_forest_importances.html) (visited on Oct. 10, 2025) (cit. on p. 57).

# List of Figures

2.1	Otter resting site imprint in sand . . . . .	13
2.2	Reference drawings of Eurasian otter feet . . . . .	17
2.3	Model insights supporting explainable FIT applications (adapted from [49]). (A) Right front footprint of an Asian small-clawed otter ( <i>Aonyx cinereus</i> ); (B) visualisation of a saliency heatmap indicating the regions emphasised by a convolutional neural network (CNN) during species classification; (C) overlay of (A) and (B). . . . .	18
2.4	Examples of footprints of varying quality . . . . .	20
2.5	WildTrackAI mobile application dashboard . . . . .	22
3.1	Sand-bed installations in otter enclosures . . . . .	31
3.2	Prepared bridge site for field footprint sampling . . . . .	32
3.3	DNA-validated otter prints under field conditions . . . . .	33
3.4	Visual examples of mammal tracks collected in the field . . . . .	34
3.5	Otter trail with characteristic 3×4 lope gait and occluded prints . . . . .	35
3.6	Example footprints from three otter species in Kistner et al. [48] . . . . .	38
4.1	Anatomical labelling of otter footprints . . . . .	42
4.2	Geometric profile from <i>FIT</i> annotation . . . . .	43
4.3	Correlation heatmaps . . . . .	44
4.4	Top features for sex and individual ID . . . . .	46
4.5	UMAP embedding by sex and data origin . . . . .	47
5.1	Pipelines for the three targets— <i>Sex</i> (left), <i>Individual</i> (middle), and <i>Population</i> (right)—constructed as stacked matrices: a header row, shared data preparation (grouped/stratified; grouped CV), and a subsequent 3×N block of steps. Optional steps are shown with dashed borders. . . . .	52
5.2	Dataset split statistics for <i>Lutra lutra</i> . . . . .	54
5.3	Historical confidence-circle rule and Reference Centroid Value-anchored pairwise embedding . . . . .	59
6.1	Tracker-based footprint characteristics of three Asian otter species . . . . .	66
6.2	Species classification (FIT): test-set confusion matrix . . . . .	67

6.3	Species classification (FIT): top features (XGBoost gain) . . . . .	68
6.4	Sex classification (Eurasian otter): <i>CV</i> and test confusion matrices . . .	69
6.5	Sex classification: top-25 feature importances (XGBoost gain) . . . . .	70
6.6	Prediction confidence (validation folds): single prints, errors, trail/in- dividual aggregation . . . . .	71
6.7	Prediction confidence (independent test set): single prints, errors, trail/individual aggregation . . . . .	72
6.8	Eurasian otter individual ID: <i>CV</i> and test Same/Diff confusion matrices	73
6.9	Population estimation (hierarchical clustering): <i>CV</i> cutoff selection . .	74
6.10	Population estimation (hierarchical clustering): sequential holdout evaluation . . . . .	75
8.1	Hula Valley floating sand rack monitoring setup . . . . .	94
8.2	Automated landmark detection in otter footprints . . . . .	96
8.3	Multiclass footprint detection examples on coastal dune images . . . .	98
8.4	Workshop at the CERES Field Centre. Participants from multiple dis- ciplines worked with experienced trackers to derive ecological infor- mation from footprints, trails, and sign, iterating between observation, inference, and verification. . . . .	99

# List of Tables

2.1	Summary of current monitoring methods in wildlife monitoring with special focus on Eurasian otters and non-invasive approaches ( <i>Lutra lutra</i> ), ordered by increasing scalability. . . . .	10
3.1	Detailed breakdown of the Eurasian otter ( <i>Lutra lutra</i> ) dataset by data origin. . . . .	37
3.2	Geometric footprint profiles used in this study . . . . .	37
6.1	Benchmarking across species . . . . .	76
6.2	Improvements to <i>FIT</i> : feature selection and sex stacking . . . . .	77
6.3	Improvements to <i>FIT</i> : imputation performance . . . . .	78
7.1	Comparison of FIT implementations for conservation monitoring: <i>JMP</i> vs. <i>Python</i> package. . . . .	90

# Appendix

The digital appendix contains a declaration of AI use as well as all peer-reviewed publications. The printed version does not include the publications.

## List of Presentations

### IUCN/OSG Otter Congresses and Associated Events

- **Kistner, F.** (2025). *From Erlinge to AI: How novel technologies are revitalizing the ancient art of otter tracking*. Oral presentation at the 16th International Otter Congress (IUCN SSC OSG), Lima, Peru, 26 February 2025. <https://www.otterspecialistgroup.org/osg-newsite/wp-content/uploads/2025/03/16th-IOC-Book-of-Abstracts.pdf>
- **Kistner, F., Slaney, L., & Morant, N.** (2022). *It's Otterly Confusing – Distinguishing Between Footprints of Three of the Four Sympatric Asian Otter Species Using Morphometrics and Machine Learning*. Poster and oral presentation at the 15th IUCN/SSC Otter Congress “Communities and Otters”, Sospel, France, 19–23 September 2022. <https://www.otterspecialistgroup.org/osg-newsite/15th-ioc/>

## Workshops and Training Sessions

- **Kistner, F.** & Slaney, L. (2022). *WildTrack Footprint Identification Technology (FIT) Training Session*. Presentation and practical session at the Malaysia Otter Workshop 2022, Kuala Selangor Nature Park, Malaysia (hybrid workshop), 25–29 April 2022. <https://www.otter.org/s/Malaysia.pdf>
- **Kistner, F.** (2021). *WildTrack: Footprint Identification Technology for Monitoring Eurasian Otters*. Invited presentation at the Eurasian Otter Workshop 2021 (virtual), 26–28 February 2021. <https://www.otterspecialistgroup.org/osg-newsite/eurasian-otter-workshop/>
- **Kistner, F.** (2020). *Presentations at WildTrack Workshop 2020*. WildTrack FIT Workshop, Cary, North Carolina, USA, 24–25 February 2020. <https://www.wildtrack.org/blog/Our-first-USA-FIT-Workshop%21>

## Other Specialist Conferences and Seminars

- **Kistner, F.** (2022). *WildTrack Footprint Identification Technology for Wildlife Monitoring*. Invited 45-minute live online presentation at the 5th Israeli Tracking Conference, Agamon Hula, Israel, 1 July 2022.
- **Kistner, F.**, Alibhai, S., & Jewell, Z. (2022). *It's Otterly Confusing! Short-Clawed, Hairy-Nosed, Smooth-Coated or Eurasian? Just ask JMP!*. Presentation at JMP Discovery Summit Europe 2022 (virtual), March 2022.
- **Kistner, F.** (2021). *Footprint Identification Technology (FIT) for Monitoring Eurasian Otters*. Invited talk at the webinar “Otterschutz in Deutschland” (Aktion Fischotterschutz e.V.), 10–11 June 2021. [https://aktion-fischotterschutz.de/fileadmin/user\\_upload/aktion\\_fischotterschutz/desktop/hauptnavigation/Unsere\\_Projekte/Tierforschung/Tagungen/2021\\_-\\_Webinar\\_\\_Otterschutz\\_in\\_Dtschl/Kistner.pdf](https://aktion-fischotterschutz.de/fileadmin/user_upload/aktion_fischotterschutz/desktop/hauptnavigation/Unsere_Projekte/Tierforschung/Tagungen/2021_-_Webinar__Otterschutz_in_Dtschl/Kistner.pdf)
- Slaney, L. & **Kistner, F.** (2023). *It's Otterly Confusing!*. Presentation at the Second International Tracking Symposium (Europe), 18–19 February 2023. <https://seratbushcraft.com/2023/02/>

## Use of Artificial Intelligence

Artificial intelligence tools were employed throughout this dissertation to support, not replace, the author's own scholarly work. Various AI systems—including OpenAI's ChatGPT, GitHub Copilot, Claude, Consensus, and Gemini—were used selectively to assist with text optimisation, code debugging, literature searches, and the refinement of methodological descriptions. AI was also used as a critical discussion partner to test arguments and improve conceptual clarity.

All outputs generated by these tools were actively reviewed, edited, and integrated under the author's direct supervision. The intellectual design, data analysis, interpretation, and final writing decisions were entirely guided and executed by the author. The use of AI thus served as an auxiliary instrument within a transparent, researcher-led workflow, ensuring that full responsibility for the content and conclusions of this dissertation remains with the author.

# IT'S OTTERLY CONFUSING – DISTINGUISHING BETWEEN FOOTPRINTS OF THREE OF THE FOUR SYMPATRIC ASIAN OTTER SPECIES USING MORPHOMETRICS AND MACHINE LEARNING

F. KISTNER<sup>\*a,b</sup>, L. SLANEY<sup>b</sup>, Z. JEWELL<sup>b,c</sup>, A. BEN-DAVID<sup>b,d</sup> and S. ALIBHAI<sup>b,c</sup>

<sup>a</sup>Institute of Photogrammetry and Remote Sensing (IPF), Institute for Technology Karlsruhe (KIT), Germany

<sup>b</sup>WildTrack Specialist Group

<sup>c</sup>Nicholas School of the Environment, Duke University, NC, USA

<sup>d</sup>School of Zoology, the George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

\*frederick.kistner@kit.edu

## ABSTRACT

*Southeast Asia is home to four sympatric otter species: Eurasian otter (*Lutra lutra*), Asian small-clawed otter (*Aonyx cinereus*), smooth-coated otter (*Lutrogale perspicillata*) and hairy-nosed otter (*Lutra sumatrana*). All species are on the IUCN's Red List of Endangered Species. In many regions, there is an overlap in the distributions of at least two species. To establish population size and range of these elusive animals, a variety of non-invasive data collection methods is available. Footprints often tend to be overlooked as valuable scientific data, which potentially leads to a wealth of information being lost.*

*The Footprint Identification Technology (FIT), developed by WildTrack, analyses morphometric data extracted from digital images of footprints and uses state-of-the-art Machine Learning models for classification tasks. In this study, we aimed to develop algorithms to distinguish between three of the four otter species of southeast Asia (*Lutra lutra*, *Aonyx cinereus* and *Lutrogale perspicillata*), using footprints.*

*For the three species, a digital footprint image database of known otters was developed with the help of several zoos. Using specific features, landmark points were placed on each footprint image and morphometric measurements in the form of distances, angles and areas were extracted and analysed in JMP software. The average classification accuracy for discriminating between the three species, using multiple analytical methods, was 97%. It is planned to develop this technology further adding the fourth species (*Lutra sumatrana*) and aim for sex and individual classification, for which more footprint data are necessary. The authors welcome contributions of footprint images from known animals.*

**Keywords:** *Eurasian otter*, *Lutra lutra*, *Asian small-clawed otter*, *Aonyx cinereus*, *smooth-coated otter*, *Lutrogale perspicillata*, *hairy-nosed otter*, *Lutra sumatrana*, *Footprint Identification Technology*, *Footprint*, *Track*, *XGBoost*

## INTRODUCTION

The development of effective conservation strategies requires reliable information on species distribution, population sizes and composition (Conde et al., 2019). These data, however, are often hard to get, especially for elusive and predominantly aquatic animals like otters.

There are four otter species with an overlapping distribution range in Asia: the Eurasian otter (*Lutra lutra*), the Asian small-clawed otter (*Aonyx cinereus*), the smooth-coated otter (*Lutrogale perspicillata*) and the hairy-nosed otter (*Lutra sumatrana*). Each species has an IUCN protection status, with the hairy-nosed otter being the most endangered species of the four (Duplaix and Savage, 2018).

There are several regions in southeast Asia where at least two of the four species are sympatric (Yoxon and Yoxon, 2017). It can be difficult to distinguish between the species visually and it is therefore critical to have reliable monitoring tools available. Wherever possible, non-invasive methods should be deployed to minimise disturbance and potential harm to the animals whilst ensuring the reliability and quality of data collected (Jewell, 2013).

Traditional tracking of animals can be seen as a non-invasive approach. Animal tracks can be collected without any disturbance to otter populations and used to inform on numbers and distribution. Data based on animal tracks is also reliable and cost-effective if the tracker has the required skills (Evans et al., 2009).

Liebenberg (2021) called tracking 'the origin of science' and with the vast development in scientific data collection and computer-assisted analysis (Petso et al., 2022), morphometric footprint analysis opens up new possibilities in wildlife monitoring. Another great advantage in using tracks as a data source for wildlife monitoring, is that indigenous communities with traditional ecological knowledge (TEK) can be involved in wildlife conservation research. Such collaborations can be of benefit to everyone involved and add valuable data for wildlife monitoring (Vieira et al., 2015; Ramos et al., 2016). In order to use tracks as a reliable data source, a strict protocol in data collection and analysis is crucial.

The Footprint Identification Technology (FIT) developed by WildTrack (www.wildtrack.org), combines traditional tracking with modern Machine Learning classification algorithms (Li et al., 2018). Following a strict protocol, it allows researchers, who might not be expert trackers, to draw dependable and unbiased information. FIT has been successfully developed for a wide range of endangered species and can classify species (Alibhai et al., 2008), sex (Alibhai et al., 2017), age

class (Li et al., 2018) and individual ID (Alibhai et al., 2008; Jewell et al., 2016; Alibhai et al., 2017; Li et al., 2018) once species-specific algorithms have been developed.

FIT classification accuracies are typically over 90% (Petso et al., 2022). In order to build such models, a reference database comprising images of known individuals, sex and/or species of a sufficient size and quality is required. Such databases are often created with the help of ex-situ organisations.

In this pilot study, we investigated the use of FIT to predict three Asian otter species based on morphometric features of their footprints. We deployed two different prediction models, Stepwise Linear Discriminant Analysis (LDA) and XGBoost.

LDA has often been the method of choice in comparable footprint studies of other species (Sharma et al., 2005; Alibhai et al., 2008; Alibhai et al., 2017; Li et al., 2018).

XGBoost is a supervised Machine Learning technique library that uses scalable, gradient boosting decision trees. It is often referred to as the best algorithm for classification and regression tasks of tabular data and has been the most successful method for the Machine Learning competition site “Kaggle” (Chen and Guestrin, 2016).

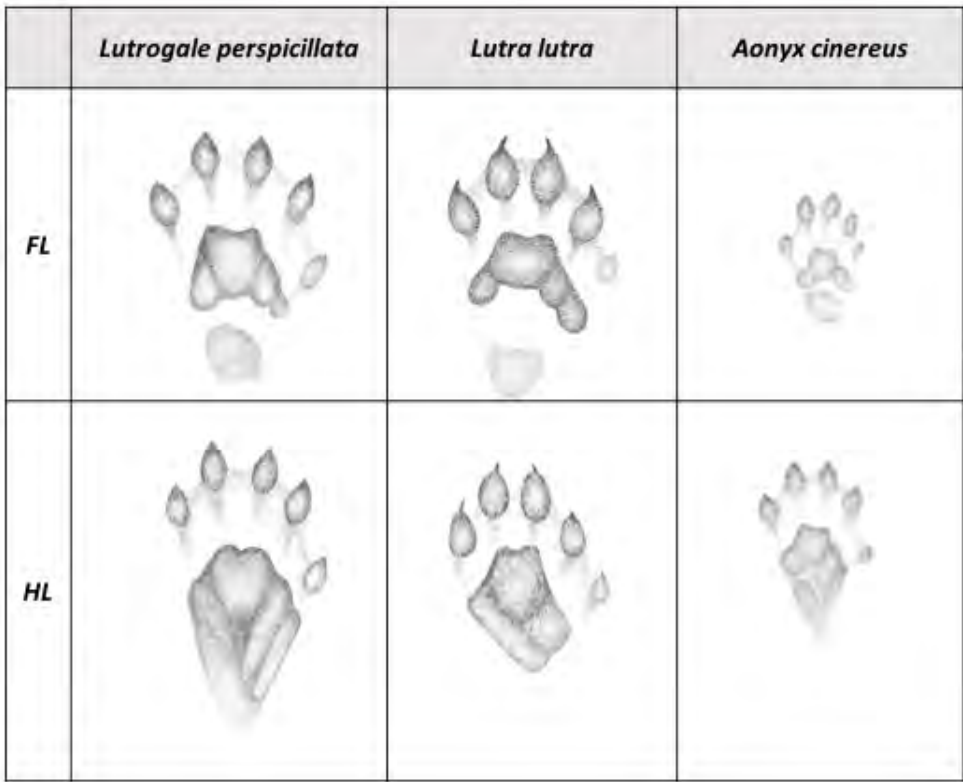
To the best of our knowledge, this is the first study that implements Machine Learning algorithms to classify otter species using biometric measurements from footprints as input data. In this report, we want to highlight the possibility of developing unbiased models with high accuracy. In addition, we investigate which features are the most important for these models in order to share insights of important areas of a footprint when species classification in situ is desired.

We would also like to encourage researchers to implement this monitoring tool in the field and help improve the classification capability and robustness of this method by contributing otter tracks to our otter footprint database.

## METHODS

### *Technical drawings*

Figure 1 shows technical drawings of otter tracks of the three species investigated in this study. The drawings are based on footprint images received for this research as well as photos of the underside of the three species' feet. Track drawings give the tracker as much detail as possible to interpret a track found in the field, but it is rare to find a track with all the components shown in the drawings. Tracks can vary in size, shape and detail depending on substrate and gait of the animal (Elbroch, 2019). As a result, certain parts of the otter's feet often do not register well in the track and the drawings reflect this by showing these parts in light grey.



**Figure 1. Technical drawings of tracks of three otter species. FL = Front Left; HL = Hind Left; Illustrations and copyright by Asaf Ben-David.**

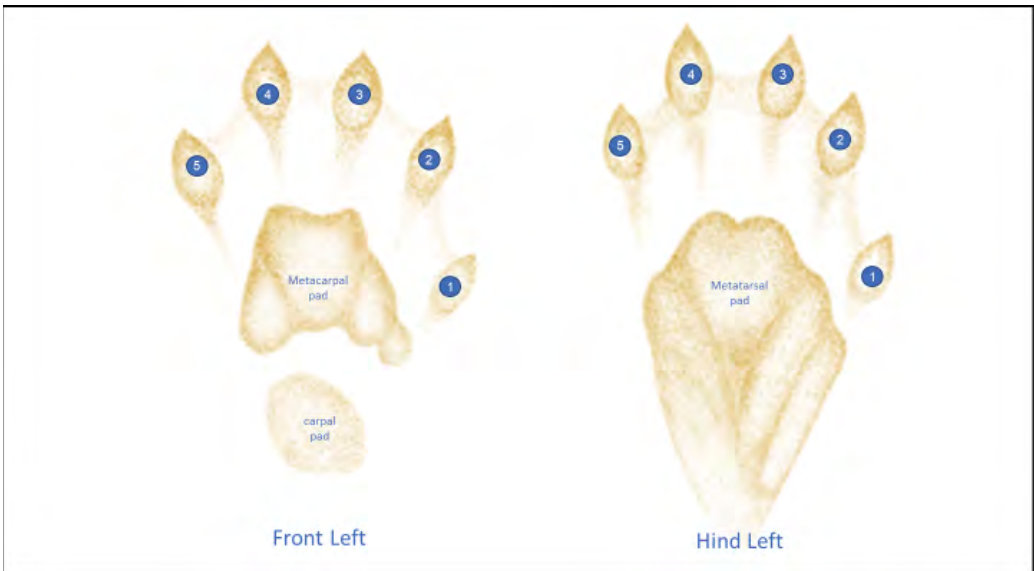
Description of the tracks by species are as follows:

- Smooth-coated otter (*Lutrogale perspicillata*): - The general impression is of a large, elongated otter track. The claws are short and blunt in comparison to the Eurasian otter. The toes are oval, and the bones between the toes and metacarpal or metatarsal pads are long and register clearly. Toe 1 (thumb) is almost straight below toe 2. The negative space between the toes and the metacarpal or metatarsal pad is 1.5 - 2 times the toe pad size. The metacarpal pad is wide and has a clear saddle in the centre top of the pad. The metatarsal pad is almost twice the size of the metacarpal pad.
- Eurasian otter (*Lutra lutra*) - The front track is less asymmetrical than the hind track and not as elongated as for smooth-coated and Asian small-clawed otters. The claw marks are small, sharp, and clear. The toe marks are large with a clear teardrop shape. The front foot's carpal pad is rounded and out of the three species it shows the most negative space to the metacarpal pad.
- Asian small-clawed otter (*Aonyx cinereus*) – This is the smallest track of the three species. The claws are very short and therefore mostly absent from the track. The toes are oval. The front foot is elongated and the carpal pad has a distinct angle towards toe 5.

Generally, the difference between these three otter species is more apparent in the front tracks than in the hind.

It was important to learn as much as possible about the footprint anatomy of these three otter species to decide on the landmark point placement for feature extraction. It was further important to compare these perfect anatomical diagrams with real footprint images to decide which footprint images were of sufficient quality for the study.

Otters leave typical mustelid tracks, which show five toes, a segmented metacarpal (palm on the front foot) and metatarsal (palm on the hind foot) pad. The front feet also have a separate carpal (heel) pad while the hind feet are clearly elongated, but do not show the same separation of metatarsal and tarsal (heel) pads. The otter species in this study all have claws, although as its name suggests the claws on the Asian small-clawed otter are small and often do not show in a print. They all have webbing between the toes and a wide gap (negative space) between toes and meta pads. The heel pad as well as webbing and toe 1 (thumb) often do not register well in the track and are difficult to see for the untrained eye, which is one of the reasons why otter tracks can be mistaken for canine tracks (Rhyder, 2021; Grolms, 2021). Figure 2 shows the numbering of toes in otter tracks as well as clearly labelled metacarpal, carpal and metatarsal pads.



*Figure 2: Illustration of otter left front and hind footprints with the appropriate labels for toes 1 to 5, metacarpal, carpal and metatarsal pads.*

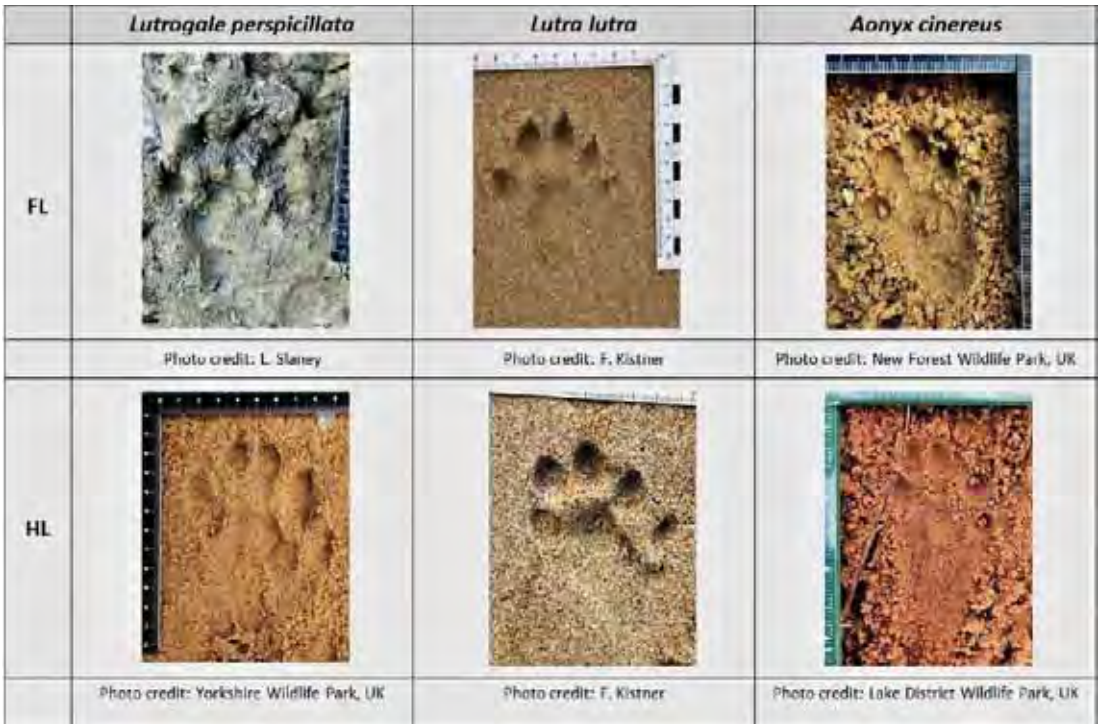
This shows how complex tracks and tracking are and to evaluate photographs of footprints can even be more challenging. In the field, well-trained trackers have a lot more information available than can be found in a photograph and they are able to take gait and behavioural signs, such as movement through habitat and sprinting

behaviour, into consideration for species classification. For most wildlife conservationists who lack such training this is not an option, but we show here that images of footprints can be analysed and are a very valuable data source, which provide a non-invasive tool for otter surveys and monitoring.

### ***Data collection***

Footprint images were collected following WildTrack's standardised footprint collection protocol for FIT (see Appendix) at eight zoos (six U.K., two Germany) by either the zoo staff or the researchers themselves. Several zoos have groups of otters in their enclosures, where keepers find it difficult to identify individuals. However, for the present study we focussed on classification at species level only, so individual identification of otters was not necessary.

To improve footprint quality, where possible, sand patches were created in the enclosures. However, prints found in muddy substrate were also included in the study. Each footprint was then photographed with a metric ruler for scale. Figure 3 shows footprint image examples of the three otter species, which show a variety of substrates and footprint qualities included in this study.



**Figure 3. Examples of otter footprints of different quality and in a variety of substrates, which were used for the study. Note: FL = Front Left; HL = Hind Left;**

The analysis is based on 100 footprint images for each species. Because of the large group size at some of the otter enclosures, we were only able to estimate a range of the number of individual otters which contributed data to the study (Table 1).

**Table 1: Range of individuals per species contributing data**

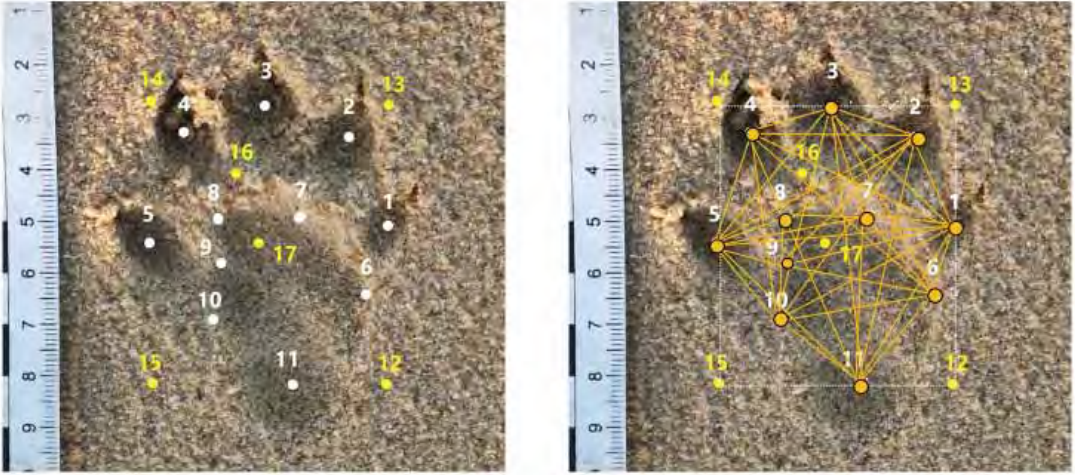
<b>Species</b>	<b>Range of individuals in study</b>
<b>Eurasian otter (<i>Lutra lutra</i>)</b>	<b>10-18</b>
<b>Asian small-clawed otter (<i>Aonyx cinereus</i>)</b>	<b>8-13</b>
<b>smooth-coated otter (<i>Lutrogale perspicillata</i>)</b>	<b>6-16</b>

**Image processing**

All subsequent image processing and statistical analyses were carried out in JMP 16.2 Pro version. The add-in, XGBoost, created for JMP software (Wolfinger, 2020) was additionally installed.

Following a standardised protocol, 300 footprints (100 per species) were analysed. This process comprises alignment and orientation of the image, calibration of scale and the manual placement of 11 specific anatomic landmark points (Figure 4). We used hind and front feet and aimed to define points that are consistently found in front and hind feet among all three species. The setting of the landmark points followed the standardised protocol for left feet. Therefore, images of the right feet were flipped horizontally prior to processing to increase the amount of processable footprints. Measurements here are taken at the centre of the toes and the heel pad, as this has proven to be more robust to variation in substrate.

A scripted routine adds a further six derived landmark points and subsequently automatically extracts a morphometric profile. This profile comprises 208 measurements per footprint, consisting of distances, angles, and areas, which is illustrated in Figure 4.



**Figure 4. Image of a left front track from a Eurasian otter *Lutra lutra*.**

*Left image: Footprints are first aligned at two rotation points at the base of the second and 4<sup>th</sup> toe and two points on a metric ruler are used for scaling of the image. Afterwards, 11 landmarks (white) are manually placed by a human operator and an additional six points are automatically derived (yellow).*

*Right image: Geometric profile of a footprint. A total of 208 measurements comprising distances, angles and areas form the morphometric profile of each footprint.*

## ***Statistical analysis***

### ***Species classification***

In a first step, the dataset of the biometric measurements was randomly split into a training set (75% of data) for the model development and a test set (25%) to evaluate the classification capability of the model. The test set was excluded from the model development.

Stepwise forward LDA and XGBoost Classifier were chosen as classification models (classifier) and trained on the training set.

To avoid overfitting of the model, an additional internal k-fold (n=5) cross-validation step was applied during training. Hyperparameters were automatically tuned within the JMP Model screening platform.

The classifiers assign probabilities between 0 and 1 to a class (here 'otter species'). The class with the highest probability is assigned the predicted species.

In a second step, the models with the highest overall classification accuracy (correct number of predictions divided by total number of footprints) in the training process were selected and applied to the test set, which can be seen as new data.

*Significance of footprint features ('Feature Importance')*

An advantage of using tree-based methods, such as XGBoost, is that the importance of features for the prediction of the model can be evaluated. Often this is done by the calculation of gain scores, with "gain" being defined as the improvement in accuracy brought by a feature to the branches it is on (Abu-Rmileh, 2019). For LDA, Feature Importance can be identified by calculating F-Ratios of the variables (Gu et al., 2014).

We analysed the most important features chosen by the best overall model and investigated how much of the variance of the whole data could be explained when focusing only on the most influential measurements. In order to achieve this, we built a decision tree model with only two splits as a general field assessment recommendation.

**RESULTS**

*Species classification*

For the whole dataset, 291 of the 300 images (training and test set) were correctly assigned to its respective species by our final XGBoost model, leading to an overall classification accuracy of 97%. The Linear Discriminant Analysis (LDA) classified 271 prints correctly and therefore had an overall accuracy on the same data of 90%.

XGBoost predicted the Asian small-clawed otter correctly at a rate of 99%, the Eurasian otter at 97% and the smooth-coated otter at 95%. This and the predictions of the LDA are shown in Table 2, which is the confusion matrix of the entire dataset.

The confusion matrix compares the model prediction with the true class (species) of each data point (footprint). A perfect model would only have non-zero values on the main diagonal of the matrix.

*Table 2. Confusion matrix of the whole dataset. Each species had one hundred footprints. The confusion matrix compares the true species of a footprint with the prediction of the XGBoost and the LDA classifier*

Species	Prediction XGBoost			Prediction LDA		
	<i>Aonyx cinereus</i>	<i>Lutra lutra</i>	<i>Lutrogale perspicillata</i>	<i>Aonyx cinereus</i>	<i>Lutra lutra</i>	<i>Lutrogale perspicillata</i>
<i>Aonyx cinereus</i>	99	0	1	96	3	1
<i>Lutra lutra</i>	1	97	2	1	92	7
<i>Lutrogale perspicillata</i>	1	4	95	3	14	83

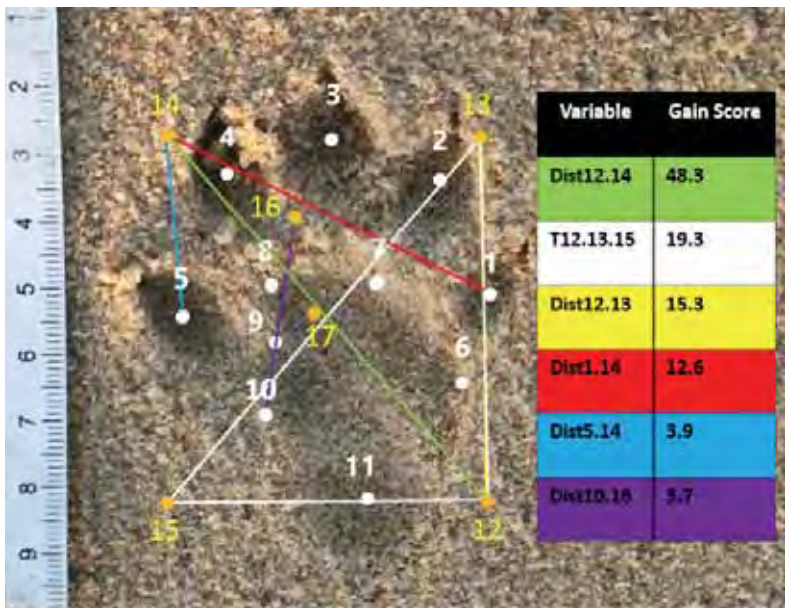
When looking solely at new data of the test set, the overall accuracy for XGBoost was 91% and 85% for the LDA model. ASCs were predicted with 96% accuracy, Eurasian otters with 92% and smooth-coated with 84% by XGboost and 92% of Asian small-clawed otters, 84% of Eurasian and 80% for the smooth-coated applying the best LDA model (Table 3).

*Table 3. Confusion matrix of the test set. The test set comprised of 25 footprints per species. XGBoost had an overall accuracy of 91% correct predictions, LDA of 85%.*

Species	Prediction XGBoost			Prediction LDA		
	<i>Aonyx cinereus</i>	<i>Lutra lutra</i>	<i>Lutrogale perspicillata</i>	<i>Aonyx cinereus</i>	<i>Lutra lutra</i>	<i>Lutrogale perspicillata</i>
<i>Aonyx cinereus</i>	24	0	1	23	2	0
<i>Lutra lutra</i>	1	23	1	1	21	3
<i>Lutrogale perspicillata</i>	1	3	21	1	4	20

**Feature importance of the XGBoost and field recommendations**

As XGBoost outperformed the LDA approach, we therefore only report the most important features of the better model. For our data, the final XGBoost classifier used 103 variables, with a gain score higher than 0. To take that number of measurements in the field is beyond practicality. Therefore, here we only want to focus on the six most important features, which are illustrated in Figure 5. It is noticeable that these six most essential measurements cover five distances and one triangle, distributed among the whole footprint.



**Figure 5. Colour-coded illustration of the six most notable features of the final XGBoost model. The implemented table reports the gain score of the respective variable. The most important feature is distance 12.14, highlighted in green.**

**Decision tree and potential field identification key**

The most influential feature of the model is the distance between point 12 and 14 with a gain score of 48.3. This measurement is equal to the square root of the sum of the squared length and the squared width of the footprint and can be measured in the field.

A further simplified model, a two split regression tree using only this measurement, misclassified 37 of the total 300 footprints, leading to an overall classification accuracy of 88% (Table 4).

**Table 4. Confusion matrix of the true species vs. the prediction of a simplified decision tree model with only the variable dist.12.14 and two splits.**

Species	Prediction decision tree model		
	<i>Aonyx cinereus</i>	<i>Lutra lutra</i>	<i>Lutrogale perspicillata</i>
<i>Aonyx cinereus</i>	98	2	0
<i>Lutra lutra</i>	3	89	8
<i>Lutrogale perspicillata</i>	4	20	76

Within our dataset, 93% of measurements smaller than 6cm belonged to Asian small-clawed otters. Measurements greater than 8.3 cm belonged to smooth-coated otters in about 90% of the cases. Measurements between 6–8.3cm had the highest probability (~80%) of belonging to Eurasian otters, however, the two other species also had measurements of footprints within this interval (18% smooth-coated otters, 2% Asian small-clawed otters). The full classification tree model with additional information is displayed in Figure 6.

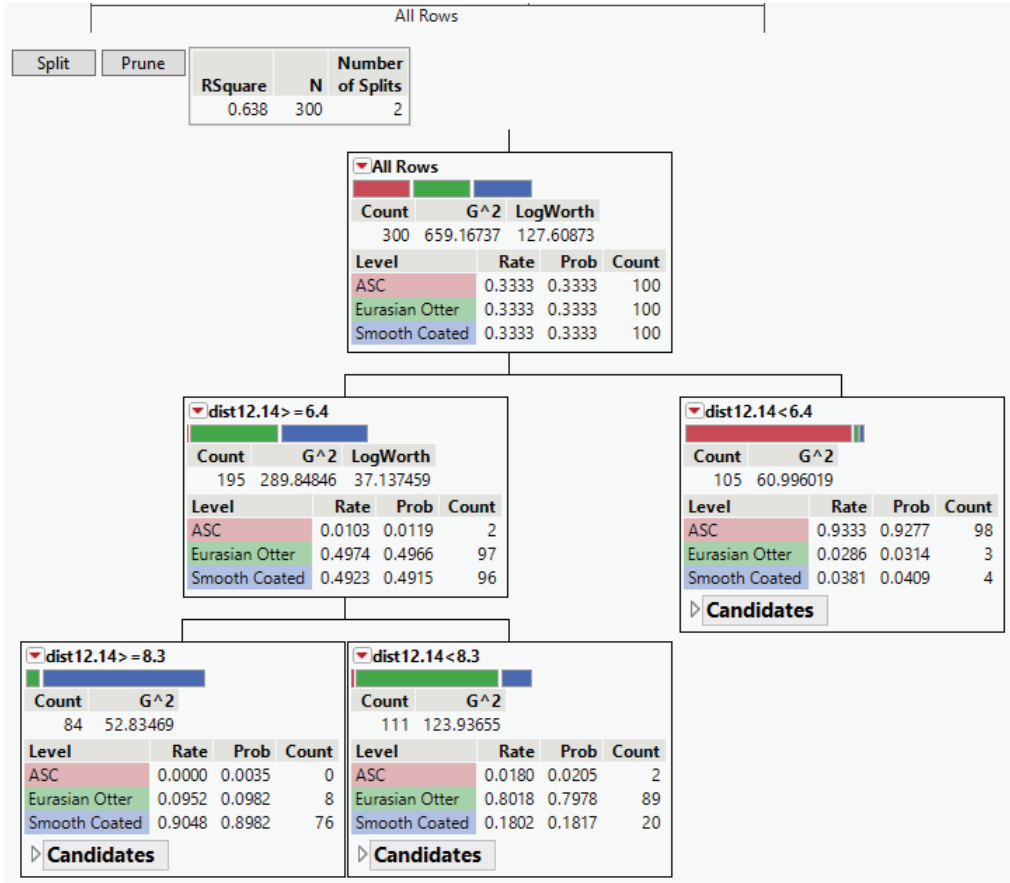


Figure 6. Output of decision tree model on the whole dataset. Target variable was species, explaining variable dist12.14. Only two splits were allowed. The model had an R<sup>2</sup> of 0.638. 88% of the data were classified correctly.

## DISCUSSION

This research shows that FIT-models can predict otter species with an excellent accuracy when Machine Learning models trained on a high number of automatically derived morphometrics from otter footprints are applied.

We also show that XGBoost outperformed Linear Discriminant Analysis and therefore recommend the use of XGBoost if a minimal misclassification rate is desired. By investigating the most important features of the best XGBoost model, it was possible to identify a single measurement that can be taken in the field and still enables a good classification accuracy on sight.

These encouraging results confirm that footprints can be a reliable source of species classification for otters. This could be a valuable, cost-effective, non-invasive, and accurate tool for field biologists, conservation policy makers and stakeholders to learn more about otter species distribution on the Asian continent. FIT could

therefore be a useful tool for otter surveys by providing data on species distribution and population size in addition to data from camera trapping and spraint analysis.

The lack of accurate data on otter species distribution and the need for otter surveys throughout southeast Asia became evident at the recent Malaysia Otter Workshop 2022, which was held jointly by the Malaysia Otter Network (MON), the Malaysia Nature Society (MNS) and the International Otter Survival Fund (IOSF) in Kuala Selangor Nature Park, Malaysia. The general consensus was that due to habitat loss, pollution and the illegal pet and fur trade, which are both on the rise in southeast Asia, otter numbers are declining (**IOSF Workshop Report, 2022**). Without accurate data, it is not known how serious the situation for each otter species is, which makes official and accurate surveys all the more important.

To survey more effectively, we need to engage a wider body of data collectors. The engagement of local communities, particularly those still holding Traditional Ecological Knowledge (TEK) could be transformational. Not only would this engage those communities as key stakeholders in the conservation process, but it would greatly augment the quality and volume of data available (**Jewell et al., 2020**). FIT was inspired by traditional tracking techniques and the data collection protocol is widely accessible. After a short training period, data collection for FIT is straightforward and only requires a (mobile phone) camera and a metric ruler as equipment.

Even though the results are already encouraging, they could be further improved with several adjustments:

- Increase the size and complexity of the dataset: The results are currently based on a small data set. The aim is to increase this to improve accuracy and test the robustness of the tool.
- In addition to improving the data set size and variation, continued re-training of classification models with clearly identifiable images from local (sub-) populations may be useful. This would likely improve accuracy and enable FIT models to address specific variations in those populations, which will most likely lead to better predictions.
- We would further like to develop models that can also predict sex and individual ID. This has been developed successfully for several other species, such as giant panda (*Ailuropoda melanoleuca*) (**Li et al., 2018**), mountain lion (*Puma concolor*) (**Alibhai et al., 2017**) and Amur tiger (*Panthera tigris altaica*) (**Gu et al., 2014**). To do so, we require many footprints clearly identifiable to the target class. This is ideally done by collecting footprint images of single, known individuals.
- To have a model that can classify all four Asian otter species, we would like to add footprint images of known hairy-nosed otters (*Lutra sumatrana*) to the database. This has proven challenging so far, as we are aware of only one

captive individual. However, for species classification, we could use good quality field data to develop the necessary algorithm as long as we can be sure that the footprints undoubtedly belong to the species. Long trails from the field, undoubtedly belonging to the same animal, could also be used to train the model for individual ID.

Zoos and field conservationists are therefore invited to contribute footprint images of known Asian otter species and, if possible, known individuals for sex and individual ID. Machine Learning models perform better with more data and otter conservationists are encouraged to further upload any images of known otter species/individuals following the link and barcode below.

All footprint images will, (given permission) also contribute to WildTrack's AI project (<https://www.wildtrack.ml>). This project aims to build computer vision models that enable classifications of footprints without the manual setting of landmarks. Once successfully developed, it will enable us to scale up this method significantly. Ultimately, we encourage field researchers to reach out and try this method and cross-validate it with other non-invasive monitoring approaches.



### **Acknowledgements**

The authors would like to thank all contributing zoos, namely Colchester Zoo, Lake District Wildlife Park, New Forest Wildlife Park, Woburn Safari Park, Yorkshire Wildlife Park, the Zoologischer Stadtgarten Karlsruhe and Zoo-Landau for either enabling us to collect images on site or directly uploading images to our WildTrack database. A special thanks goes to the International Otter Survival Fund (IOSF) and Prof Dr. Ing. Stefan Hinz from the IPF for their continued support.

We would also like to thank JMP Software for kindly providing a complementary licence.

### **Funding**

The authors would like to thank the Karlsruhe House of Young Scientists (KHYS) for the funding support for this work.

### **Disclosure Statement**

No potential conflict of interest was reported by the authors.

## Author Biographies

FREDERICK KISTNER has a degree in Environmental Science and a strong interest in applied conservation biology and the usage of Machine Learning for ecological problems. He has been tracking and studying the tracks and signs of coastal otters in Portugal for many years, is a member of the IUCN Otter Specialist Group and a founding member of the WildTrack Specialist Group. He is currently doing his PhD at the Karlsruhe Institute of Technology where he is working on the development of new monitoring tools for elusive species.

LARISSA SLANEY is a founding member of the WildTrack Specialist Group, who specialise in the use of non-invasive methods for wildlife monitoring. She has a BSc (Honours) in Life Sciences, a 1st State Exam in German Law and is currently doing her PhD in cheetah footprint analysis at Heriot-Watt University in Scotland, U.K. She also develops science workshops and has worked internationally as a science communicator. Her interests are in non-invasive wildlife monitoring, education and traditional tracking.

ZOE JEWELL is the co-founder and Executive Director of WildTrack. She has a BSc in Zoology/Physiology, an MSc in Medical Parasitology from the London School of Tropical Medicine and Hygiene (London University) and a degree in Veterinary Medicine from the University of Cambridge, U.K. Jewell is responsible for WildTrack's field operations, research and development. She publishes on wild animal monitoring using non-invasive approaches and has a particular interest in animal monitoring ethics. She is an adjunct associate professor at the Nicholas School of the Environment at Duke University, NC, and an associate academic at the Centre for Compassionate Conservation at the University of Technology in Sydney.

ASAF BEN-DAVID is a qualified Level 3 tracker with the Cybertracker system. He teaches a Wildlife Tracking Course at Tel-Chai College in Israel, and is the author of the "Israeli Tracks Field Guide". He specialises in technical drawings of track images. He is also currently a PhD student at Tel Aviv University at the School of Zoology, specialising in citizen science with tracking data.

SKY ALIBHAI is the co-founder and Director of WildTrack. He has a BSc in Zoology from Makerere University, Uganda, and a DPhil in small mammal population dynamics from the University of Oxford, U.K. He is a fellow of the Zoological Society of London. Alibhai leads WildTrack's analytical platform and publishes on wild animal monitoring using non-invasive approaches. He is currently an adjunct associate professor at the Nicholas School of the Environment at Duke University, NC, and an associate academic at the Centre for Compassionate Conservation at the University of Technology in Sydney.

## REFERENCES

- Abu-Rmileh, A, 2019.** The multiple faces of 'feature importance' in XGBoost. *Towards Data Science*, 08 February 2019, <https://Towardsdatascience.Com/Be-Careful-When-Interpreting-Your-Features-Importance-in-Xgboost-6e16132588e7>. Accessed 10 May 2022
- Alibhai, SK, Jewell, Z and Law, PR, 2008.** A footprint technique to identify white rhino *Ceratotherium simum* at individual and species levels. *Endangered Species Research* 4, 1–2, 205–218.
- Alibhai, SK, Jewell, Z and Evans, J, 2017.** The challenge of monitoring elusive large carnivores: An accurate and cost-effective tool to identify and sex pumas (*Puma concolor*) from footprints. *PLoS ONE* 12, 3.

- Chen, T and Guestrin, C, 2016.** XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Conde, DA, Staerk, J, Colchero, F, da Silva, R, Schöley, J, Baden, HM, Jouvet, L, Fa, JE, Syed, H, Jongejans, E, Meiri, S, Gaillard, J-M, Chamberlain, S, Wilcken, J, Jones, OR, Dahlgren, JP, Steiner, UK, Bland, LM, Gomez-Mestre, I, Lebreton, JD, González Vargas, J, Flesness, N, Canudas-Romo, V, Salguero-Gómez, R, Byers, O, Berg, TB, Scheuerlein, A, Devillard, S, Schigel, DS, Ryder, OA, Possingham, HP, Baudisch, A and Vaupel, JW, 2019.** Data gaps and opportunities for comparative and conservation biology. *Proceedings of the National Academy of Sciences of the United States of America*, 19 Apr 2019, 116(19):9658-9664. <https://europepmc.org/article/pmc/pmc6511006#free-full-text>. Accessed on 12 May 2022.
- Duplaix, N and Savage, M, 2018.** The global otter conservation strategy. *IUCN/SSC Otter Specialist Group, Salem, Oregon, USA*
- Elbroch, M, 2019.** *Mammal tracks and signs: a guide to North American species*. Second Edition, Stackpole Books, Guilford, Connecticut.
- Evans, JW, Evans, CA, Packard, JM, Calkins, G and Elbroch, M, 2009.** Determining observer reliability in counts of river otter tracks. *Journal of Wildlife Management*, 73,3, 426–432.
- Grolms, J, 2021.** *Tierspuren Europas: spuren und zeichen bestimmen und interpretieren*. Verlag Eugen Ulmer, Stuttgart.
- Gu J, Alibhai SK, Jewell ZC, Jiang G and Ma J, 2014.** Sex determination of Amur tigers (*Panthera tigris altaica*) from footprints in snow. *Wildlife Society Bulletin 04/2014*, 38, 3, 495–502.
- IOSF Malaysian Workshop Report ,2022.** *Otter, Journal of the International Otter Survival Fund*. Current issue.
- Jewell, Z, 2013.** Effect of monitoring technique on quality of conservation science. *Conservation Biology*, 27, 3, 501–508.
- Jewell, ZC, Alibhai, SK, Weise, F, Munro, S, van Vuuren, M and van Vuuren, R, 2016.** Spotting cheetahs: identifying individuals by their footprints. *Journal of Visualized Experiments*, 111, 1–11.
- Jewell, ZC, Alibhai, S, Law, PR, Uiseb, K and Lee, S, 2020.** Monitoring rhinoceroses in Namibia’s private custodianship properties. *Peer Journal* 8:e9670.
- Li, BV, Alibhai, S, Jewell, Z, Li, D, and Zhang, H, 2018.** Using footprints to identify and sex giant pandas. *Biological Conservation*, 218, 83–90.
- Liebenberg, L, 2021.** The origin of science: the evolutionary roots of scientific reasoning and its implications for tracking science. *Cape Town, South Africa: CyberTracker*.
- Petso, T, Jamisola, RS and Mpoeleng, D, 2022.** Review on methods used for wildlife species and individual identification. *European Journal of Wildlife Research*, 68, 1, 1–18.
- Ramos, SC, Shenk, TM and Leong, KM, 2016.** Introduction to traditional ecological knowledge in wildlife conservation. *Natural Resource Report NPS/NRSS/BRD/NRR—2016/1291*. National Park Service, Fort Collins, Colorado.
- Rhyder, J, 2021.** *Tracks and signs: a guide to the field signs of mammals and birds of the U.K.* The History Press, Cheltenham, Gloucestershire.

**Sharma, S, Jhala, Y and Sawarkar, VB, 2005.** Identification of individual tigers (*Panthera tigris*) from their pugmarks. *Journal of Zoology* 267, 1, 9-18.

**Vieira, MAR, von Muhlen, EM and Shepard, GH Jr., 2015.** Participatory monitoring and management of subsistence hunting in the Piagaçu-Purus Reserve, Brazil, *Conservation and Society*, 13, 3, 254-264.

**Wolfinger, R, 2020.** XGBoost Add-In for JMP Pro, *JMP User Community blog on topic of JMP Add-Ins*, 28 April 2022, <https://community.jmp.com/t5/JMP-Add-Ins/XGBoost-Add-In-for-JMP-Pro/ta-p/319383>. Accessed 01 June 2022.

**Yoxon, P and Yoxon, GM, 2017.** *Otters of the World*. Whittles Publishing Ltd, Dunbeath, Caithness, Scotland.


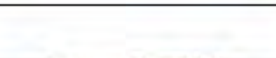



## Appendix

### FIT Zoo Protocol for Otter

You need:

- sand, rake, trowel or similar
- metric(!) ruler (an L-shaped ruler/tri-square is ideal)
- label
- smartphone camera/digital camera
- watering can if necessary

Sand patches should be strategically placed at frequently used areas, e.g. the entrance of nest boxes, near feeding places, at water entry points or at the otters' patrol routes along fence lines.

<p>1. <b>Prepare</b> the ~1cm deep sand patch by raking and levelling with trowel, add some water with watering can first if sand too dry. The sand should be level but not too compressed in the end.</p>	
<p>2. Let <b>one individual walk across</b> the sand patch so that you know for sure which otter left the footprints.</p>	
<p>3. - Take <b>overhead image of the trail*</b> (several tracks)          - Locate complete <b>left front (LF) footprints</b> (see image on right) for FIT and circle them.          - Any other good quality footprints of the <b>other feet</b> (RF, LH, RH) can be used for <a href="#">WildTrack's AI project</a>.</p>	
<p>4. Place a <b>metric(!) L-shaped ruler</b> approximately <b>2cm above the footprint</b>, framing the track as shown in the image on the right (and making sure not to touch the edges of the footprint).</p>	
<p>5. Add a <b>label</b> stating the date and place of where the footprint was taken. Also note the number or name of the animal and the name of the photographer. When following a trail*, number it and mark the tracks* with letters 'A', 'B', 'C' etc. Also give foot ID as LF (left front) (or RF, LH, RH if you are also collecting for AI project)</p>	<div style="border: 1px solid black; padding: 5px;"> <p>DATE: 04 April 2019    ANIMAL ID: AIJ05/name              LOCATION: name of zoo _____              FOOTPRINT #: 1 A, B, C _____              FOOT ID: LH (left hind) _____              PHOTOGRAPHER: I.S.</p> </div>
<p>6. Take a <b>close-up photo</b> of the footprint, incl. the label and enough of the ruler to clearly see a measurement of at least 5cm. It is important that the camera is absolutely parallel to the footprint (ensure right angle of ruler compliments right angle of camera screen). Photos of completely shaded footprints normally give the clearest image. Use an umbrella or similar to create shade if necessary.</p>	

\***track** = a footprint; **trail** = unbroken line of tracks belonging to one individual

## ARTICLE

# CAN YOU TELL THE SPECIES BY A FOOTPRINT? - IDENTIFYING THREE OF THE FOUR SYMPATRIC SOUTHEAST ASIAN OTTER SPECIES USING COMPUTER VISION AND DEEP LEARNING

Frederick KISTNER\*<sup>1,2</sup>, Larissa SLANEY<sup>2,3</sup>, Nicolas MORANT<sup>4,5</sup>

<sup>1</sup>*Institute for Photogrammetry (IPF), Institute for Technology Karlsruhe (KIT), Germany*

<sup>2</sup>*WildTrack Specialist Group (WSG)*

<sup>3</sup>*Institute for Life and Earth Sciences, School of Energy, Geosciences, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK*

<sup>4</sup>*Harvard University*

<sup>5</sup>*WildTrack*

\*Corresponding Author: Postal Address: Photogrammetrie und Fernerkundung (IPF) Karlsruher Institut für Technologie (KIT)  
Email: frederick.kistner@kit.edu



(Received 12<sup>th</sup> May 2023, accepted 21<sup>st</sup> July 2023)

**Abstract:** Southeast Asia is home to four otter species, all with decreasing population trends. All four of Southeast Asia's otter species can coexist sympatrically and are on the International Union for Conservation of Nature's (IUCN) Red List of Endangered Species. There are knowledge gaps in the distribution range and population sizes of these elusive species, which is essential information for the implementation of effective conservation measures. Footprints can be a cost-effective, non-invasive way to collect relevant data. WildTrack has developed a Footprint Identification Technology (FIT) classification model that uses landmark-based measurements as input data. This model is highly accurate at distinguishing between three of the four otter species in Southeast Asia. In this study, we propose a deep learning-based approach that automates the classification of species by analyzing the area within bounding boxes placed around footprints. The method significantly reduces the processing time and eliminates the need for highly skilled operators placing landmark points on footprints.

To train the model, 2,562 images with 3,895 annotated footprints were used, which resulted in an impressive accuracy, precision, and recall of 99% on both training and test sets. Furthermore, the model's performance was tested on a new set of 431 footprints, which were not used in the training process, and only 4 of them were incorrectly classified, demonstrating the effectiveness of the proposed approach on unseen data. The research findings of this study confirm the viability of using a machine learning model-based approach to accurately identify otter species through their footprints. This approach is both reliable and cost-effective, which makes it an attractive tool for otter monitoring and conservation efforts in Southeast Asia. Additionally, the method has significant potential for application in community-based citizen science monitoring programs. Further research could focus on expanding the scope of the study by adding footprints from hairy-

nosed otters, as well as sympatric non-otter species, to the training database. Furthermore, this study suggests developing an object detection model and training new classification models that predict sex or re-identify individuals using a larger set of images of known (captive) individuals.

**Citation:** Kistner, F., Slaney, L. and Morant, N. (2023). Can You Tell the Species by a Footprint? - Identifying Three of the Four Sympatric Southeast Asian Otter Species using Computer Vision and Deep Learning. *IUCN Otter Spec. Group Bull.* 40 (4): 197 - 210

**Keywords:** Eurasian otter, *Lutra lutra*, Small-clawed otter, *Aonyx cinereus*, Smooth-coated otter, *Lutrogale perspicillata*, tracks, non-invasive animal monitoring

## INTRODUCTION

There are four species of otter that share the same general geographic range in Southeast Asia. These are the Asian small-clawed otter (*Aonyx cinereus*), the Eurasian otter (*Lutra lutra*), the Hairy-nosed otter (*Lutra sumatrana*), and the Smooth-coated otter (*Lutrogale perspicillata*). They are listed on the IUCN's Red List of Endangered Species and all four species are suffering from declining population trends (Duplaix and Savage, 2018).

Otters are an important flagship species for conservation (Stevens et al., 2011) and indicator species for water and wetland health (Bhandari and GC, 2008a). Some even describe them as keystone species (Basnet et al., 2020; Bhandari and GC, 2008b). In Southeast Asia, otters face numerous threats to their survival, including habitat loss and degradation (de Silva, 2011; Foster-Turley, 1992), hunting, poaching, illegal wildlife trade (Feeroz, 2015; Soe, 2022), and climate change (Cianfrani et al., 2018), as well as pollution and human-otter conflicts (Gomez and Bouhuys, 2018; Yoxon and Yoxon, 2017). There are knowledge gaps of the distribution range and population sizes of these elusive species in Southeast Asia, which is essential information for the implementation of effective conservation measures (Duplaix and Savage, 2018).

Effective conservation efforts require knowledge of a species' distribution and abundance to understand their ecological needs and develop targeted conservation strategies. This is particularly important in areas where little or no data is available and therefore conservation needs are difficult to judge (Duplaix and Savage, 2018). Under Objective 4 of the Global Otter Conservation Strategy by the IUCN's Otter Specialist Group, noninvasive survey methods, such as camera-trapping, spraint analysis and environmental DNA (eDNA) analysis are recommended (Duplaix and Savage, 2018). In addition, a novel, non-invasive and cost-effective PCR-RFLP eDNA method has recently been developed for otter, which relies on finding otter spraints for analysis (Sharma et al., 2022).

Despite the benefits of noninvasive survey methods, challenges remain in identifying otter species. Tracks can be a valuable and cost-effective source of noninvasive data and have the benefit that they can be collected by local communities and citizen scientists, which can enhance public awareness and engagement in conservation (Danielsen et al., 2005). They can however be difficult to find and interpret, particularly in dense or rugged environments. Even experienced field observers can misidentify tracks, with a misclassification rate of 44% reported for North American river otters (Evans et al., 2009). However, digital images of tracks can be analyzed later by an expert and, when taken in a standardized way with a scale, can be used to extract morphometric measurements for classification purposes.

WildTrack is a prominent research group in footprint identification and has published extensively on footprint analysis for many endangered and elusive species using their highly accurate Footprint Identification Technology (FIT). Several studies have demonstrated the effectiveness of morphometric analysis using digital images of tracks for species identification (S. K. Alibhai et al., 2008; De Angelo et al., 2010;

Kistner et al., 2022), sex determination (S. Alibhai et al., 2017; Gu et al., 2014; Li et al., 2018), individual identification (S. Alibhai et al., 2017; S. K. Alibhai et al., 2008, 2023; Jewell et al., 2016; Li et al., 2018), and population size estimation (S. K. Alibhai et al., 2023; Jewell et al., 2020).

In a previous study, we used morphometric measurements derived from footprints of *Lutra lutra*, *Lutrogale perspicillata*, and *Aonyx cinereus* to build an XG boost classifier and achieved an overall species classification accuracy of 91% on new unseen test data (Kistner et al., 2022).

However, even though these morphometrics models perform well with a small amount of data, they have limitations. These include the requirement of an expert to set the morphometric extraction landmark points and the inability of the model to account for variations within a species that are not represented within the previously defined measurements. To address these limitations, computer vision techniques such as convolutional neural networks (CNNs) have been utilized successfully in various wildlife conservation efforts, including age prediction of pandas (Zang et al., 2022), species identification in camera trap images (Carl et al., 2020; Wägele et al., 2022; Willi et al., 2019) and individual animal recognition (Chen et al., 2020; Hansen et al., 2018). These techniques offer advantages in speed, automation, and cost-effectiveness for species identification in conservation applications (Wäldchen and Mäder, 2018).

This study builds on previous research (Kistner et al., 2022) and investigates the possibility of distinguishing between three otter species by their footprints with a deep learning approach. Here we aim to address the limitations of previous morphometric approaches and explore the use of CNNs for the identification of otter species in Southeast Asia using digital images of tracks. As far as we know, this is the first study to use a method that combines traditional track identification with computer vision and machine learning to classify otter tracks with high accuracy, allowing for intra-species variation. We demonstrate the potential of this technique as a valuable tool for otter conservation efforts in the region by achieving accurate species identification.

## **METHODS**

### **Ethics Statement**

The majority of data collection for this study was carried out in zoos and otter sanctuaries by either professional wild animal keepers or under their supervision.

Keepers followed their normal working practice and followed working risk assessments. Data collection involved the non-invasive collection of digital otter footprint images found on substrates in the animals' enclosures. All animals walked across these natural substrates out of their own free will. This non-invasive approach avoided any direct contact with the animals to minimize any potential risks associated with handling them. This also ensured the animals left the footprints whilst displaying their natural behavior. Footprint data found in the field was collected after the wild otters had left the area, thus avoiding any stress on the wild animals and keeping both researchers and animals safe. Therefore, our data collection method was ethical and non-invasive, aligning with the principles of animal welfare.

### **Data Collection**

In this study, we only used images with a known species identity. Therefore, we collected images from captive otters and added field prints from Eurasian otters in Portugal where only one species of the three is present. Smooth-coated otter prints were collected *in-situ* after visual identification of the species. In total, we received a dataset of 2,562 images from over 20 participants, each contributing one or more otter

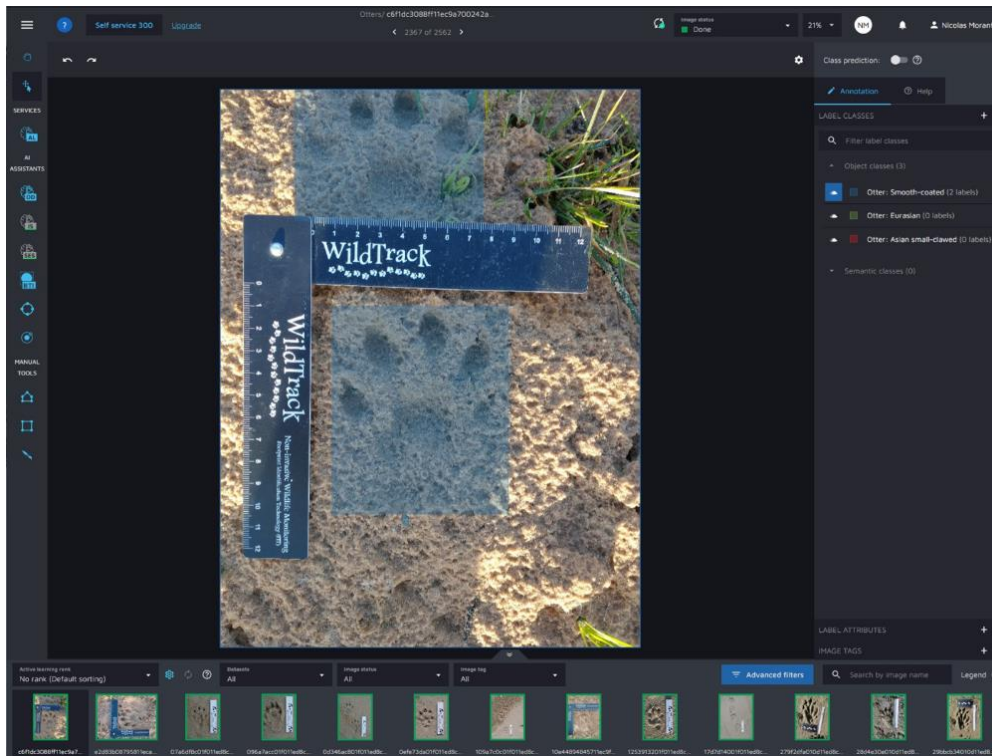
footprints. The data collection involved 17 zoos and otter sanctuaries in Germany, the UK, Austria and France. Figure 1 shows data collection of otter footprint images following the FIT protocol at a UK zoo. To ensure the quality of the images, we rated them on a scale of 1 (poor) to 5 (excellent). Most of the images were collected following the FIT protocol for collecting otter footprints (Kistner et al., 2022) and included a metric ruler for scale, but images with multiple footprints taken from various heights, some of which did not feature a scale, were also included.



**Figure 1.** Images of Asian small-clawed otter footprints collected at a UK zoo following the FIT protocol. Pictures are taken directly above and should contain a metric ruler. Photo credit: Woburn Safari Park, UK

### Data Annotation

To annotate images, we used Hasty, a CloudFactory company, which helped us streamline the process of labeling images by creating an accurate ground truth of our data at a much faster rate. The platform's user-friendly interface and built-in tools made it easy for our team to preprocess and label the data efficiently, thereby reducing the amount of time and resources required for this crucial step in training our models. Specifically, we annotated 3,895 footprints with known species identities by placing bounding boxes around them. This included 1,138 *Aonyx cinereus*, 479 *Lutrogale perspicillata*, and 2,278 *Lutra lutra*. We annotated all four feet in the images as part of the annotation process. Figure 2 shows a screenshot illustrating data annotation in Hasty.



**Figure 2.** Bounding Box labeling of an otter footprint image using Hasty.

## Model Training

Additionally, Hasty's support for a wide range of algorithms and frameworks, as well as its ability to train models on cloud-based resources, has allowed us to experiment with different approaches and find the best solution for our needs. Furthermore, the monitoring and analyzing model performance feature allowed us to track progress and identify areas for improvement. This helped us to optimize model training and ultimately improve the accuracy of our models.

We developed an image label classification model using the labeled bounding boxes containing the otters' footprints with the objective of identifying the specific otter species. By providing only the bounding boxes image area, the model could focus on this specific region of interest for making the classification more accurate. This technique is especially useful as a lot of our images contain multiple footprints within one image.

To train our models, we divided the data into three sets: a training set (80%), a validation set (10%), and a test set (10%). To ensure that our data was split in a representative manner, we employed stratification, which divided the data into approximately equal proportions of the target classes. The training set was used to train the different models, while the validation set was used to evaluate performance and optimize parameters. Finally, the test set was used to assess the overall performance of the models on unseen data.

We selected the ResNet-18 CNN architecture (He et al., 2016) due to its good trade-off between accuracy, recall, and training duration. This architecture is suitable for small datasets like ours since it has fewer parameters in comparison to other CNN architectures like ResNet34, ResNet50, or larger models, which require more resources and longer training time ultimately resulting in higher costs for model training. We used the AdamW optimizer (Loshchilov and Hutter, 2017) which uses the Adam algorithm with weight decay regularization to prevent overfitting. We also applied, as our only

data transformation step, resizing the input images to 512x512 pixels. It is beyond the scope of this paper to go further into detail of CNN architecture and computational need for model training.

In addition, to improve the overall performance of our models, we tuned various hyperparameters. Firstly, we adjusted the datasets by selecting different ratings of the images used in our model. Secondly, we adjusted the base learning rate of our model to slowly tune its parameters and optimize its performance. Additionally, we used a scheduler to adjust the learning rate over time to allow our model to adjust to the changing data and improve its efficiency. Lastly, we adjusted the batch size to fine tune the model as shown in Table 1. These combined adjustments resulted in us being able to produce accurate multi-label classification models.

**Table 1.** Summary of model experiments with different hyperparameters under evaluation

Image Rating	Split	Base Learning Rate	Scheduler	Train Batch Size	Test Batch Size	Accuracy Train	Accuracy Val	Avg Loss Train	Avg Loss Val
5,4,3	80/10/10	0.0001	ReduceLR OnPlateau	64	24	100.00	99.66	0.00	0.01
All	80/10/10	0.0001	ReduceLR OnPlateau	64	24	99.44	99.07	0.00	0.02
5,4,3	80/10/10	0.0001	ReduceLR OnPlateau	64	24	99.75	98.99	0.00	0.01
5,4,3	80/10/10	0.0001	ReduceLR OnPlateau	64	24	99.66	97.54	0.01	0.05
5,4,3	80/10/10	0.0001	ReduceLR OnPlateau	128	24	99.44	97.64	0.01	0.05
5,4,3	80/10/10	0.001	ReduceLR OnPlateau	64	24	95.28	90.91	0.09	0.25
5,4,3	80/10/10	0.001	None	24	8	93.75	86.20	0.18	0.36
5,4,3	80/10/10	0.001	ReduceLR OnPlateau	24	8	89.92	82.15	0.30	0.54
All	80/10/10	0.001	None	24	8	91.50	63.81	0.25	1.29
All	80/10/10	0.001	None	4	1	58.17	56.15	0.95	0.95

## RESULTS

### Model Selection

The results of the different models are presented in Table 1, with the “Accuracy Train” column displaying the accuracy on the training data, the “Accuracy Val” column showing the accuracy on the validation data, the “Avg Loss Train” column displaying the average loss on the training data, and the “Avg Loss Val” column showing the average loss on the validation data. We selected the best model for images rated 5, 4, and 3 and the best model for all images, based on their performance on the validation set and were further evaluated on the holdout test set.

The model that performed the best for images with ratings 5, 4, and 3 had a base learning rate of 0.0001, employed the ReduceLROnPlateau scheduler, had a train batch size of 64, and a test batch size of 24. This model displayed exceptional results, achieving an accuracy of 100% on the training set, 99.66% on the validation set, and a negligible average loss of 0.00 on the training set and 0.01 on the validation set.

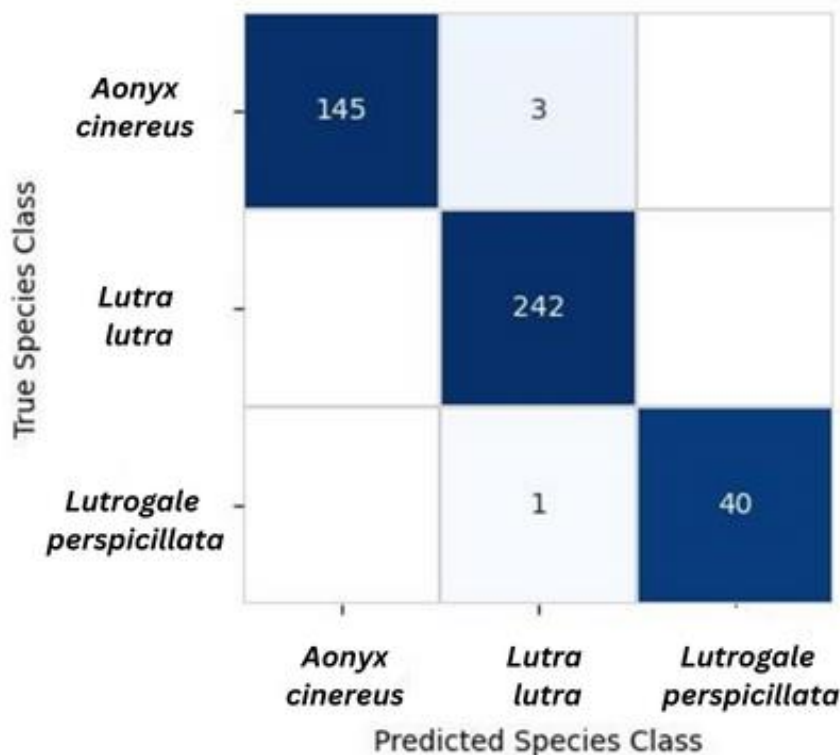
The best model for all images had a base learning rate of 0.0001, used the ReduceLROnPlateau scheduler, had a train batch size of 64, and a test batch size of 24. This model had a mAP accuracy of 99.44% on the training set, 99.07% on the validation set, and an average loss of 0.00 on the training data and 0.02 on the validation data. A full display of all further hyperparameters is displayed in the table above.

The model that performed the best for all images had comparable results to the model using only high-quality images. Therefore, we only present the performance of the model using all images on the test set, as it includes a broader range of images and is likely to have a greater level of generalizability.






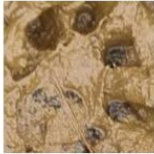
**Model Evaluation Test Set**

The performance of the model on the test set, which consists of previously unseen data, is displayed in this confusion matrix (Figure 3). It illustrates the number of accurate and inaccurate predictions made by the model for each species. For instance, the model correctly identified 40 instances of the Otter: Smooth-coated species, and it made one misclassification of this species. Additionally, it correctly identified 242 instances of the Otter: Eurasian species and had no incorrect predictions of this species. Likewise, it correctly identified 145 instances of the Otter: Asian small-clawed species but made 3 erroneous predictions of this species. Examples of these predictions are shown in Figure 4. Figure 5 is a saliency map, illustrating an exemplary visual representation of regions of importance for the classification model within this particular footprint.

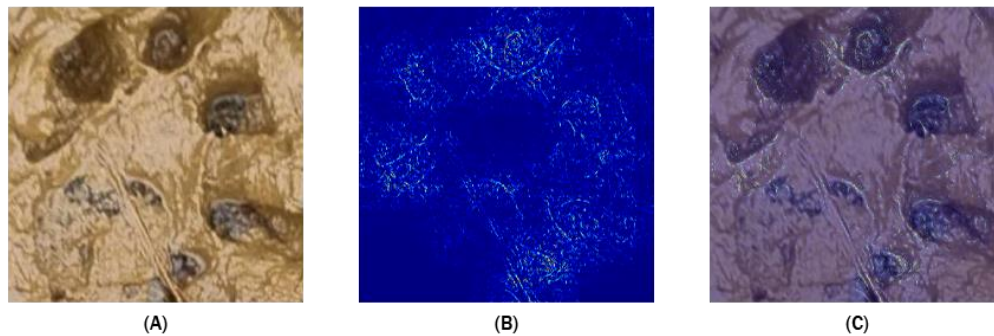
In conclusion, the model’s performance on the test set is relatively strong, with a high number of correct predictions for each species, but it also highlights the need for potential improvements in regard to the low number of incorrect predictions.



**Figure 3.** Confusion matrix for our best image classification model on unseen data: evaluating model performance

			
<b>True Label:</b>	Otter: Eurasian	Otter Smooth-coated	Otter: Asian small-clawed
<b>Predicted Label:</b>	Otter: Eurasian	Otter Smooth-coated	Otter: Asian small-clawed
<b>Confidence:</b>	100%	100%	98%
			
<b>True Label:</b>	Otter: Eurasian	Otter: Smooth-coated	Otter: Asian small-clawed
<b>Predicted Label:</b>	Otter: Eurasian	Otter: Smooth-coated	Otter: Asian small-clawed
<b>Confidence:</b>	100%	100%	100%

**Figure 4.** Predictions on unseen data using our best image label classification model.



**Figure 5.** Saliency map overlay on the original image. (A) Original image of a right front footprint from an Asian small-clawed otter. (B) Saliency map highlighting the most visually salient regions of the image. (C) Overlay of the saliency map on the original image, emphasizing the most visually significant regions of the footprint. The saliency map was generated using our best model. The overlay provides a clear visualization of the model output, highlighting the regions that help to predict and classify the otter species.

Table 2 displays the performance of the top model on the test set, which was trained using all images. The first column displays the number of times the model correctly identified each species (True Positive count). The second column displays the number of times the model incorrectly identified each species (True False count). The third column displays the number of times the model failed to identify each species (False Negative count). The fourth column shows the precision of each species, which is the percentage of correctly identified instances among all instances that were identified as that species. The fifth column shows the recall of each species, which is the percentage of correctly identified instances among all instances that actually belong to that species. The sixth column shows the micro-averaged F1 score, which represents the overall performance of the model. The seventh column shows the macro-averaged F1 score, which represents the model's performance across all species. The last column shows the weighted F1 score, which takes into account the relative importance of each species. The model performed well overall, with all the above metrics achieving a score of 0.99.

**Table 2.** Best Model (All Images) Classification Performance Table on Unseen Data

	<b>TP</b>	<b>TF</b>	<b>FN</b>	<b>Precision</b>	<b>Recall</b>	<b>Micro F1</b>	<b>Macro F1</b>	<b>Weighted F1</b>
<b>Model</b>	427	4	4	0.99	0.99	0.99	0.99	0.99

Table 3 shows the performance of our top classification model, which was trained on all images, in identifying three of the four Southeast Asian otter species. It reveals that the model has a high level of precision and recall for each species. The F1-Score, a measure of the balance between precision and recall, is also high, which means that the model is performing well in both identifying the species and minimizing false identifications. Overall, the table indicates that the model is effectively distinguishing between the three different types of otter species.

**Table 3.** Best Model (All Images) Classification Performance Table on Unseen Data

	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
<b>Otter: Smooth-coated</b>	0.98	1.00	0.99
<b>Otter: Eurasian</b>	1.00	0.98	0.99
<b>Otter: Asian small-clawed</b>	0.98	1.00	0.99

## DISCUSSION

The recent Malaysia Otter Workshop 2022, jointly organized by the Malaysia Otter Network, Malaysian Nature Society and the International Otter Survival Fund, highlighted the need for conducting otter surveys across Southeast Asia in light of the pressing issues of habitat degradation, pollution, human-otter conflict and the illegal trade of otter pets and fur (Yoxon, 2022). Given the scarcity of precise information on otter species distribution, it is crucial to carry out official and accurate surveys.

Our study demonstrates that using this non-invasive, low-cost and effective method of standardized, machine-learning-based footprint analysis provides an accurate survey method. Furthermore, to enhance the quality and quantity of available data, engaging local communities, especially those with Traditional Ecological Knowledge (TEK), could be a valuable approach. By incorporating the insights of these communities, a more comprehensive understanding of the current status of otters in the region could be attained.

Accurate baseline data is crucial for conservation efforts, but identifying elusive species like otters can be challenging. To tackle this challenge, we have developed a promising novel approach that leverages the advantages of CNN-based computer vision and ResNet models for identifying otter species using their footprints. Our approach has yielded promising results, outperforming our previous study (Kistner et al., 2022) in terms of performance scores on a larger dataset as well as reducing the time required for data labeling. Our study found that it is possible to distinguish between three sympatric otter species by analyzing their footprints using a partially automated computer vision approach. Furthermore, we believe the use of bounding boxes instead of morphometric landmark points has potentially reduced operator bias, thereby enhancing the reliability of our findings. We are optimistic that this innovative technique can pave the way for more effective monitoring and conservation of otters and other elusive species, while linking with TEK across communities globally, and linking in-situ and ex-situ species conservation and research. However, it should be noted that our current model is limited to predicting the three otter species it is trained on, even if a track originates from a non-otter species. To overcome this limitation, future research could focus on training the model on a comprehensive regional database

that includes tracks and signs of all species present in that specific region. Alternatively, a model could be developed to identify new classes once the data discrepancy reaches a certain threshold. Additionally, our approach is not entirely automated, as bounding boxes need to be manually set. Future research could enhance the methodology by incorporating object detection, footprint quality, and scale estimation models built into an entire machine learning pipeline, enabling a fully automated method. The integration of such an approach into a smartphone-based application with inference could potentially enhance user experience.

On the technical side, future research could explore the minimum quality criteria for footprints necessary for successful application of our approach. Additionally, our technique could be extended to benefit other species beyond otters, and data collection could involve not only professional trackers and zoos, but also local/indigenous people and citizen scientists. The participation of indigenous people and locals in conservation efforts is essential as they are key stakeholders in these areas and their involvement could potentially lead to job creation and the incorporation of Traditional Ecological Knowledge (TEK) into conservation initiatives (Danielsen et al., 2014; Ponce-Martins et al., 2022). This would not only foster greater community engagement but also promote the preservation of local ecosystems and their biodiversity.

From an ecological perspective, our approach could be used to look for new target classes, such as sex, age class, and individual identification, as demonstrated previously for other species with morphometrics (S. Alibhai et al., 2017; S. K. Alibhai et al., 2023; Jewell et al., 2016, 2020; Li et al., 2018). For instance, we could integrate hairy-nosed otters for the Southeast Asian region, create a method for individual identification, and develop an approach for other otter species. Footprint analysis could also aid in mitigating human-otter conflict by identifying the specific otter species or even the individual otter that is part of the conflict situation, such as conflict with commercial and subsistence fish farming (Duplaix and Savage, 2018; Shrestha, M.B., Shrestha et al., 2022) This would not only facilitate better management of human-otter interactions but also enhance our understanding of otter behavior and ecology.

## **CONCLUSION**

In conclusion, our approach has significant potential to improve our understanding of otter species distribution and behavior, thus providing valuable insights for conservation efforts aimed at safeguarding these charismatic and ecologically important species. By enabling more accurate and efficient identification of otter species, our approach could potentially facilitate more targeted conservation strategies, such as habitat protection and restoration, and the identification and mitigation of threats like pollution and poaching. Ultimately, this could help to preserve these animals' populations and the ecosystems they inhabit for future generations.

Currently, public use of the model is not possible due to concerns about protecting endangered species from potential misuse by poachers. However, we are working on developing a secure platform where verified biologists can register, submit images for inference, and receive results within seconds. In the future, this functionality will also be available through a mobile app, even in areas without internet access. This development is still in progress and will take time before it becomes available to users.

**Acknowledgements** - We would like to express our deep gratitude to Prof. Dr. Ing. Stephan Hinz for his invaluable guidance and support throughout the course of this study. Furthermore, we extend our sincere thanks to Deutsche Bundesstiftung Umwelt (DBU) and Karlsruhe House of Young Scientists (KHYS) for their generous financial support, which made this research possible. We express our utmost gratitude

to Zoe Jewell and Sky Alibhai from WildTrack for their invaluable support and the remarkable WildTrack AI footprint database they have initiated.

We extend our heartfelt appreciation to the zoos that have generously contributed to our research. We would like to express our gratitude to Colchester Zoo, Flamingo Land Resort, Lake District Wildlife Park, New Forest Wildlife Park, Woburn Safari Park, and Yorkshire Wildlife Park in the UK; Otter-Zentrum Hankensbüttel, Tiergarten Nürnberg, Tiergarten Straubing, Wiesbaden, Wildpark Pforzheim, Zoologischer Stadtgarten Karlsruhe, and Zoo-Landau in Germany; Alpenzoo Innsbruck and Salzburg Zoo in Austria; and NaturOparC in France. They have either facilitated our image collection on-site or directly uploaded images to our WildTrack database. Their valuable contributions have been instrumental in our research efforts.

We would like to extend a special thanks to the International Otter Survival Fund (IOSF) for their continued support and advice. Further thanks also to Amy Fitzmaurice and Huayu Tsu for their invaluable advice when reviewing the draft article.

Finally, we would like to acknowledge the contributions of Hasty, a CloudFactory company, which provided us with the necessary data labeling infrastructure and model training computational capacity, without which this would not have been possible.

## REFERENCES

- Alibhai, S., Jewell, Z., and Evans, J. (2017).** The challenge of monitoring elusive large carnivores: An accurate and cost-effective tool to identify and sex pumas (*Puma concolor*) from footprints. *PLoS ONE*, **12**(3). <https://doi.org/10.1371/journal.pone.0172065>
- Alibhai, S. K., Gu, J., Jewell, Z. C., Morgan, J., Liu, D., and Jiang, G. (2023).** ‘I know the tiger by his paw’: A non-invasive footprint identification technique for monitoring individual Amur tigers (*Panthera tigris altaica*) in snow. *Ecological Informatics*, **73** (December 2022): 101947. <https://doi.org/10.1016/j.ecoinf.2022.101947>
- Alibhai, S. K., Jewell, Z. C., and Law, P. R. (2008).** A footprint technique to identify white rhino *Ceratotherium simum* at individual and species levels. *Endangered Species Research*, **4**(1–2): 205–218. <https://doi.org/10.3354/esr00067>
- Basnet, A., Prashant, G., Timilsina, Y., and Bist, B. (2020).** Otter research in Asia: Trends, biases and future directions. *Global Ecology and Conservation*, **24**. <https://doi.org/10.1016/j.gecco.2020.e01391>
- Bhandari, J., and GC, D. B. (2008a).** Preliminary Survey and Awareness for Otter Conservation in Rupa Lake, Pokhara, Nepal. *Journal of Wetlands Ecology*, **2**. <https://doi.org/10.3126/jowe.v1i1.1570>
- Bhandari, J., and GC, D. B. (2008b).** Preliminary survey and awareness for otter conservation in Rupa Lake, Pokhara, Nepal. *Journal of Wetlands Ecology*, **2**. <https://doi.org/10.3126/jowe.v1i1.1570>
- Carl, C., Schönfeld, F., Profft, I., Klamm, A., and Landgraf, D. (2020).** Automated detection of European wild mammal species in camera trap images with an existing and pre-trained computer vision model. *European Journal of Wildlife Research*, **66**(4): 62. <https://doi.org/10.1007/s10344-020-01404-y>
- Chen, P., Swarup, P., Matkowski, W. M., Kong, A. W. K., Han, S., Zhang, Z., and Rong, H. (2020).** A study on giant panda recognition based on images of a large proportion of captive pandas. *Ecology and Evolution*, **10**(7): 3561–3573. <https://doi.org/10.1002/ece3.6152>
- Cianfrani, C., Broennimann, O., Loy, A., and Guisan, A. (2018).** More than range exposure: Global otter vulnerability to climate change. *Biological Conservation*, **221**: 103–113. <https://doi.org/10.1016/j.biocon.2018.02.031>
- Danielsen, F., Jensen, P. M., Burgess, N. D., Altamirano, R., Alviola, P. A., Andrianandrasana, H., Brashares, J. S., Burton, A. C., Coronado, I., and Corpuz, N. (2014).** A multicountry assessment of tropical resource monitoring by local communities. *BioScience*, **64**(3) : 236–251. <https://doi.org/10.1093/biosci/biu001>
- De Angelo, C., Paviolo, A., and Di Bitetti, M. S. (2010).** Traditional versus multivariate methods for identifying jaguar, puma, and large canid tracks. *The Journal of Wildlife Management*, **74**(5): 1141–1151. <https://doi.org/10.2193/2009-293>
- De Silva, P. K. (2011).** Status of Otter Species in the Asian Region Status for 2007: Proceedings of XIth International Otter Colloquium. *IUCN Otter Specialist Group Bulletin*, **28**: 97–107. [http://www.iucnosg.org/Bulletin/Volume28A/de\\_Silva\\_2011.html](http://www.iucnosg.org/Bulletin/Volume28A/de_Silva_2011.html)
- Duplaix, N., and Savage, M. (2018).** The global otter conservation strategy. In *IUCN/SSC Otter Specialist Group, Salem*. eScholarship, University of California. <https://escholarship.org/uc/item/12d608qf>

- Evans, J. W., Evans, C. A., Packard, J. M., Calkins, G., and Elbroch, M. (2009). Determining Observer Reliability in Counts of River Otter Tracks. *Journal of Wildlife Management*, **73**(3): 426–432. <https://doi.org/10.2193/2007-514>
- Feeroz, M. M. (2015). *Lutrogale perspicillata*. IUCN Bangladesh. Red List of Bangladesh Volume 2: Mammals. IUCN, International Union for Conservation of Nature, Bangladesh Country Office, Dhaka, Bangladesh. p.76  
<https://docslib.org/doc/418767/red-list-of-bangladesh-volume-2-mammals>
- Foster-Turley, P. (1992). Conservation Aspects of the Ecology of Asian Small- Clawed and Smooth Otters on the Malay Peninsula. *IUCN Otter Specialist Group Bulletin*, **7**: 26–29. [http://www.otterspecialistgroup.org/Bulletin/Volume7/Foster\\_Turley\\_1992.pdf](http://www.otterspecialistgroup.org/Bulletin/Volume7/Foster_Turley_1992.pdf)
- Gomez, L., and Bouhuys, J. (2018). Illegal otter trade in Southeast Asia. *TRAFFIC, Petaling Jaya, Selangor, Malaysia*. <https://www.traffic.org/publications/reports/illegal-otter-trade-in-southeast-asia/>
- Gu, J., Alibhai, S. K., Jewell, Z. C., Jiang, G., and Ma, J. (2014). Sex determination of Amur tigers (*Panthera tigris altaica*) from footprints in snow. *Wildlife Society Bulletin*, **38**(3): 495–502. <https://doi.org/10.1002/wsb.432>
- Hansen, M. F., Smith, M. L., Smith, L. N., Salter, M. G., Baxter, E. M., Farish, M., and Grieve, B. (2018). Towards on-farm pig face recognition using convolutional neural networks. *Computers in Industry*, **98**: 145–152. <https://doi.org/10.1016/j.compind.2018.02.016>
- He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 770–778. <https://doi.org/10.1109/CVPR.2016.90>
- Jewell, Z. C., Alibhai, S. K., Weise, F., Munro, S., Van Vuuren, M., and Van Vuuren, R. (2016). Spotting Cheetahs: Identifying Individuals by Their Footprints. *Journal of Visualized Experiments*, **III**, 1–11. <https://doi.org/10.3791/54034>
- Jewell, Z. C., Alibhai, S., Law, P. R., Uiseb, K., and Lee, S. (2020). Monitoring rhinoceroses in Namibia’s private custodianship properties. *PeerJ*, **8**, e9670. <https://doi.org/10.7717/peerj.9670>
- Kistner, F., Slaney, L., Jewell, Z., Ben-david, A., and Alibhai, S. (2022). It’s Otterly Confusing-Distinguishing Between Footprints of Three of the Four Sympatric Asian Otter Species Using Morphometrics and Machine Learning. *Otter (Journal of the International Otter Survival Fund)*, **8**: 108 – 124. <https://www.wildtrack.org/assets/documents/IOSF-JOURNAL-2022-Kistner-et-al-108-125-1.pdf>
- Li, B. V., Alibhai, S., Jewell, Z., Li, D., and Zhang, H. (2018). Using footprints to identify and sex giant pandas. *Biological Conservation*, **218**: 83–90. <https://doi.org/10.1016/j.biocon.2017.11.029>
- Loshchilov, I., and Hutter, F. (2017). Decoupled weight decay regularization. *ArXiv Preprint ArXiv:1711.05101*. <https://doi.org/10.48550/arXiv.1711.05101>
- Ponce-Martins, M., Lopes, C. K. M., de Carvalho-Jr, E. A. R., dos Reis Castro, F. M., de Paula, M. J., and Pezzuti, J. C. B. (2022). Assessing the contribution of local experts in monitoring Neotropical vertebrates with camera traps, linear transects and track and sign surveys in the Amazon. *Perspectives in Ecology and Conservation*, **20**(4): 303-313. <https://doi.org/10.1016/j.pecon.2022.08.007>
- Sharma, S., Chee-, W., Kannan, A., Rama, S., Patah, P. A.-, and Ratnayeke, S. (2022). Identification of three Asian otter species (*Aonyx cinereus*, *Lutra sumatrana*, and *Lutrogale perspicillata*) using a novel noninvasive PCR- RFLP analysis. *Ecology and Evolution*, **12**, e9585. <https://doi.org/10.1002/ece3.9585>
- Shrestha, M.B., Shrestha, G., Reule, S., Oli, S. and Ghartimagar, T. B. (2022). Presence of Evidence and Factors Affecting Distribution of Eurasian Otter (*Lutra lutra*) in the Pelma River, Rukum East, Nepal. *Otter (Journal of the International Otter Survival Fund)*, **8**: 70–83. [https://otter.org/documents/journals/IOSF\\_Journal\\_Vol8\\_2022.pdf](https://otter.org/documents/journals/IOSF_Journal_Vol8_2022.pdf)
- Soe, M. M. (2022). Otter Records in Myanmar. *Otter (Journal of the International Otter Survival Fund)*, **8**: 43–50. [https://otter.org/documents/journals/IOSF\\_Journal\\_Vol8\\_2022.pdf](https://otter.org/documents/journals/IOSF_Journal_Vol8_2022.pdf)
- Stevens, S., Organ, J.F. and Serfass, T.L. (2011) Otters as Flagships: Social and Cultural Considerations. Proceedings of Xth International Otter Colloquium, *IUCN Otter Spec. Group Bull.* **28A**: 150 – 161 [https://www.iucnosgbull.org/Volume28A/Stevens\\_et\\_al\\_2011.html](https://www.iucnosgbull.org/Volume28A/Stevens_et_al_2011.html)
- Wägele, J. W., Bodesheim, P., Bourlat, S. J., Denzler, J., Diepenbroek, M., Fonseca, V., Frommolt, K. H., Geiger, M. F., Gemeinholzer, B., Glöckner, F. O., Haucke, T., Kirse, A., Kölpin, A., Kostadinov, I., Kühl, H. S., Kurth, F., Lasseck, M., Liedke, S., Losch, F., Wildermann, S. (2022). Towards a multisensor station for automated biodiversity monitoring. *Basic and Applied Ecology*, **59**: 105–138. <https://doi.org/10.1016/j.baae.2022.01.003>

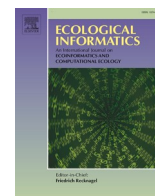
- Wäldchen, J., and Mäder, P. (2018). Machine learning for image based species identification. *Methods Ecol Evol.*, **9**: 2216–2225. <https://doi.org/10.1111/2041-210X.13075>
- Willi, M., Pitman, R. T., Cardoso, A. W., Locke, C., Swanson, A., Boyer, A., Veldthuis, M., and Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, **10**(1): 80–91. <https://doi.org/10.1111/2041-210X.13099>
- Yoxon, P. (2022). Conservation of Endangered Otters and their habitats in Malaysia through education and a Training Workshop. *Otter (Journal of the International Otter Survival Fund)*, **8**:14–25. [https://otter.org/documents/journals/IOSF\\_Journal\\_Vol8\\_2022.pdf](https://otter.org/documents/journals/IOSF_Journal_Vol8_2022.pdf)
- Yoxon, P., and Yoxon, G. M. (2017). *Otters of the world*. Whittles Publishing. ISBN-13 978-1849951296
- Zang, H.-X., Su, H., Qi, Y., Feng, L., Hou, R., He, M., Liu, P., Xu, P., Yu, Y., and Chen, P. (2022). Ages of giant panda can be accurately predicted using facial images and machine learning. *Ecological Informatics*, **72**: 101892. <https://doi.org/10.1016/j.ecoinf.2022.101892>

## RESUME

### **POUVEZ-VOUS RECONNAÎTRE UNE ESPÈCE A SON EMPREINTE ? - IDENTIFICATION DE TROIS DES QUATRE ESPÈCES DE LOUTRES SYMPATRIQUES D'ASIE DU SUD-EST À L'AIDE DE LA VISION PAR ORDINATEUR ET DE L'APPRENTISSAGE EN PROFONDEUR**

L'Asie du Sud-Est abrite quatre espèces de loutres, présentant toutes une tendance démographique à la baisse. Les quatre espèces de loutres d'Asie du Sud-Est peuvent coexister en sympatrie et figurent sur la Liste rouge des espèces menacées de l'Union Internationale pour la Conservation de la Nature (UICN). Il existe des lacunes dans les connaissances de l'aire de répartition et la taille des populations de ces espèces furtives, informations essentielles à la mise en œuvre de mesures de conservation efficaces. Les empreintes peuvent constituer un moyen rentable et non invasif de collecter des données pertinentes. Wild Track a développé un modèle de classification de la Technologie d'Identification des Empreintes (TIE) qui utilise des mesures basées sur des points de repère comme données de départ. Ce modèle est très précis et permet de distinguer trois des quatre espèces de loutres d'Asie du Sud-Est. Dans cette étude, nous proposons une approche basée sur l'apprentissage en profondeur qui automatise la classification des espèces en analysant la zone des points de repère de délimitation placés autour des empreintes. Le procédé réduit considérablement le temps de traitement des données et permet de se passer d'opérateurs hautement qualifiés pour la localisation des points de repère sur les empreintes.

Pour tester le modèle, 2.562 images comportant 3.895 empreintes annotées ont été utilisées, ce qui a abouti à une exactitude, une précision et un retour impressionnants de 99 % sur les groupes d'entraînement et de test. De plus, les performances du modèle ont été testées sur un nouveau groupe de 431 empreintes, qui n'ont pas été utilisées dans le processus de formation, et seulement 4 d'entre elles n'ont pas été bien identifiées, démontrant l'efficacité de l'approche proposée sur des données inconnues. Les résultats de cette étude confirment la viabilité d'utilisation d'une approche basée sur un modèle d'apprentissage automatique pour identifier avec précision les espèces de loutres grâce à leurs empreintes. Cette approche est à la fois fiable et rentable, ce qui en fait un outil attrayant pour les efforts de suivi et de conservation des loutres en Asie du Sud-Est. De plus, la méthode présente un potentiel d'application important dans les programmes communautaires de suivi scientifique citoyen. Des recherches plus approfondies pourraient se concentrer sur l'élargissement de la portée de l'étude en ajoutant à la base de données de formation des empreintes de loutres de Sumatra ainsi que d'espèces sympatriques autres que les loutres. En outre, cette étude suggère de développer un modèle de détection d'objets et de former de nouveaux modèles de classification qui



## Enhancing endangered species monitoring by lowering data entry requirements with imputation techniques as a preprocessing step for the footprint identification technology (FIT)

Frederick Kistner<sup>a,b,\*</sup>, Justus Tulowietzki<sup>c</sup>, Larissa Slaney<sup>b,d</sup>, Sky Alibhai<sup>b,e,f</sup>, Zoe Jewell<sup>b,e,f</sup>, Burim Ramosaj<sup>g</sup>, Markus Pauly<sup>c,h</sup>

<sup>a</sup> Institute of Photogrammetry and Remote Sensing (IPF), Institute for Technology Karlsruhe (KIT), Germany

<sup>b</sup> WildTrack Specialist Group (WSG)

<sup>c</sup> Department of Statistics, TU Dortmund University, Germany

<sup>d</sup> Institute for Life and Earth Sciences, School of Energy, Geosciences, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK.

<sup>e</sup> WildTrack, Durham, NC, United States of America

<sup>f</sup> Nicholas School of the Environment, Duke University, Durham, NC, USA

<sup>g</sup> Google, Zurich, Switzerland

<sup>h</sup> Research Center Trustworthy Data Science and Cybersecurity, University Alliance Ruhr, Germany

### ARTICLE INFO

#### Keywords:

Endangered species  
Footprint identification technology  
Non-invasive monitoring  
Imputation  
Missing values  
Random Forest

### ABSTRACT

Numerous species of Earth's biota are at risk of extinction and wildlife conservation is more important than ever. Reliable baseline data are essential for wildlife management to inform on the numbers and distribution of endangered species. A promising non-invasive and cost-effective method for monitoring endangered species is the Footprint Identification Technology (FIT). It lends itself to both conservation research as well as citizen science and can be combined with other data collection methods. FIT extracts and uses morphometrics from animal footprints to create geometric profiles that are analyzed in customized classification models. These can identify species, sex and individual. The ability to identify individuals can then be used to predict various population parameters including size, distribution, and growth rate. FIT has been developed and published for several species, but it requires high quality footprints. Perfect prints are not always easy to find in the field as various factors can influence their quality. In this paper, we demonstrate that geometrical profiles derived from poor quality footprints can be seen as datasets with missing values. Missing values are a common problem in various disciplines, and well-established strategies to impute missing values are widely available. We conducted two experiments to see whether such an approach could widen the application of the FIT method. The experiments were designed to test the hypothesis that population sizes can be underestimated when incomplete footprints are discarded from the data. We artificially introduced different proportions of missing values in datasets with geometric profiles of five different species for which FIT models have been published. We also analyzed a new dataset of geometric profiles of cheetah (*Acinonyx jubatus*) footprints not meeting the standard FIT requirements. We demonstrated that excluding incomplete footprints led to an underestimation of the known population. As an alternative to discarding footprints, we compared different imputation techniques as data pre-processing steps by comparing the performance of resulting FIT models.

When imputation was chosen instead, we could show that FIT models with imputed geometric profiles were not significantly less accurate in predicting individual ID or population size even with high rates of missing values. We believe that our findings can be generalized, and the results indicate that imperfect footprints can contribute to the robustness of the FIT method and that this approach is particularly applicable when few good-quality footprints are available. We therefore highly recommend including imputation of imperfect footprints as a data pre-processing step.

\* Corresponding author at: Karlsruhe Institute of Technology (KIT) Campus South Institute of Photogrammetry and Remote Sensing (IPF), 76128 Karlsruhe, Germany.

E-mail address: [frederick.kistner@kit.edu](mailto:frederick.kistner@kit.edu) (F. Kistner).

<https://doi.org/10.1016/j.ecoinf.2024.102676>

Received 17 July 2023; Received in revised form 5 June 2024; Accepted 7 June 2024

Available online 8 June 2024

1574-9541/© 2024 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 1. Introduction

Earth's biota is entering the 6th mass extinction in 4.5 billion years of history Ceballos et al. (2015). Species are disappearing at unprecedented rates due to human activities Ceballos et al. (2015); Tollefson (2019). This biodiversity loss threatens ecosystem services and necessitates urgent global action Ceballos et al. (2020). The UN's 2022 Kunming-Montreal Global Biodiversity Framework (GBF) and Goal 15 of the Sustainable Development Goals emphasize urgent actions to address biodiversity loss and extinction risk Convention on Biological Diversity (CBD) (2022); United Nations Department of Economic and Social Affairs (2023).

Reliable species demographic data is crucial for evaluating extinction risk and managing human-wildlife conflicts Conde et al. (2019); United Nations Department of Economic and Social Affairs (2023). However, comprehensive demographic data is available for only 1.3% of tetrapod species, impeding evidence-based conservation policies Conde et al. (2019). For animal ethics and to assure high-quality data, the use of non-invasive monitoring techniques should always be the preferred choice to collect such data Jewell (2013) as they neither disturb nor harm the animal and have little to no effect on animal behavior. Non-invasive methods include DNA-based approaches Bohmann et al. (2014); Thomsen and Willerslev (2015); Lefort et al. (2022), camera traps Agha et al. (2018), tracks and signs Liebenberg (2014); Elbroch and McFarland (2019); Brown et al. (2021); Grolms (2021); Rhyder (2021), and bioacoustic monitoring Penar et al. (2020); Darras et al. (2024), as detailed in a comprehensive overview by Zemanova Zemanova (2020).

This study focuses on improving population estimates using footprints through Footprint Identification Technology (FIT) Jewell et al. (2001). Footprints are a cost-effective method for monitoring elusive species and can provide consistent data for conservation planning Alibhai et al. (2017); Seidlitz et al. (2021); Ponce-Martins et al. (2022); Dorfman et al. (2023). Citizen science projects like iNaturalist and WildTrackAI enable users to collect and share images of animal footprints, significantly boosting available data for conservation Evans (2023); WildTrack (2023).

### 1.1. Related work

#### 1.1.1. Footprints and tracking as data source for animal conservation science

Tracking animals is as old as humanity and may, in fact, be the origin of science Liebenberg (2013). In early hunter-gatherer cultures, this ability gave successful hunters access to protein-rich foods. This skill is no longer widespread in modern societies. In science, animal tracks are often only used for presence-absence studies of species, although it is possible to get much more information from them.

A wide array of scientific studies has demonstrated that footprints can provide valuable information for modeling and predicting various response variables of species of conservation concern. For instance, footprints have been utilized to distinguish between sympatric otter species Kistner et al. (2022, 2023) and to differentiate trackplate prints of elusive carnivores like fisher (*Pekania pennanti*) and Pacific marten (*Martes caurina*) originating from trackplate stations in the southern Sierra Nevada Tucker et al. (2024). Trackplates were also used often in pest control operations, and models using footprints were trained to distinguish between different small mammal and insect species Hasler et al. (2004). Additionally, such data have been employed to predict gender, e.g., in pine martens (*Martes martes*) Zalewski (1999), Bengal and Amur tigers (*Panthera tigris tigris* and *Panthera tigris altaica*) Sharma et al. (2003); Alibhai et al. (2022), and mountain lions (*Puma concolor*) Alibhai et al. (2017). Footprint analysis has also facilitated the modeling and prediction of age classes, as demonstrated with giant pandas (*Ailuropoda melanoleuca*) Li et al. (2018). Moreover, several successful attempts have been made to re-identify individual animals across

different taxonomic families, including black and white rhinos (*Diceros bicornis* and *Ceratotherium simum*) Jewell et al. (2001); Alibhai et al. (2008), tigers (*Panthera tigris*) Sharma et al. (2005), tapirs (*Tapirus terrestris*) Moreira et al. (2018), and cheetahs (*Acinonyx jubatus*) Jewell et al. (2016). These methods have proven effective in estimating the minimum population sizes of targeted species, a finding corroborated by field studies Alibhai et al. (2022); Moreira et al. (2018); Jewell et al. (2020). Various classification techniques have been employed in these studies, including stepwise discriminant analysis Palma and Gurgel-Gonçalves (2007); De Angelo et al. (2010); Tucker et al. (2024), Principal Component analyses combined with a simple Naive Bayes classifier Hasler et al. (2004), Decision Trees and XGBoost Kistner et al. (2022), Random Forest Awais et al. (2018), Convolutional Neural Networks Kistner et al. (2023), geometric morphometrics using 3D photogrammetry for lion (*Panthera leo*) paws and pads Marchal et al. (2017), and the Footprint Identification Technology (FIT) Alibhai et al. (2008); Jewell et al. (2016); Alibhai et al. (2017); Moreira et al. (2018); Li et al. (2018); Kistner et al. (2022); Alibhai et al. (2022); Tucker et al. (2024).

FIT is a software package in JMP® statistical software with an integrated landmark labeling tool, customized classification models, and report templates, which will be described in more detail in the methods section. It has been documented as an accurate and cost-effective wildlife monitoring tool for conservation, providing valuable and reliable data for policymakers to develop conservation strategies Pimm et al. (2015).

FIT analyzes images of footprints following a standardized protocol. A species-specific, landmark-based feature extraction script calculates morphometric measurements of distances, angles, and areas of the footprint image Jewell et al. (2016). These measurements are the foundation of all subsequent modeling steps that are described in the methods section below. Fig. 1 illustrates a cheetah (*Acinonyx jubatus*) footprint with landmarks and shows the auto-generated geometrical profile of the footprint.

Working with footprints has one obvious limitation: the substrate must permit a clear foot impression, as poor-quality or incomplete prints provide insufficient details. FIT requires a continuous set of 6–8 high-quality footprints (a trail) from the same foot for individual identification Alibhai et al. (2008). Additionally, a field study on cheetahs in Namibia demonstrated that FIT performs more reliably when more prints from longer trails, separated into sub-trails, are included. Conversely, the performance worsens when prints below a certain quality threshold are used Laity (2015). In studies using the FIT method or other classification approaches that utilize geometric profiles of footprints, it is commonly advised to either refrain from collecting or discard prints that do not meet predefined quality criteria. Footprints found in the field are often imperfect and quite often do not look like the ones from textbooks. The quality of prints can vary significantly between different sites Laity (2015).

Fig. 2 shows examples of footprints of various qualities from two species included in our analyses, white rhinos (*Ceratotherium simum*) and mountain lions (*Puma concolor*). A multitude of factors, such as the substrate, the animal's gait, changes in direction, weather conditions, or objects on the ground, influence the respective shape and depth of a footprint Grolms (2021). The quality of prints can vary significantly between different sites Laity (2015).

Of course, it is reasonable to select and compare the footprints that show the most detail. However, there is currently no recommendation on how to proceed when there are not enough high-quality footprints available in a trail. In studies that utilize footprints as input data, it is commonly advised to either refrain from collecting or to discard prints that do not meet predefined quality criteria.

We hypothesize that, instead of discarding imperfect footprints, exploring potential techniques to lower the quality requirements of footprints could broaden the applicability of footprint-based monitoring methods such as FIT, making it a cost-effective and reliable conservation tool.

1.1.2. Missing values: Origins and handling strategies

Missing data can come from three different sources, which are responsible for why some data, in our case landmark points, is not available. These sources are called missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR) mechanisms Garcia-Laencina et al. (2010). MCAR means that the missing data is random and not related to any other observed or unobserved data. MAR means that the missing data is related to other observed data, while NMAR means that the missing data is related to both observed and unobserved data. Understanding the source is important for selecting appropriate methods to handle missing data Garcia-Laencina et al. (2010). One approach to handling missing values is imputation, which fills in missing instances with reasonable values that are based on the observed components. Potential dependencies among the different variables in the dataset are taken care of by using Bayesian or Machine Learning (ML)-based algorithms Garcia-Laencina et al. (2010). Recent research in statistical missing value imputation has been focusing on ML-based methods such as tree-based Random Forest and variants of it Stekhoven and Bühlmann (2012); Mayer (2021); Buczak et al. (2023); Schwerter et al. (2024). Moreover, the Multivariate Imputation by Chained Equations (MICE) approach has been shown to be effective in accurately filling in missing data and producing reliable statistical results Doove et al. (2014); Ramosaj and Pauly (2019); Ramosaj et al. (2020, 2022); Thurrow et al. (2021, 2024). The techniques applied in this study are further described in the Methods section.

1.2. Aims and significance

In this study, we examined subpar and incomplete footprints—specifically, their geometric profiles—from the perspective of a data scientist. An incomplete footprint produces a data row with missing values, which can be analyzed using statistical techniques designed to impute these missing entries. To the best of our knowledge, this is the first study that combines well-established missing value imputation

techniques with morphometric data derived from animal footprints. We assessed the effectiveness of various missing value handling strategies by evaluating the performance of established FIT models in terms of their ability to re-identify individual animals and estimate population sizes. We hypothesize that our proposed integration of these methods could significantly expand the application spectrum of non-invasive animal tracking as a cost-effective and scientifically robust conservation tool.

2. Methods

In this study, we conducted two experiments: the first with artificially introduced missing values, and the second with real data to further evaluate our findings. Our intention was to explore:

- a) to what extent discarding of incomplete footprints can lead to an under- or overestimation of the investigated true population size;
- b) if well-established missing value imputation methods can be used as an alternative to omitting incomplete footprints;
- c) which imputation methods as a pre-processing step might lead to improved FIT model performance.

In the first experiment, the focus lies on missing values from the MCAR mechanism. Although in practice there is no direct control over the mechanism, in our experiment we randomly and artificially generated missing instances to analyze their impact on the final classification. Statistical analysis with partially observed data under the MCAR mechanism allows several analyses to be inferentially valid. The reason for this is that missing instances do not need to be modelled precisely, instead, they occur randomly, independently of the data itself.

We then evaluate our findings in a second experiment, where we analyzed a new dataset of tagged geometrical profiles originating from images of various qualities. Therefore, we created a new dataset of cheetah footprints of different qualities. Here, the quality of each landmark point was assessed beforehand by an expert operator, so that

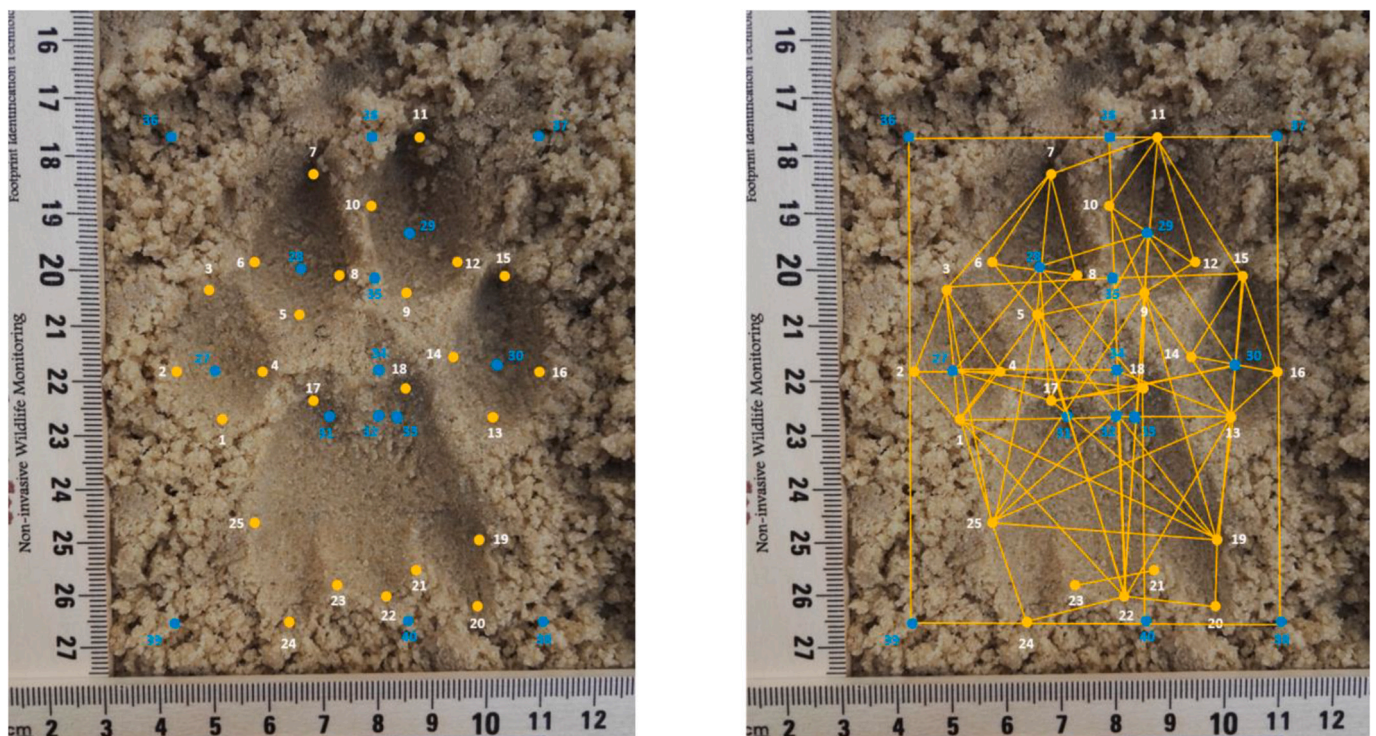


Fig. 1. Left hind footprint of a Southern African cheetah (*Acinonyx jubatus jubatus*). The FIT protocol for this species comprises of 25 manually placed landmark points (yellow) and 15 auto-generated derived points (blue). The image on the right is an illustration of the geometrical profile of this footprint. It comprises of distances, angles and areas, which were calculated from the landmark points with the addition of two scaling points placed on a metric ruler. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

real missing values, which were potentially not missing at random, could be analyzed in this dataset. We then applied and compared the same imputation techniques as in the first experiment and evaluated the performance of FIT predictions for this new dataset.

### 2.1. Datasets

For the first experiment, we used datasets containing only high-quality footprints from five different endangered species for which the FIT method has been developed, published, and described in scientific journals Alibhai et al. (2008); Gu et al. (2014); Jewell et al. (2016); Alibhai et al. (2017); Li et al. (2018); Alibhai et al. (2022). Each dataset comprised a known number of animals with a known number of footprints per individual. The footprints of individuals were grouped into trails. A trail is a series of footprints along an unbroken line that was made by one individual. Each species has a defined number of morphometric variables containing distances, angles, and areas that are automatically computed from a species-specific set of manually placed landmark points. These manually placed landmarks were used to add additional rule-based species-specific derived points as illustrated in

Fig. 1. An overview can be found in Table 1.

The digitization of animal footprint images into numerical datasets was achieved through a systematic process. Initially, high-quality images of animal footprints were collected. These images were then subjected to precise preparation techniques including resizing, rotating to ensure standard orientation, and scaling based on a known reference to guarantee uniformity and accuracy across all samples.

Critical to this process was the manual identification and marking of anatomical landmarks on the footprints by species experts, utilizing a custom package developed for JMP® statistical software. This meticulous procedure involved placing landmarks on specific locations, such as toes and pads, tailored to the species under study to accurately capture the unique morphology of each animal's track.

Subsequent to the initial marking of landmarks by experts, JMP's species-specific feature extraction scripts, JMP's jsl coding language, were employed to automatically generate additional derived points. This significantly enriched the dataset with detailed geometric measurements like distances, angles, and areas, thereby ensuring a rigorous, consistent, and replicable method for transforming visual footprint data into a structured numerical format. This methodology is finely tuned for



Fig. 2. Examples of footprints from white rhino (*Ceratotherium simum*) top and mountain lion (*Puma concolor*) bottom. Morphometric data of both species was used in the analyses. The quality of the prints is increasing from left to right. Certain parts of the footprints in the left and middle images are unclear and therefore not possible to process for operators with limited experience. Images on the left are discarded following current FIT protocols. Images in the center can be processed but require the researcher to possess species-specific anatomical knowledge. The images on the right fulfill the FIT image criteria and can be analyzed by a less experienced operator.

**Table 1**

Overview of the dataset of geometrical profiles used for the study. The first five datasets originate from previous FIT publications. The dataset of the South African Cheetah is a subset of a yet unpublished study that has been enriched with geometrical profiles of lower-quality footprints for the second experiment of our study. Each dataset contains multiple individuals per species and each individual is represented by different footprint images that each generate a unique geometric profile. Trails used for FIT modeling are groups of geometrical profiles from usually 6–8 footprint images that were found in line. A species-specific set of automatically created measurements (distances, areas, and angles) is derived from the X-Y-coordinates of geometrical landmarks. Additional information regarding the datasets utilized in this study can be found in the literature cited.

Species (Latin name)	No. of animals	No. of footprints and trails	No. of landmarks	No. of measurements
Amur tiger (c.f. <a href="#">Gu et al. (2014)</a> ; <a href="#">Alibhai et al. (2022)</a> ) ( <i>Panthera tigris altaica</i> )	44	605 / 85	25	128
cheetah (c.f. <a href="#">Jewell et al. (2016)</a> ) ( <i>Acinonyx jubatus</i> )	38	781 / 109	25	136
giant panda (c.f. <a href="#">Li et al. (2018)</a> ) ( <i>Ailuropoda melanoleuca</i> )	42	523 / 77	25	124
mountain lion (c.f. <a href="#">Alibhai et al. (2017)</a> ) ( <i>Puma concolor</i> )	35	535 / 77	25	133
white rhino (c.f. <a href="#">Alibhai et al. (2008)</a> ) ( <i>Ceratotherium simum</i> )	41	1273 / 159	13	112
cheetah (experiment 2) ( <i>Acinonyx jubatus</i> )	17	231 / 33	25	117

analyzing the distinctive characteristics of various animal species' footprints, thereby addressing the concern raised regarding the digitization process.

Within a footprint image, any landmark could be missing, and we assume for these cases that there is no systematic underlying missing pattern, thus we assume MCAR for the first experiment. For datasets initially lacking X-Y coordinate data, a workaround was implemented by focusing on the derived morphometric measurements directly. For consistency, we therefore only used the morphometrics instead of the X-Y-coordinates across all datasets. If the random generator selected a landmark point as missing, we multiplied it with a matrix that deleted all morphometric variables directly connected to this landmark point. The datasets analyzed in this study might slightly differ from the ones used in the publications referred to above, as over time, several new animals have been added to the FIT database.

For the second experiment, we engaged with a dataset of cheetah (*Acinonyx jubatus*) footprints that were processed by an operator experienced in this specific area. Unlike the first experiment, which utilized high-quality images, this dataset included images of lesser quality that would normally be excluded based on the FIT standard protocol. Due to the limited availability of lower-quality images, we did not differentiate whether missing features were the result of distorted footprints or general image quality issues, such as resolution and lighting. The inclusion of these images allowed us to explore the robustness of our methodology in handling data of varying quality.

All landmarks within each footprint image were meticulously rated according to their visibility and clarity: 0 = clearly identifiable, 1 = unclear, and 2 = fully missing. This rating system was crucial, as landmarks rated with a 1 were visible to the experienced operator but might

pose identification challenges to less experienced individuals. In instances where a landmark was completely missing, the operator employed an educated guess approach, leveraging the geometrical shape of the footprint and the relative positions of other identifiable landmarks to infer the missing data points.

This nuanced approach enabled us to generate two sub-datasets: one excluding all landmarks rated as 2 (fully missing), and another further excluding those rated as 1 (unclear). This stratification was instrumental in evaluating the impact of landmark clarity on the accuracy of our footprint analysis, thereby testing the limits of our digitization process under less-than-ideal conditions.

## 2.2. Imputation and prediction models

In this paper, the methods used were both intuitive (omitting missing values; mean imputation) and machine-learning-based. These procedures and the FIT method are briefly presented below.

### • Control Group

In the first experiment, the control group consisted of the unmodified datasets of each species. In the second experiment, we defined the control group as the geometric profile extracted by our experienced operator.

### • Delete Case

All data rows containing missing values were removed.

### • Mean imputation

For metric features, the mean value of the corresponding variable is computed among the observed components and imputed for the missing instances at that variable.

### • missForest

missForest is a Random Forest-based algorithm developed by ([Stekhoven and Bühlmann, 2012](#)), which imputes missing values iteratively using Random Forests trained on subsets of the dataset. In contrast to the standard Random Forest methodology, as introduced by ([Breiman, 2001](#)), where the focus is primarily on predictive modeling or classification tasks, missForest uses the Random Forest algorithm for the specific task of imputation. In missForest, each variable containing missing values is treated as a target variable. For every target variable, a dedicated Random Forest model is trained, utilizing the remaining variables as predictors. This targeted model is then employed to predict missing values within subsets of the dataset. This procedure is repeated with the updated imputed missing values until a predefined stopping criterion is reached ([Stekhoven and Bühlmann, 2012](#)). The iterative approach of missForest allows it to progressively enhance its imputation accuracy.

### • missRanger

missRanger is implemented similarly to missForest but additionally allows the incorporation of some sampling uncertainty. This is done by integrating predictive mean matching into the Random Forest model. Instead of predicting missing values using a Random Forest, missing values are imputed by randomly drawing among  $k$  observed values that are closest to the final Random Forest prediction ([Mayer, 2021](#)).

### • Multivariate imputation using chained equation (MICE)

MICE is a multiple imputation technique developed by ([Rubin, 2004](#)). This method comprises three essential steps: generating multiple

imputations, analyzing the imputed data, and pooling the analyzed results in the form of a final statistic, see (Van Buuren and Oudshoorn, 1999). At its core, the MICE procedure relies on a Gibbs Sampling mechanism, where missing values are drawn from univariate distribution models conditioned on other observed features. In an iterative approach, the method approximates the conditional distribution function using a suitable number of iteration steps to simulate the burn-in period of a Markov chain. After obtaining sufficiently stable random draws, MICE enables the inclusion of multiple imputations since the variation between the different imputation values originates from the independent draws acquired through the (stable), conditional distribution Van Buuren and Groothuis-Oudshoorn (2011b).

According to Van Buuren and Groothuis-Oudshoorn (2011b), different variants of the MICE procedure are implemented which are using the framework of a Gibbs Sampler as described above. Regarding the context, we focus on two built-in imputation techniques in this paper: predictive mean matching (PMM) and Random Forest with PMM (RF). Using PMM assumes a Bayesian linear regression with suitable priors for every model parameter, where the normality assumption is set as the core distribution for the regression coefficients.

Missing values are then imputed by extracting random draws from the model. The final imputation is based on observed values that are closest to the extracted random draws. Setting the number of potential candidates to  $k$  corresponds to the number of closest observed values, from which one is randomly drawn and is defined as the imputation value. A completely different option is the usage of Random Forests (RF) for imputation in MICE Doove et al. (2014). The latter is similar to missForest but allows for additional variations by initially filling missing instances in the dataset with random draws, and then fitting Random Forest models iteratively throughout the different variables. Imputation is then conducted by randomly drawing observations among the leaf nodes.

After generating potentially different datasets, all of them are analyzed using standard statistical procedures. Then, they are combined into one final statistic, which depends on the complete case method used in every dataset Van Buuren and Oudshoorn (1999). According to Rubin (2004), the imputation of  $m$  potentially different values for missing instances reflects the uncertainty arising from the presence of missing values. Pooling them into one final statistic takes this uncertainty effect into account Rubin (2004). In this paper, we use  $m = 3$ .

### 2.2.1. Footprint identification technology (FIT)

FIT uses a customized model that classifies animal trails by employing a robust cross-validated discriminant analysis. In an iterative approach, each pair combination of trails is once held back as the test set, with the rest of the remaining trails used as the training set model. Using stepwise forward regression, the top explanatory variables are selected based on their F-ratio Alibhai et al. (2008). Trails are mapped in two-dimensional space using the first two canonical variates. Multivariate least-square means, from now on referred to as centroid values, and their respective 95% confidence ellipses are then plotted for each trail. The distance between the centroids is relative, depending on the (relative) position vector of the centroids and the matrix of the within-group variation. Therefore, adding or removing animals from the test set alters the position of the centroid values and its confidence ellipses Li et al. (2018).

To face these challenges, a modification first proposed by Jewell et al. (2001) was applied. The centroid plot technique was used on a pairwise basis, by comparing only two trails at a time with an additional third group, the Reference Centroid Value (RCV). The RCV acts as a geometrical anchor of inter-canonical space, as it comprises the other known individuals and is several magnitudes larger than the compared trails of the pairwise comparisons. This leads to a stabilization of the location of any test group in the canonical space Jewell et al. (2001).

The output of this pairwise comparison model is a confusion matrix and a matrix that stores the Euclidean distances between the trails. In

previously cited FIT publications, positives in the confusion matrix were defined as Self, and negatives as non-Self classifications and the overall classification accuracies were reported in most FIT studies Alibhai et al. (2008).

In total, we had 540 individual animals over all 6 datasets and an average of 2.4 (sd = 0.8) trails per animal. Therefore, the number of pairwise comparisons of trails originating from different animals exceeds the number of pairwise comparisons of trails originating from the same animal by several orders of magnitudes, making it an imbalanced classification task. However, we consider the ability to distinguish between different individuals equally important as the ability to identify several trails originating from a single individual.

### 2.2.2. Classification and prediction accuracy

To address this imbalance while maintaining equal classification importance, other than using the overall accuracy, we used the Balanced Classification Rate (BCR). The BCR is defined as the average of the True Positive Rate (TPR) and the True Negative Rate (TNR), as a target metric and has a Range from 0 to 1 Tharwat (2020).

$$BCR = \frac{TPR + TNR}{2}$$

The second output of a FIT model run is a cluster dendrogram. The dendrogram is derived from the distances between the trails from the pairwise comparison model. At a defined cut-off distance, the number of formed clusters represents the predicted number of animals. Since we know the true number of clusters, we can define a Dendrogram Error Rate (DER) as the absolute difference between the Predicted Number of Animals (PNA) and the True Number of Animals (TNA) divided by the TNA which is equivalent to the number of animals in the control set, i.e.

$$DER = \frac{|PNA - TNA|}{TNA}$$

We obtained rank values for the BCR and DER for each species and for each fraction  $p$  of missing landmarks. This resulted in 30 rankings for each method and target metric. We then summed the rankings for both the BCR and DER and calculated the average to obtain a final rank-based performance evaluation metric with a range of 1–6. Lower numbers on this scale indicate the better overall performance of the FIT model.

### 2.3. Study design

During the study, two different statistical software were utilized. For the generation and imputation of missing values, the statistical software R-version 3.6.3 R Core Team (2020) was used together with the packages missMethods Rockel (2020), missForest Stekhoven (2013), missRanger Mayer (2021) and MICE van Buuren and Groothuis-Oudshoorn (2011a) as imputation algorithms. FIT is implemented as a package in JMP® statistical software SAS Institute Inc (2021). We used version 16.2pro for all FIT analyses.

For the first experiment, we simulated missing landmarks by artificially introducing different percentages  $0 < p < 1$  of missing values. We used the function prodNA() from the missForest package in R to simulate missing values under the MCAR mechanism. Then, we generated missing values in the morphometric variables. These variables are defined, and scaled measurements are derived from the corresponding X-Y coordinates of the landmark points. To analyze the effect of different missing rates on the prediction accuracy, several missing rates  $p \in \{0.01, 0.025, 0.05, 0.10, 0.15, 0.20\}$  are used for all species datasets. This leads to  $6$  (number of  $p$ ) \*  $5$  (number of species) =  $30$  cases to be considered in the first experiment.

We then compared the number of animals remaining in the dataset after the deletion of missing cases with the number of individuals in the control dataset. All data rows left after the deletion step were then included in the FIT analyses and its performance was evaluated as described above.

For each case containing missing values, we then compared five different imputation methods in addition to the delete case with a species-specific control that did not contain any missing values. As imputation methods, we chose mean imputation, MICE PMM ( $m = 3$ ), MICE RF ( $m = 3$ ), missForest and missRanger imputation. To ensure a standardized comparison, we used the default settings in R for all methods. In the case of the Giant Panda dataset, MICE RF returned errors when permuting the variable *Ar.1.V17* due to too small variance ( $< 0.0001$ ). To obtain imputations for this variable using MICE RF, we multiplied the observations of the variable with a factor of 3 before imputation to increase variance and transformed everything back after imputation.

For all FIT models analyzed in this study, we used the RCV of the control that does not contain any imputed missing values and is consistent among any treatments of a species.

The delete case strategy can be seen as the currently deployed strategy of a FIT analysis, and we have chosen the mean imputation as a baseline to which more advanced imputation strategies could be compared.

The classification accuracy of a species FIT algorithm is dependent on three model parameters, which are summarized for each species in Table 6 in the appendix.

These parameters are the number of selected variables in the model, the size of the confidence ellipsoid (counter probability), and the cut-off threshold value of the distances between the geometrical mean of the trails (Wards-cut-off distance). No hyperparameter tuning has been conducted and in order to assure uniformity, all models of the same species were executed with the default FIT parameters on the full dataset.

For the second experiment, we applied the same approach to real missing data points.

In summary, we analyzed 37 FIT runs per species and compared 185 model outputs in the first experiment and a total of 13 FIT runs for the second experiment.

### 3. Results

#### 3.1. a) Effects of deleting missing values on estimating the true population size

Firstly, the effect of the delete case scenario on the respective population sizes was investigated. This simple descriptive analysis is shown in Fig. 3, in which the percentage of missing values is plotted against the respective relative proportions of missing individuals after the deletion of observations with missing values. This figure contains plots for individual species and a combined plot for all datasets (All Species). As can be observed in Fig. 3, 5% of randomly introduced missing landmarks bring an underestimation of the true population size in three out of the five investigated datasets. 10% of missing landmarks resulted in an underestimation of the true population size in all datasets. There is a linear relationship between the percentage of the introduced missing landmarks and the number of missing individuals. For the white rhino dataset, the slope of the regression was smaller in comparison to the other species.

#### 3.2. b) + c) Performance of FIT models for different imputation strategies

Fig. 4 shows the box plots of the Balanced Classification Rate (BCR),

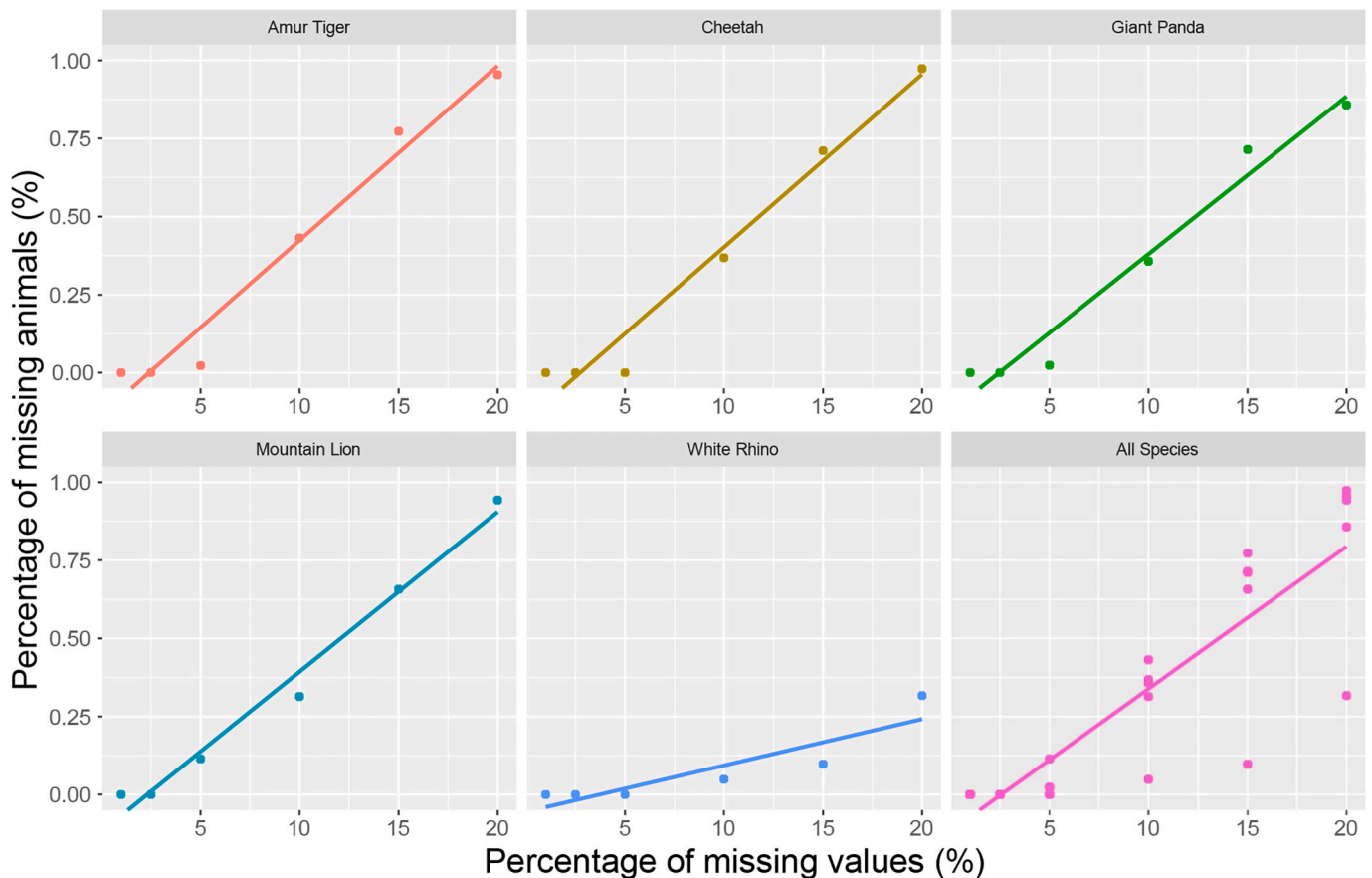
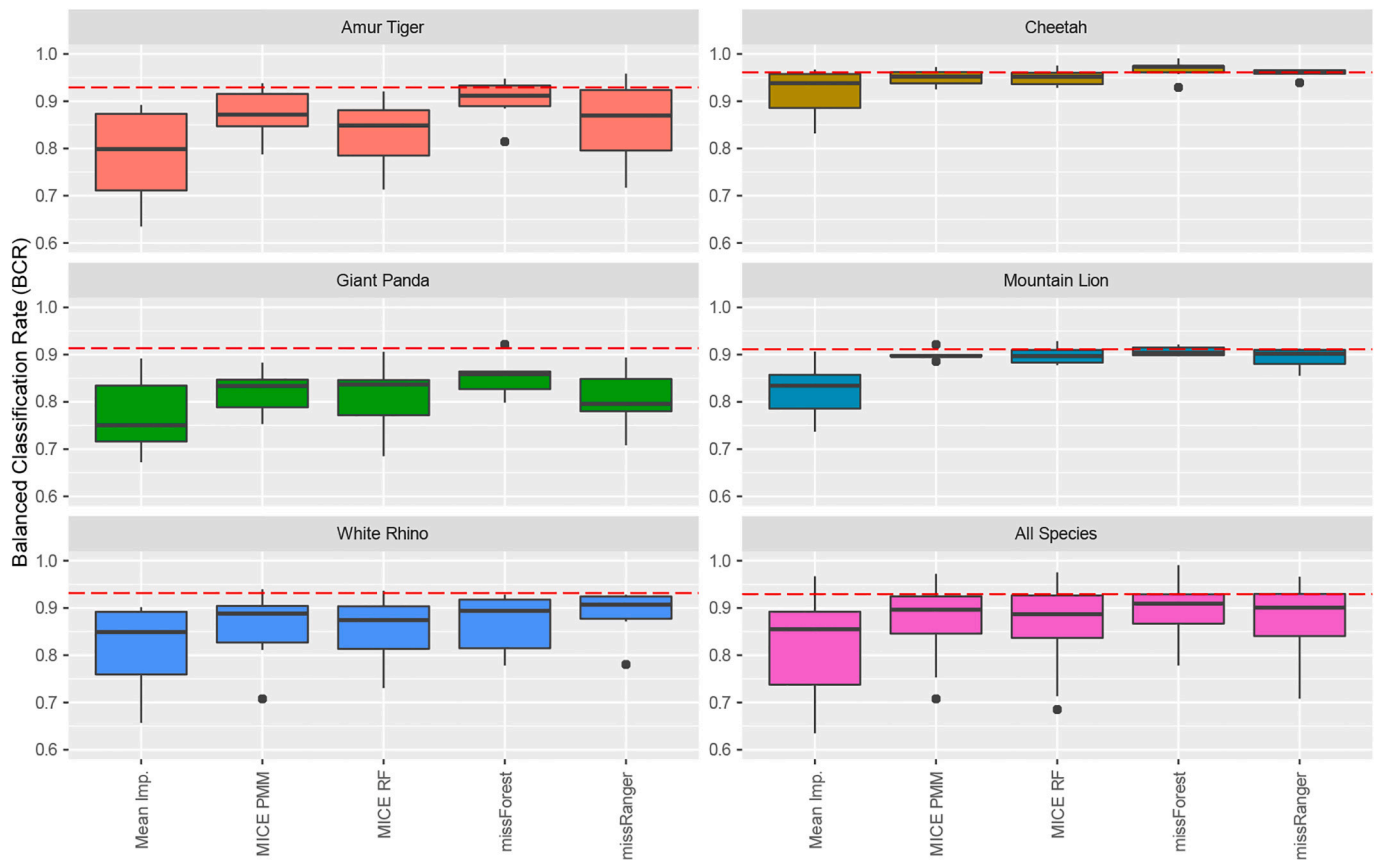


Fig. 3. Relationship between the percentage of missing values and the relative number of individuals missing in the dataset. There seems to be a linear relationship between the percentage of missing values and the relative number of missing individuals. This is consistent for all five species although the linear increase for the white rhino is lower.



**Fig. 4.** This figure shows the boxplots for the Balanced Classification Rate (BCR) of the five imputation methods per species and across all five species. The boxplots of each animal are composed of the BCR values of the six different artificially introduced missing value mechanisms. The value of the respective control group is indicated by the red line. For the results of all animals, the boxplots consist of all results and the red line corresponds to the average BCR value of all control groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

while Fig. 5 shows the Dendrogram Error Rate (DER) for the individual species and across all five species. The respective box plots of the imputation methods are based on the corresponding results of the six considered percentages of missing landmarks. The box plot of the species-specific control contains only one value.

The highest BCR was obtained by missForest with 0.90, only 0.03 points lower than the control group. The lowest DER with 0.07 was achieved by a simple mean imputation. It is noticeable that this DER is 0.01 points smaller than the control group. Among the imputation methods, missRanger had the lowest rank with a score of 2.72 and therefore the best overall performance. In general, all imputation methods performed better than the Delete Case method, with more advanced methods achieving lower average ranks than a simple mean imputation (as seen in Table 2).

Comparing the results between species, it is noticeable that the box plots of all imputation methods for White Rhino are larger compared to the other species. The adverse effect can be observed in the delete case, where the box plot for the white rhino is the smallest. Accordingly, the results differ more for the various proportions of missing landmarks. Furthermore, the results scatter more when deleting the footprints containing missing values, in contrast to the results when using imputation methods. These findings relate to both BCR and DER.<sup>1</sup>

To further investigate the influence of the proportion of missing values on the performance of the different methods, we plotted the

<sup>1</sup> Note that the box plot of the delete case was excluded after the revision on the recommendation of the reviewer to more clearly demonstrate the differences between the methods.

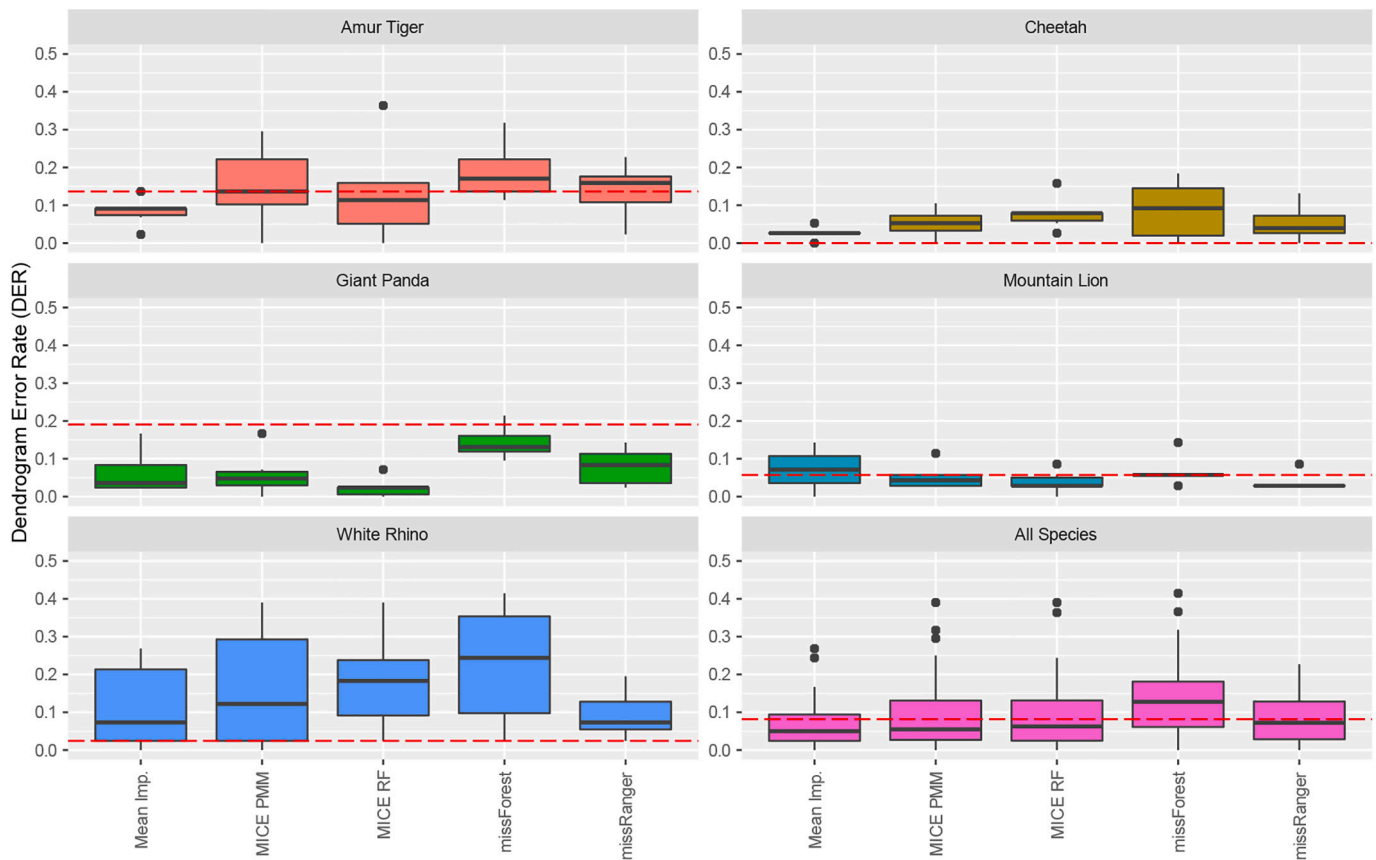
results in Fig. 6. The blue regression line shows that the Balanced Classification Rate (BCR) for all methods decreases as the proportion of missing values increases, with the lowest slope for the missForest method and the highest slope for the delete case. As the percentage of missing values increased, the Dendrogram Error Rate (DER) increased for all methods except for missRanger, which was almost unaffected and even had a slightly negative slope. These performance metrics indicated that the delete case method was the most negatively affected by an increase in the percentage of missing values. The BCR decreased at a higher rate for the delete case compared to other imputation methods, and the DER also increased at a higher rate for the delete case. This suggests that the delete case method was more sensitive to missing data compared to the imputation methods.

### 3.3. Results from uncurated cheetah footprints

In this second experiment, we investigated the effect of different imputation strategies on real data originating from footprint images of various image qualities with rated landmark qualities. For our uncurated data, a distinction is made between clear, missing and unclear landmarks. In total, 0.7% of all landmarks are marked as missing and 19.5% as unclear. This results in 2.0% and 41.2% of footprints being discarded respectively.

In the first case, the removal of only clearly missing landmark points does not lead to an animal removal from the dataset. In contrast, the exclusion of all unclear landmarks leads to the removal of six individuals, leaving only 61.0% of animals in the study.

Table 3 shows the three top missing landmarks of our real data cheetah study. Landmark 20 was affected the most, missing in 6% of all



**Fig. 5.** This figure shows the boxplots for the Dendrogram Error Rate (DER) of the five imputation methods per species and across all animals (All Species). The boxplots of each animal are composed of the BCR values of the six different artificially introduced missing values. The value of the respective control group is indicated by the red line. For the results of all animals, the boxplots consist of all results and the red line corresponds to the average BCR value of all control groups. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 2**

Overview of the results of the two target FIT metrics Balanced Classification Rate (BCR) and Dendrogram Error Rate (DER). The table shows the mean of these metrics with their standard deviation (SD) of the six imputation methods over the five different species datasets, each with six different percentages of artificially introduced missing values. Furthermore, the results of the control groups are shown. The last column (Avg. Rank) shows the average rank value of the BCR and DER results compared to the other imputation methods. The smaller this average rank value, the better the overall performance of the method.

Imputation method	No. of Datasets	BCR (SD)	DER (SD)	Avg. Rank
Control	5	0.93 (0.02)	0.08 (0.06)	–
Delete Case	30	0.51 (0.34)	0.51 (0.38)	5.63
Mean Imp.	30	0.82 (0.09)	0.07 (0.07)	3.51
MICE PMM	30	0.84 (0.09)	0.09 (0.08)	2.92
MICE RF	30	0.80 (0.10)	0.11 (0.10)	3.08
missForest	30	0.90 (0.06)	0.14 (0.10)	3.15
missRanger	30	0.88 (0.07)	0.08 (0.06)	2.72

footprints and unclear or missing in 50% of all footprints.

It is noticeable that all the percentages of missing and/or unclear values were predominantly found in the metatarsal part of the footprint and fewer in the toe areas. See Table 4 for details.

The performance of the FIT models for the different imputation strategies was comparable to the first experiment and the results are shown in Table 5. In the first scenario the deletion of clearly missing landmarks lead to a higher BCR than the control for all imputation strategies, including the delete case. The highest BCR was achieved for missForest and missRanger. The DER of the first scenario was 0 and therefore equal to the control for MICE RF and missForest. missForest

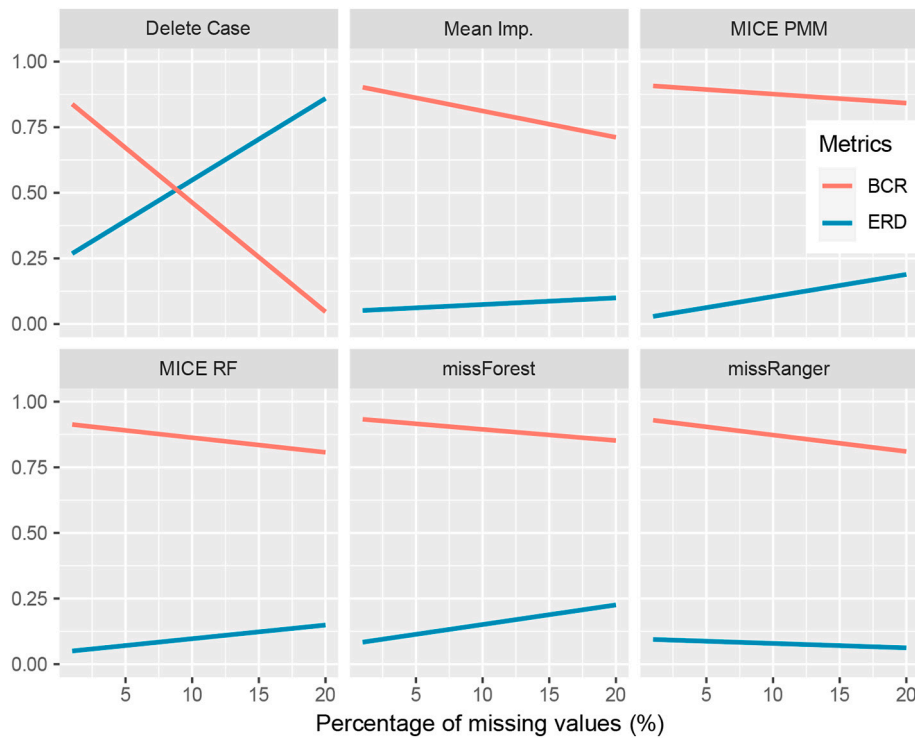
outperformed the results of the control group in this scenario. For the second scenario, MICE PPM had the highest BCR with 0.88 and the second lowest DER with 0.06.

According to the results of the second experiment, the delete case strategy had the worst performance among all the imputation strategies in both target metrics. This poor performance was observed consistently across all the six missing value handling strategies. However, the negative impact of the delete case strategy was more pronounced in the scenario where both missing and unclear landmarks were removed from the dataset.

#### 4. Discussion

This study introduces a novel combination of well-established methods that enhances a footprint dataset by avoiding the removal of incomplete footprints. It demonstrates that discarding imperfect footprints can lead to an underestimation of the true population for Footprint Identification Technique (FIT) prediction models. Instead, by imputing the missing values using as a preprocessing step to FIT, the study shows the following:

1. Imputing missing values allows us to preserve the true population size of the sample, regardless of the sample's representativeness of the entire population, rather than discarding incomplete data points.
2. The best imputation methods yield comparable model performance to the control group.



**Fig. 6.** Relationship between the Balanced Classification Rate (BCR) (red regression line) and the Dendrogram Error Rate (DER) (blue regression line) and the percentage of missing values artificially introduced in the five different species datasets for the six different missing value handling strategies. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 3**

Top three most missing landmark points, divided into not missing (= 0), unclear (= 1) and missing (= 2) and the area of the foot (illustrated in Fig. 1 above).

Landmarks	0	1	2	Area of foot
L20	115	101	15	Metatarsal
L24	126	97	8	Metatarsal
L18	145	81	5	Metatarsal

**Table 4**

Relative distribution of affected landmarks categorized by its region in a footprint, divided into not missing (= 0), unclear (= 1), and missing (= 2).

Area of foot	0	1	2
Metatarsal	27.88%	70.93%	75.56%
Toe	72.12%	29.07%	24.44%

3. The best imputation methods can handle a high proportion of missing values while maintaining overall high performance in the target metrics.

The novelty of the study lies not in introducing a new method but in the combination of these well-established methods.

To further enhance the Footprint Identification Technique (FIT) method, we recommend collecting imperfect footprints when high-quality footprints are insufficient during field surveys. Instead of discarding incomplete prints, we advocate using missing value imputation as part of the data preprocessing pipeline. This approach preserves valuable data and improves the accuracy of FIT predictions when sub-quality footprints are included. Our findings are supported by a study using deep learning models to classify three sympatric otter species, where a model trained on images of all qualities was only marginally outperformed by a model trained on high-quality prints, highlighting the potential of utilizing lower quality data effectively Kistner et al.

**Table 5**

Results of the second experiment were presented in a table with two scenarios. The left scenario had clearly missing landmarks removed, while the right scenario had both missing and unclear landmarks removed. The table displayed six different strategies for handling missing values, along with the FIT target metrics BCR and DER, and the average rank of both metrics. The strategies were compared to a control group where all landmarks were placed by an experienced operator, and missing landmarks were set based on the most likely position in the context of other features of the footprint.

Imputation method	Missing delete			Missing and unclear delete		
	BCR	DER	Avg. Rank	BCR	DER	Avg. Rank
Control	0.80	0.00	–	0.80	0.00	–
Delete Case	0.81	0.12	5.75	0.60	0.18	5.75
Mean Imp.	0.90	0.06	3.00	0.86	0.12	2.75
MICE PMM	0.84	0.12	5.25	0.88	0.06	1.50
MICE RF	0.87	0.00	2.75	0.75	0.00	2.50
missForest	0.90	0.00	1.25	0.81	0.12	3.25
missRanger	0.90	0.06	3.00	0.65	0.18	5.25

(2023). Additionally, a publication analyzing fisher (*Pekania pennanti*) and Pacific marten (*Martes caurina*) footprints demonstrated that defining feature extraction points on frequently visible landmarks, rather than on the entire foot, allowed the usage of more footprints Tucker et al. (2024).

Comparing the results of the five datasets, we observe that the white rhino set exhibits a more gradual slope in the linear relationship between missing animals and the percentage of missing values. Additionally, there is higher variance in the FIT target metrics for this dataset. We attribute these observations to the higher number of prints per animal in this specific dataset. To test this assumption, we randomly selected a subset of the white rhino dataset with sample sizes similar to the other datasets and verified our hypothesis. Despite these minor discrepancies, we interpret our findings as consistent among all five species for the artificially introduced missing values and for the cheetah dataset with

real missing values. We, therefore, conclude that our findings are indicative of a more generalized relationship.

In the first experiment, we assumed that there was no underlying mechanism responsible for the occurrence of missing values (MCAR). This is a legitimate assumption for incompleteness, which is due to external factors such as changes in weather conditions and heterogeneous substrates. Nevertheless, there might be additional systematic causes for missing values introduced by other factors occurring in real-life data.

Although we did not investigate the causation of introduced missing patterns in this study, we have several explanations for the potential general occurrence of systematically missing features in a footprint:

- **Missing values due to substrate type:** One explanation could be that several features of a footprint are differently affected by erosion. Deeper parts of a footprint, especially in regions of the track where the depth gradient between the negative space and the print is high, erode faster than parts with lower depth gradients. This effect can be magnified in dry and wet substrates.
- **Effects of gait, speed, and movement:** Faster traveling speeds, such as when galloping, usually cause more disturbance within footprints. This disturbance caused by movement is often most visible at points where most of the pressure was “released” into the substrate. Similar effects can be seen when an animal changes course from the initial direction of travel [Brown et al. \(1993\)](#). At slower speeds (e.g., walk or trot), animals often step with the hind foot in the track of the front foot (double register), which is common in many species [Elbroch and McFarland \(2019\)](#). This can cause systematic missing patterns in particular and therefore problems, especially when front feet are used for the analysis such as in the giant panda (*Ailuropoda melanoleuca*) [Li et al. \(2018\)](#) and Eurasian otters (*Lutra lutra*) [Kistner et al. \(2022\)](#).
- **Species-specific variation:** Most species have specific gait patterns, which can be influenced by changes in weight distribution when an individual is moving. This can lead to incomplete footprints. Some species have characteristic missing points in their footprints (e.g., otters and coypu (*Myocastor coypus*)) that can be a useful classification criterion for the identification of certain species [Brown et al. \(1993\)](#). If a species has characteristic missing patterns and if they are not addressed correctly, it is unlikely that FIT can be deployed as a successful method for such a species. However, if the missing patterns are consistent, then the feature extraction protocol can be designed accordingly for that particular species.
- **Individual missing pattern:** This is another likely source for missing values. Life history events like injuries can leave individual-specific variation that results in systematic missing values in a footprint. As an extreme example, if an individual lost a toe, this would always be visible in a footprint, and if strictly following the FIT protocol, an animal with such a history might be overlooked because of insufficient track quality. Even though we did not further investigate this in this study because of insufficient data, we hypothesize that a matrix containing information on the quality of the landmarks displayed in a footprint could be a valuable additional feature for individual identification.

## 5. Conclusion

In summary, we conclude that a quality assessment of each landmark with a focus on understanding systematically missing features in footprints could improve FIT. We, therefore, recommend conducting studies to investigate this, as well as applying appropriate imputation methods.

The ability to accurately count and identify wild individual animals using footprints is of considerable value to censusing and monitoring in wildlife conservation [Alibhai et al. \(2008\)](#). There are, however, some limitations when working with footprints [Jewell et al. \(2001\)](#); [Alibhai et al. \(2017, 2008\)](#); [Li et al. \(2018\)](#); [Gu et al. \(2014\)](#); [Jewell et al. \(2016,](#)

[2020\)](#). The most obvious one is the ability of the tracker to find and correctly identify them. Assuming equal tracking skills, we argue that substrate, the size of an animal, population densities, and the weather have the biggest impact on the likelihood of finding high-quality footprints. For instance, we assume it is easier to track a large mammal such as a white rhino in wet sand near a waterhole than it is to follow a mountain lion through dense, rocky forest habitat on a dry, windy summer day. Therefore, it can be assumed that our findings might have a different impact on different species and in different habitats. Most of the studies we referred to reported that they only collected high-quality footprints [Jewell et al. \(2001\)](#); [Alibhai et al. \(2017, 2008\)](#); [Li et al. \(2018\)](#); [Gu et al. \(2014\)](#); [Moreira et al. \(2018\)](#); [Jewell et al. \(2016\)](#). When the abundance of high-quality footprints is a limiting factor, our approach can transform the ability to use data of lower quality. This is of particular value for population estimate surveys. Therefore, our proposed approach could widen the range of FIT and we encourage future development of new species algorithms for which the number of high-quality prints in a trail found in situ is usually lower.

A further limitation when working with footprints is an operator's experience and subjectivity for placing the landmark points [Alibhai et al. \(2008\)](#). Although FIT has a rigorous, standardized feature extraction protocol that aims to minimize this operator bias [Alibhai et al. \(2008\)](#), we assume some landmarks might still appear clearer to some operators than others. Encouraged by the high performance of the best imputation methods, we argue that the ability to skip unclear landmarks might further increase the objectivity of FIT. We propose that this should also be further investigated when comparing the output of different operators.

Regarding the usage of appropriate imputation methods before applying the FIT algorithm, we recommend tree-based nonparametric imputation methods like missForest, missRanger or MICE RF. This is in line with recent findings on imputing before applying classification algorithms from machine learning [Buczak et al. \(2023\)](#). According to [Breiman \(2001\)](#), the advantages of Random Forests are, besides the handling of mixed-type data, an accurate prediction performance under high dimensions and non-linear data structures. These are often found in real datasets and should be considered when imputing missing values.

In conclusion, FIT is a valuable, non-invasive, cost-effective and highly applicable tool for wildlife conservation, which provides reliable data to inform wildlife management actions as required in Targets 4 and 21 of the UN's Kunming-Montreal GBF 2022 [Convention on Biological Diversity \(CBD\) \(2022\)](#). This study adds a novel improvement to this established method and highlights the usefulness of poor quality and incomplete footprints. It demonstrates that the problem of missing values in animal tracks can be overcome and that a larger variety of track qualities for footprint analysis can be used as a valuable data source.

We strongly believe that the utilization of professional tracking combined with community involvement through citizen science data collection of footprints and state-of-the-art machine learning holds immense potential as an additional non-invasive method in the toolbox of every conservationist for monitoring endangered and elusive species. This integrated approach allows for the collection of valuable data from both experts and the general public, fostering a collaborative effort in wildlife conservation.

## Declaration of competing generative AI in scientific writing

During the preparation of this work, the author(s) used OpenAI's ChatGPT for spell checking, grammar checking, and debugging in code to improve readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## CRediT authorship contribution statement

**Frederick Kistner:** Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Methodology, Visualization. **Justus Tulowitzki:** Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Larissa Slaney:** Data curation, Visualization, Writing – original draft, Writing – review & editing. **Sky Alibhai:** Software, Supervision, Validation, Writing – original draft, Writing – review & editing, Data curation. **Zoe Jewell:** Resources, Supervision, Writing – original draft, Writing – review & editing. **Burim Ramosaj:** Methodology, Supervision, Writing – review & editing. **Markus Pauly:** Funding acquisition, Methodology, Supervision, Writing – review & editing.

## Declaration of competing interest

None.

## Data availability

Data found on Github and referred in Manuskript. [https://github.com/TuloJustus/RCode\\_Datasets-Imputing\\_missing\\_values\\_in\\_footprint\\_metrics\\_for\\_wildlife\\_monitoring](https://github.com/TuloJustus/RCode_Datasets-Imputing_missing_values_in_footprint_metrics_for_wildlife_monitoring)

## Appendix A

**Table 6**

FIT default parameters for five different species. FIT models contain three tunable hyperparameters: 1. The number of variables used in the stepwise forward discriminant analyses selected based on the F-Ratios; 2. The counter probability that defines the prior likelihood of trails originating from the same or from different animals; 3. The Ward's cut-off distance, predicts the number of clusters (= number of predicted animals). All hyperparameters have been previously tuned for all species, and we have selected the default parameters within our experiments to assure comparability between the methods.

Species (latin name)	No. of variables	Counter probability	Wards cut-off distance
Amur Tiger ( <i>Panthera tigris altaica</i> )	18	0.50	1.95
Cheetah ( <i>Acinonyx jubatus</i> )	18	0.50	1.90
Giant Panda ( <i>Ailuropoda melanoleuca</i> )	12	0.50	1.50
Mountain Lion ( <i>Puma concolor</i> )	18	0.50	1.90
White Rhino ( <i>Ceratotherium simum</i> )	12	0.50	1.95
Cheetah (experiment 2)	18	0.50	1.90

## References

- Agha, M., Batter, T., Bolas, E., Collins, A., Gomes da Rocha, D., Montez-Moreno, C., Preckler-Quisquater, S., Sollmann, R., 2018. A review of wildlife camera trapping trends across africa. *Afr. J. Ecol.* 56, 694–701. URL: <https://onlinelibrary.wiley.com/doi/10.1111/aje.12565>.
- Alibhai, S., Jewell, Z., Law, P., 2008. A footprint technique to identify white rhino (*ceratotherium simum*) at individual and species levels. *Endanger. Species Res.* 4, 205–218. URL: <https://www.int-res.com/abstracts/esr/v4/n1-2/p205-218/>.
- Alibhai, S., Jewell, Z., Evans, J., 2017. The challenge of monitoring elusive large carnivores: an accurate and cost-effective tool to identify and sex pumas (*puma concolor*) from footprints. *PLoS One* 12, e0172065. URL: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0172065>.
- Alibhai, S., Gu, J., Jewell, Z., Morgan, J., Liu, D., Jiang, G., 2022. 'I know the tiger by his paw': a non-invasive footprint identification technique for monitoring individual Amur tigers (*panthera tigris altaica*) in snow. *Eco. Inform.* 101947 URL <https://doi.org/10.1016/j.ecoinf.2022.101947> <https://www.sciencedirect.com/science/article/pii/S1574954122000064>.
- Awais, M., Naeem, F., Rasool, N., Mahmood, S., 2018. Identification of sex from footprint dimensions using machine learning: a study on population of Punjab in Pakistan. *Egypt. J. Forensic Sci.* 8, 1–9. URL: <https://ejfs.springeropen.com/articles/10.1186/s41935-018-0081-1>.
- Bohmann, K., Evans, A., Gilbert, M.T.P., Carvalho, G.R., Creer, S., Knapp, M., Douglas, W.Y., De Bruyn, M., 2014. Environmental dna for wildlife biology and biodiversity monitoring. *Trends Ecol. Evol.* 29, 358–367. URL: <https://www.sciencedirect.com/science/article/pii/S0169534714000725>. <https://doi.org/10.1016/j.tree.2014.04.003>.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32. URL: <https://link.springer.com/article/https://doi.org/10.1023/A:1010933404324>.
- Brown, R., Lawrence, M., Pope, J., 1993. *Animals: Tracks, Trails & Signs*, Hamlyn.
- Brown, R., Lees, D., Ferguson, J., Lawrence, M., 2021. *Tracks and Signs of the Birds of Britain and Europe*. Bloomsbury Publishing. URL: <https://www.bloomsbury.com/uk/tracks-and-signs-of-the-birds-of-britain-and-europe-9781472961008/>.
- Buczak, P., Chen, J.J., Pauly, M., 2023. Analyzing the effect of imputation on classification performance under mcar and mar missing mechanisms. *Entropy* 25, 521. URL: <https://www.mdpi.com/1099-4300/25/3/521>. <https://doi.org/10.3390/e25030521>.
- Ceballos, G., Ehrlich, P., Barnosky, A.D., García, A., Pringle, R., Palmer, T.M., 2015. Accelerated modern human-induced species losses: entering the sixth mass extinction. *Science. Advances* 1. <https://doi.org/10.1126/sciadv.1400253>. URL: <https://www.science.org/doi/10.1126/sciadv.1400253>.

- Ceballos, G., Ehrlich, P., Raven, P., 2020. Vertebrates on the brink as indicators of biological annihilation and the sixth mass extinction. *Proc. Natl. Acad. Sci.* 117, 13596–13602. URL: <https://www.pnas.org/content/117/24/13596>. <https://doi.org/10.1073/pnas.1922686117>.
- Conde, D., Staerck, J., Colchero, F., da Silva, R., Schöley, J., Baden, M., Jouvett, L., Fa, J., Syed, H., Jongejans, E., Meiri, S., Gaillard, J.M., Chamberlain, S., Wilcken, J., Jones, O., Dahlgren, J., Steiner, U., Bland, L., Gomez-Mestre, I., Lebreton, J.D., González Vargas, J., Flesness, N., Canudas-Romo, V., Salguero-Gómez, R., Byers, O., Bjørnboe Berg, T., Scheuerlein, A., Devillard, S., Schigel, D., Ryder, O., Possingham, H., Baudisch, A., Vaupel, J., 2019. Data gaps and opportunities for comparative and conservation biology. *Proc. Natl. Acad. Sci.* 116, 9658–9664. URL: <https://www.pnas.org/content/116/19/9658>. <https://doi.org/10.1073/pnas.1816367116>.
- Convention on Biological Diversity (CBD), 2022. Kunming–montreal global biodiversity framework: Draft decision submitted by the president cbd/cop/15/l. 25, 18 december 2022. Convention on Biological Diversity. URL: <https://www.cbd.int/doc/c/9f1b/ce93/a0d4b06a303ae878c9535b1/cop-15-l-25-en.pdf>. draft decision submitted by the President of the Convention on Biological Diversity Conference of the Parties (COP) at its 15th meeting.
- Darras, K., Rountree, R., Van Wilgenburg, S., Dong, L., Gasc, A., Pitz, F., Chen, Y., Lin, T. H., Diaz, P., Wu, S.H., Salton, M., Marley, S., Cord, A., Acevedo-Charry, O., Adam, M., Do Nascimento, L., Astaras, C., Atamasov, A., Barbaro, L., Batist, C., Lopez Baucells, A., Bayle Sempere, J., Bellisario, K., Ben David, A., Berger-Tal, O., Betts, M., Bhalla, I., Bicudo, T., Bolgan, M., Bombaci, S., Boullhesen, M., Bradfer-Lawrence, T., Briars, R., Budka, M., Calvente, A., Cerezo-Araujo, M., Cerwen, G., Desjournes, C., Diniz, P., Duarte, A., Enari, H., Enari, H., Ferrari, N., Freitas, B., Friedman, N., Froidevaux, J., Gogoleva, S., Gonzalez Correa, J.M., Goodale, E., Greenhalgh, J., Hache, S., Hage, J., Halliday, W., Hanf-Dressler, T., Jacot, A., Jahn, O., Kepfer-Rojas, S., Lehnhardt, Y., Leon, J., Lloyd, K., Looby, A., Lopez-Bosch, D., Mammides, C., Marccaci, M., Markolf, M., Marques, M., Martin, C., Martin, D., Meyer, C., Muller, S., Navarro, M., Nouioua, R., Opper, S., Osiecka, A., Parsons, M., Pereira Samarra, F., Pradervand, J.N., Perez-Granados, C., Reynolds, L., Riede, K., Rocha, R., Ross, S., Scharffenberg, K., Schuchmann, K.L., Sebastian-Gonzalez, E., Singer, D., Traba, J., Tremblay, J., Tseng, S., Valverde, M., Vernasco, B., Ward, M., Weldy, M., Wensveen, P., Wood, H., da Silva, L., de Oliveira Tissiani, A., de Zwaan, D., Li, S., Sousa-Lima, R., Wanger, T., 2024. Worldwide landscape ecology patterns across realms. *bioRxiv*. <https://doi.org/10.1101/2024.04.10.588860>, 2024–04URL:<https://www.biorxiv.org/content/10.1101/2024.04.10.588860>.
- De Angelo, C., Paviolo, A., Di Bitetti, M., 2010. Traditional versus multivariate methods for identifying jaguar, puma, and large canid tracks. *J. Wildl. Manag.* 74, 1141–1151. URL: <https://www.jstor.org/stable/40665189>.
- Doove, L., Van Buuren, S., Dusseldorp, E., 2014. Recursive partitioning for missing data imputation in the presence of interaction effects. *Comp. Stat. Data Anal.* 72, 92–104. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0167947313003134>. <https://doi.org/10.1016/j.csda.2013.10.025>.
- Dorfman, A., Subach, A., Scharf, I., 2023. Tracking spoor and field signs as a zoological research method. *Anim. Behav.* 203, 133–141. URL: <https://www.sciencedirect.com/science/article/pii/S0003347223001793>.
- Elbroch, M., McFarland, C., 2019. *Mammal Tracks & Sign: A Guide to North American Species*. Rowman & Littlefield.
- Evans, J., 2023. North american animal tracks database. <https://www.inaturalist.org/projects/north-american-animal-tracks-database> Accessed: 2024-05-28.
- Garcia-Laencina, P., Sancho-Gomez, J., Figueiras-Vidal, A., 2010. Pattern classification with missing data: a review. *Neural Comput. & Applic.* 19, 263–282. URL: <https://link.springer.com/article/https://doi.org/10.1007/s00521-009-0365-0>.
- Grolms, J., 2021. *Tierspuren Europas: Spuren und Zeichen bestimmen und interpretieren. Mit Spuren und Zeichen von Säugetieren, Vögeln, Reptilien, Amphibien und Wirbellosen*. Ulmer.
- Gu, J., Alibhai, S., Jewell, Z., Jiang, G., Ma, J., 2014. Sex determination of Amur tigers (*panthera tigris altaica*) from footprints in snow. *Wildl. Soc. Bull.* 38, 495–502. URL: <https://wildlife.onlinelibrary.wiley.com/doi/10.1002/wsb.438>.
- Hasler, N., Klette, R., Agnew, W., 2004. *Footprint Recognition of Rodents and Insects*. Technical Report.. CINTR, The University of Auckland, New Zealand.
- Jewell, Z., 2013. Effect of monitoring technique on quality of conservation science. *Conserv. Biol.* 27, 501–508. URL: <https://onlinelibrary.wiley.com/doi/10.1111/cobi.12066>.
- Jewell, Z., Alibhai, S., Law, P., 2001. Censusing and monitoring black rhino (*Diceros bicornis*) using an objective spoor (footprint) identification technique. *J. Zool.* 254, 1–16. URL: <https://zslpublications.onlinelibrary.wiley.com/doi/abs/10.1017/S0952836901000516>.
- Jewell, Z., Alibhai, S., Weise, F., Munro, S., Van Vuuren, M., Van Vuuren, R., 2016. Spotting cheetahs: identifying individuals by their footprints. *J. Visual. Exp.* <https://doi.org/10.3791/54033>. URL: <https://www.jove.com/v/54033/spotting-cheetahs-identifying-individuals-by-their-footprints>.
- Jewell, Z., Alibhai, S., Law, P., Useib, K., Lee, S., 2020. Monitoring rhinoceroses in namibia's private custodianship properties. *PeerJ* 8, e9670. URL: <https://peerj.com/articles/9670/>.
- Kistner, F., Slaney, L., Jewell, Z., Ben-David, A., Alibhai, S., 2022. It's otterly confusing - distinguishing between footprints of three of the four sympatric asian otter species using morphometrics and machine learning. *Otter* 8, 108–125. URL: <https://www.wildtrack.org/assets/documents/IOSF-JOURNAL-2022-Kistner-et-al-108-125-1.pdf>.
- Kistner, F., Slaney, L., Morant, N., 2023. Can you tell the species by a footprint? Identifying three of the four sympatric southeast asian otter species using computer vision and deep learning. *IUCN Otter Spec. Group Bull.* 40, 197–210. URL: [https://iucnsgbull.org/Volume40/Kistner\\_et\\_al\\_2023.pdf](https://iucnsgbull.org/Volume40/Kistner_et_al_2023.pdf).
- Laity, K., 2015. *The Field Validation of Footprint Identification Technology on Free-roaming Cheetahs in Namibia*. Master's project. Duke University. URL: <https://hdl.handle.net/10161/9651>. accessed: 2024-05-14.
- Lefort, M.C., Cruickshank, R.H., Descovich, K., Adams, N.J., Barun, A., Emami-Khoyi, A., Riddien, J., Smith, V.R., Sprague, R., Waterhouse, B., et al., 2022. Blood, sweat and tears: a review of non-invasive DNA sampling. *Peer Commun. J.* 2 <https://doi.org/10.24072/pcjournal.86>. URL: <https://peercommunityjournal.org/articles/10.24072/pcjournal.86>.
- Li, B., Alibhai, S., Jewell, Z., Li, D., Zhang, H., 2018. Using footprints to identify and sex giant pandas. *Biol. Conserv.* 218, 83–90. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0006320717315030>.
- Liebenberg, L., 2013. *The Origin of Science*. CyberTracker. URL: <https://www.cybertacker.org/science/the-origin-of-science>.
- Liebenberg, L., 2014. *Photographic Guide to Tracks & Tracking in Southern Africa*. Penguin Random House South Africa.
- Marchal, A., Lejeune, P., De Bruyn, P.N., 2017. Identification of the anteroposterior and mediolateral position of lion paws and tracks using 3d geometric morphometrics. *African Journal of Wildl. Res.* 47, 106–113. URL: <https://www.ajol.info/index.php/ajwr/article/view/170973>.
- Mayer, M., 2021. *missRanger: Fast Imputation of Missing Values*. URL: <https://CRAN.R-project.org/package=missRanger>. r package version 2.1.3.
- Moreira, D., Alibhai, S., Jewell, Z., da Cunha, C., Seibert, J., Gatti, A., 2018. Determining the numbers of a landscape architect species (tapirus terrestris), using footprints. *PeerJ* 6. <https://doi.org/10.7717/peerj.4591>. URL: <https://peerj.com/articles/4591/>.
- Palma, A., Gurgel-Gonçalves, R., 2007. Morphometric identification of small mammal footprints from ink tracking tunnels in the brazilian cerrado. *Rev. Brasil. Zool.* 24, 333–343. URL: <https://www.scielo.br/rbzoil/a/kxTV4cQ68gPv96X8FK9f6vN/a/abstract/?lang=en>. <https://doi.org/10.1590/S0101-81752007000200012>.
- Penar, W., Magiera, A., Klocek, C., 2020. Applications of bioacoustics in animal ecology. *Ecol. Complex.* 43 <https://doi.org/10.1016/j.ecocom.2020.100870>. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1476945X20300865>.
- Pimm, S., Alibhai, S., Bergl, R., Dehgan, A., Giri, C., Jewell, Z., Joppa, L., Kays, R., Loarie, S., 2015. Emerging technologies to conserve biodiversity. *Trends Ecol. Evol.* 30, 685–696. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0169534715000224X>. <https://doi.org/10.1016/j.tree.2015.08.008>.
- Ponce-Martins, M., Lopes, C.K.M., Carvalho-Jr, E.A.R.D., Castro, F.M.D.R., Paula, M.J.D., Brito, J.C., Pezzuti, J.C.B., 2022. Assessing the contribution of local experts in monitoring neotropical vertebrates with camera traps, linear transects and track and sign surveys in the amazon. *Perspect. Ecol. Conserv.* 20, 303–313. URL: <https://www.perspectecolconserv.com/en-assessing-contribution-local-experts-in-articulo-S253006442200058X>. <https://doi.org/10.1016/j.pecon.2022.08.002>.
- R Core Team, 2020. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <https://www.R-project.org/>.
- Ramosaj, B., Pauly, M., 2019. Predicting missing values: a comparative study on non-parametric approaches for imputation. *Comput. Stat.* 34, 1741–1764. URL: <https://link.springer.com/article/https://doi.org/10.1007/s00180-019-00877-y>.
- Ramosaj, B., Amro, L., Pauly, M., 2020. A cautionary tale on using imputation methods for inference in matched-pairs design. *Bioinformatics* 36, 3099–3106. URL: <https://academic.oup.com/bioinformatics/article/36/10/3099/5735363>. <https://doi.org/10.1093/bioinformatics/btaa074>.
- Ramosaj, B., Tulowitzki, J., Pauly, M., 2022. On the relation between prediction and imputation accuracy under missing covariates. *Entropy* 24, 386. URL: <https://www.mdpi.com/1099-4300/24/3/386>. <https://doi.org/10.3390/e24030386>.
- Rhyder, J., 2021. *Tracks and Signs: A Guide to the Field Signs of Mammals and Birds of the U.K.* The History Press Ltd. URL: <https://www.alibris.com/booksearch?keyword=9780750996143>.
- Rockel, T., 2020. *missMethods: Methods for Missing Data*. URL: <https://CRAN.R-project.org/package=missMethods>. r package version 0.3.0.
- Rubin, D.B., 2004. *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons, Hoboken, NJ. URL: <https://www.wiley.com/en-us/Multiple+Imputation+for+Nonresponse+in+Surveys-p-9780471655749>.
- SAS Institute Inc, 2021. *JMP®: Version 16.2*. SAS Institute Inc, Cary, NC, pp. 1989–2021. URL: <https://www.jmp.com/support/help/en/16.2/>.
- Schwerter, J., Gurtškaia, K., Romero, A., Zeyer-Gliozzo, B., Pauly, M., 2024. Evaluating tree-based imputation methods as an alternative to mice pmm for drawing inference in empirical studies. *arXiv preprint. arXiv:2401.09602*. URL: <https://arxiv.org/abs/2401.09602>.
- Seidlitz, A., Bryant, K.A., Armstrong, N.J., Calver, M.C., Wayne, A.F., 2021. Sign surveys can be more efficient and cost effective than driven transects and camera trapping: a comparison of detection methods for a small elusive mammal, the numbat (*myrmecobius fasciatus*). *Wildl. Res.* 48, 491–500. URL: <https://bioone.org/journals/wildlife-research/volume-48/issue-6/WR20020/sign-surveys-can-be-more-efficient-and-cost-effective-than-driven-transects-and-camera-trapping-a-comparison-of-detection-methods-for-a-small-elusive-mammal-the-numbat-myrmecobius-fasciatus>.
- Sharma, S., Jhala, Y., Sawarkar, V., 2003. Gender discrimination of tigers by using their pugmarks. *J. Wildl. Res.* 8, 37–42. URL: <https://www.jstor.org/stable/3784382>.
- Sharma, S., Jhala, Y., Sawarkar, V., 2005. Identification of individual tigers (*panthera tigris*) from their pugmarks. *J. Zool.* 267, 9–18. URL: <https://www.cambridge.org/core/journals/journal-of-zoology/article/identification-of-individual-tigers-panthera-tigris-from-their-pugmarks/10.1017/S0952836905006657>.
- Stekhoven, D., 2013. *missForest: Nonparametric Missing Value Imputation using Random Forest*. URL: <https://CRAN.R-project.org/package=missForest>. r package version 1.5.

- Stekhoven, D., Bühlmann, P., 2012. Missforest — non-parametric missing value imputation for mixed-type data. *Bioinformatics* 28, 112–118. URL: <https://academic.oup.com/bioinformatics/article/28/1/112/290015>. <https://doi.org/10.1093/bioinformatics/btr597>.
- Tharwat, A., 2020. Classification assessment methods. In: *Applied Computing and Informatics*. <https://doi.org/10.1016/j.aci.2020.05.001>. URL: <https://www.sciencedirect.com/science/article/pii/S2210832720301327>.
- Thomsen, P., Willerslev, E., 2015. Environmental dna—an emerging tool in conservation for monitoring past and present biodiversity. *Biol. Conserv.* 183, 4–18. URL: <https://www.sciencedirect.com/science/article/pii/S0006320714003734>.
- Thurrow, M., Dumpert, F., Ramosaj, B., Pauly, M., 2021. Imputing missings in official statistics for general tasks—our vote for distributional accuracy. *Stat. J. IAOS* 37, 1379–1390. URL: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji210901>. <https://doi.org/10.3233/SJI-210901>.
- Thurrow, M., Dumpert, F., Ramosaj, B., Pauly, M., 2024. Assessing the multivariate distributional accuracy of common imputation methods. *Stat. J. IAOS* 40, 99–108. URL: <https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji230007>. <https://doi.org/10.3233/SJI-230007>.
- Tollefson, J., 2019. Humans are driving one million species to extinction. *Nature* 569, 171–172. URL: <https://www.nature.com/articles/d41586-019-01448-4>. <https://doi.org/10.1038/d41586-019-01448-4>.
- Tucker, J., King, C., Lekivetz, R., Murdoch, R., Jewell, Z., Alibhai, S., 2024. Development of a non-invasive method for species and sex identification of rare forest carnivores using footprint identification technology. *Eco. Inform.* 79, 102431. URL: <https://www.sciencedirect.com/science/article/pii/S1574954123004600>.
- United Nations Department of Economic and Social Affairs, 2023. The Sustainable Development Goals Report 2023. Technical Report. United Nations. URL: <https://unstats.un.org/sdgs/report/2023/The-Sustainable-Development-Goals-Report-2023.pdf>. prepared in collaboration with over 50 international and regional agencies, based on data from more than 200 countries and territories.
- van Buuren, S., Groothuis-Oudshoorn, K., 2011a. mice: multivariate imputation by chained equations in r. *J. Stat. Softw.* 45, 1–67. URL: <https://www.jstatsoft.org/v45/i03/>.
- Van Buuren, S., Groothuis-Oudshoorn, K., 2011b. mice: Multivariate imputation by chained equations in r. *J. Stat. Softw.* 45, 1–67. URL: <https://www.jstatsoft.org/v045/i03>.
- Van Buuren, S., Oudshoorn, K., 1999. Flexible Multivariate Imputation by MICE. Technical Report, TNO Prevention and Health, Leiden.
- WildTrack, 2023. Wildtrackai. Available on the Apple App Store. URL: <https://apps.apple.com/ch/app/wildtrackai/id1625129285>. accessed: 2024-05-28.
- Zalewski, A., 1999. Identifying sex and individuals of pine marten using snow track measurements. *Wildl. Soc. Bull.* 27, 28–31. URL: <https://www.jstor.org/stable/3784121>.
- Zemanova, M., 2020. Towards more compassionate wildlife research through the 3rs principles: moving from invasive to non-invasive methods. *Wildl. Biol.* 2020, 1–17. URL: <https://bioone.org/journals/wildlife-biology/volume-2020/issue-1/wlb.00607/towards-more-compassionate-wildlife-research-through-the-3rs-principles-moving/https://doi.org/10.2981/wlb.00607.full>.