

Earthworm Image Recognition Project Activity Report

Keiron Derek Brown Version 5 25 November 2024

Produced for







Contents

Ac	kno	owledgements	3
1	h	ntroduction	4
2	Р	Project Summary	4
3	Р	Project Timeline	5
4	E	Earthworm Specimen Photography Protocol	6
4	4.1	Photography Protocol Issues	7
5	Е	Earthworm Sampling Day Programme	8
ļ	5.1	Earthworm Sampling Day Format and Guidance	9
į	5.2	Secondary Achievements of Earthworm Sampling Days	10
6	Т	Fraining Image Library Summary	11
7	A	Algorithm Development	13
	7.1	Algorithm performance by image	14
•	7.2	2 Algorithm performance by species	15
8	C	Conclusions	17
9	В	Bibliography	17

This project received funding from the Department for Environment, Food and Rural Affairs (Defra) as part of the <u>Natural Capital and Ecosystem Assessment (NCEA) programme</u>. The NCEA is undertaking a nationwide survey of England's land, coast and sea with the aim to transform environmental decision making by building a 'whole system' picture of the state of our natural environment.





This document is distributed under a Creative Commons Attribution 4.0 International Licence.

Suggested citation: Brown, K.D. (2024) Earthworm Image Recognition Project Activity Report [Version 5]. London: Biological Recording Company



Acknowledgements

The author would like to thank the following groups and people for their contribution to this project:

- Defra and the Joint Nature Conservation Committee for sitting on a steering group for the project.
- Tom August, Alba Gomez Segura and Richard Pywell for helping to put together the photography protocol and their guidance on the technical side of the project.
- Rachel Davies, Leanna Dixon, Katie Last, Kate Lemon, Charlie Linton, Nick Marriner, Elliot Newton, Anna Rebman and Elspeth Robinson for helping to organise the sampling permissions and access to sites.
- The many volunteers that took part in the Earthworm Sampling Day programme and helped to collect and photograph the specimens that were used within this project.
- Douglas Nunes for his efforts sorting the photos received from these events.
- The Tees-Swale Naturally Connected Project staff and volunteers for assisting with the ID of specimens from the sampling in North Yorkshire.
- Sheila Warbus for her attendance of all of the London sampling and assistance in identifying the specimens that resulted from this.
- Aidan Keith for providing additional data by sampling and photographing earthworms according to the project protocol outside of project sampling events.
- Emma Sherlock for assisting with the identification and verification of the many specimens that required a second opinion.



Photo 1: Volunteer surveyors at Grove Farm Nature Reserve (London). Image: Anna Rebmann.





Photo 3: Volunteer earthworm surveyors at Rothwell Country Park (West Yorkshire). Image: Keiron Brown.



Photo 4: Volunteer earthworm surveyors at Sefton Park (Liverpool). Image: Peter Brash.

1 Introduction

Earthworms are widely regarded to be of great ecological importance, with different ecological categories of earthworms contributing to soil processes and resulting in a number of ecosystem services (see **Figure 1**) (Keith & Robinson, 2012).

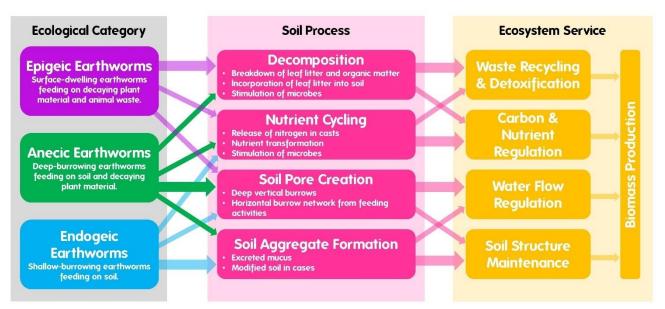


Figure 1: Earthworm ecosystem services adapted from (Keith & Robinson, 2012).

Despite recognition of their role as ecosystem engineers, earthworms have previously been relatively under-recorded in comparison to other species groups (Carpenter, et al., 2012). 33 species are known to occur in the UK and Ireland. Earthworms are difficult to identify, requiring the collection of preserved specimens and observation of microscopic features in order to reach an accurate species identification.

2 Project Summary

This project received funding from the Department for Environment, Food and Rural Affairs (Defra) as part of the Natural Capital and Ecosystem Assessment (NCEA) programme. The NCEA is undertaking a nationwide survey of England's land, coast and sea with the aim to transform environmental decision making by building a 'whole system' picture of the state of our natural environment.

In this proof-of-concept study, the Biological Recording Company and UK Centre for Ecology and Hydrology investigated the possibility of using an image recognition algorithm and machine learning to support identification of live earthworm specimens in the field.

The project aimed to build an image library of live earthworm specimens that had been subsequently identified to species by an earthworm specialist using traditional methods (i.e. observation of morphological features through microscopy).

The image library was then used to train, test and validate an earthworm identification algorithm and the results of this proof-of-concept study were reported back to the funder.

3 Project Timeline

The project commenced in July 2023 and the final action for this proof-of-concept study was completed on 14^{th} October 2024.

A summary of project activities and timeline is provided in Table 1

Table 1: Summary table of project activities and timeline.

Trial undertaken photographing earthworms in the field Draft protocol for earthworm sampling and photography Meet with CEH on 26/07/23 to finalise protocol Aug 2023 Design Earthworm Sampling Day survey plan and identify sample sites Schedule 8 Earthworm Sampling Days across Great Britain Promote Earthworm Sampling Days and manage volunteer bookings Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 species occurrence records to iRecord Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH Schedule 3 Earthworm Sampling Days across London	
Meet with CEH on 26/07/23 to finalise protocol Aug 2023 Design Earthworm Sampling Day survey plan and identify sample sites Schedule 8 Earthworm Sampling Days across Great Britain Promote Earthworm Sampling Days and manage volunteer bookings Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Aug 2023 Design Earthworm Sampling Day survey plan and identify sample sites Schedule 8 Earthworm Sampling Days across Great Britain Promote Earthworm Sampling Days and manage volunteer bookings Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Schedule 8 Earthworm Sampling Days across Great Britain Promote Earthworm Sampling Days and manage volunteer bookings Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	\rightarrow
Schedule 8 Earthworm Sampling Days across Great Britain Promote Earthworm Sampling Days and manage volunteer bookings Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Prepare equipment and consumables for Earthworm Sampling Days Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	· · · · · · · · · · · · · · · · · · ·
Delivery of Earthworm Sampling Day Programme 1 (Great Britain) events ID of Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	· · · · · · · · · · · · · · · · · · ·
Oct 2023 Dof Earthworm Sampling Day Programme 1 earthworm specimens Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord	· · · · · · · · · · · · · · · · · · ·
Process Earthworm Sampling Day Programme 1 images and submit to CEH Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	· · · · · · · · · · · · · · · · · · ·
Submit Earthworm Sampling Day Programme 1 species occurrence records to iRecord Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	· · · · · · · · · · · · · · · · · · ·
Nov 2023 Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓ ✓
Nov 2023 Schedule 3 Earthworm Sampling Days on farms within Chilterns National Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Nov 2023 Landscape Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Dec 2024 Dec 20	
Delivery of Earthworm Sampling Day Programme 2 (Chilterns National Landscape) events ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Dec 2024 ID of Earthworm Sampling Day Programme 2 earthworm specimens Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Process Earthworm Sampling Day Programme 2 images Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Jan 2024 Submit Earthworm Sampling Day Programme 2 species occurrence records to iRecord Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Jan 2024 Process live specimen images collected to date Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	
Confirm uncertain species determinations with Natural History Museum Submit updated image library to CEH	✓
Feb 2024	✓
Schedule 3 Farthworm Sampling Days across London	✓
Schedule 3 Earthworth Sampling Day's across Eoridon	✓
Mar 2024 Delivery of Earthworm Sampling Day Programme 3 (London) events	~
Detivery of Earthworth Sampting Day Programme 3 (London) events	
ID of Earthworm Sampling Day Programme 3 earthworm specimens	_
Apr 2024 Process Earthworm Sampling Day Programme 3 images and submit to CEH	\ \ \ \
Submit Earthworm Sampling Day Programme 3 species occurrence records to	\ \ \
iRecord	*
May 2024 Algorithm development	>
Algorithm accuracy review	
Jun 2024 Review of project outputs	\ \ \ \
Tuli 2024 Neview of project outputs	· ·
Jul 2024 Production on proof-of-concept study report for client	
Delivery of 1-hour webinar to present project challenges and lessons learned to	~
Oct 2024 project volunteers	✓ ✓

4 Earthworm Specimen Photography Protocol

Earthworms were photographed in a standard 'white tub' container with the following dimensions: 18 cm length x 11.5 cm width x 7 cm height. The specific units used within this project were purchased from Parkers Packaging Limited: https://www.parkerspackaging.com/product/1-litre-rectangular-ice-cream-container-lid/

During sampling it was noted that mobile phone cameras often had difficulty focusing on the earthworm so the protocol was modified in February 2024 to use black containers so that the project could investigate the effectiveness of a white background versus a black background. Ideally, this would have involved the use of a black plastic container with the same shape and dimensions. However, it was not possible to source this product in black so an alternative product was used with the following dimensions: 19.5 cm length x 13.5 cm width x 3.5 cm height. The specific units used for this project were purchased from Deco Dine Limited: https://decodine.co.uk/categories/814-711-categories-16-oz-rectangular-container-with-lid.html

Using a standard container for photographing the specimens ensures a consistent background and enables a reference for the size of the specimen.



Figure 2: Details of the standard sampling containers used for the Earthworm Image Recognition Project.

When photographing earthworms for the Earthworm Image Recognition Project:

- The **whole container should appear within the image**. This helps the algorithm estimate the size of the earthworm in comparison to the standard size of the container.
- Labels and other objects should not be photographed within the container. Additional objects will be detected by the image recognition technology and create additional admin to remove.
- The **container should be kept clean of debris and soil** as this can also be detected by the image recognition technology. Where necessary, the container should be wiped clean between specimens.
- Photos should be taken from directly above the container and not at an angle with care taken not to create shadows in the image.
- Multiple photos of each specimen should be taken to account for the constantly changing shape of the specimen as it stretches, contracts and twists. At least one of the photos should have the specimen centrally in the container.



Figure 3: Acceptable and unacceptable examples of photographs for the earthworm specimen photography protocol.

Once photographed, earthworms were collected into labelled tubes of preservative and identified by Keiron Derek Brown using a microscope and the *Key to the Earthworms of the UK and Ireland (2nd ed.)* (Sherlock, 2018). Some specimens were identified by other trusted earthworm recorders and the determinations were all checked by Keiron Derek Brown. Where there was any uncertainty over the species identification, the specimens were submitted to the Natural History Museum (London) to be checked by Emma Sherlock.

Photographs of the specimen were then added to the relevant species folder on the Earthworm Images Google Drive.

4.1 Photography Protocol Issues

A number of factors were noted during sampling that could potentially be problematic for generating suitable images:

- **Precipitation**: Rainy conditions can result in water gathering in the containers and cause issues with glare and reflections.
- **Shadows**: Both the sides of the container and photographer were found to cast shadows over part or all of the image.
- **Soil**: Earthworms are found within the soil, so it was unsurprising that the sampling containers became dirty relatively quickly. This was particularly evident in muddy conditions.
- Excretions: Earthworms that have been collected may excrete liquid as a defence mechanism. This can impact the background of the image where containers have been marked and it was noted that this could also change the colour and patterns on the earthworm. For example, Aporrectodea icterica is known for having a yellow mottled appearance and it was noted that this appeared to fade out and disappear as specimens excreted mucus from their skin.
- **Photographer inconsistency**: Despite step-by-step instructions and a demonstration, many participants still failed to photograph specimens according to the protocol and a significant proportion of images were discarded (though the actual proportion discarded was not recorded).

5 Earthworm Sampling Day Programme

Earthworm Sampling Days are 1-day events where members of the public are invited to participate in earthworm surveys of a site. These events generate earthworm species occurrence records and engage local communities with earthworm research and soil health.

A total of 16 Earthworm Sampling Days were scheduled at sites across England and Wales, covering a variety of habitats. The Earthworm Sampling Day Programme aimed to cover a relatively wide geographic area and sampling sites were selected across a north to south gradient in England, from North Yorkshire to London. One site was also selected in Wales. No sites were selected in Scotland or Northern Ireland due to budget and logistical constraints. Earthworm Sampling Days undertaken in 2023 used the white container in the photography protocol, whereas Earthworm Sampling Days undertaken in 2024 used the black container.

A summary of the Earthworm Sampling Day Programme is provided in **Table 2**.



Earthworm
Sampling Days
delivered



Earthworm Sampling Day attendees



Earthworm specimens photographed and collected



8,862

Earthworm

photographs with

white background

generated



2,341
Earthworm
photographs with
white background
generated

Table 2: Summary of Earthworm Sampling Days.

Date	Venue	County	Attendees	Specimens	Images
26/09/2023	Paradise Fields	London	7	51	737
10/10/2023	Nant Mill	Denbighshire	5	51	1203
16/10/2023	Woodhouse Farm	North Yorkshire	10	43	1619
21/10/2023	Rothwell Country Park	West Yorkshire	4	34	462
23/10/2023	Erewash Meadows Nature Reserve	Derbyshire	5	43	576
24/10/2023	Erewash Meadows Nature Reserve	Derbyshire	2	17	202
25/10/2023	Woodside Nature Reserve	Derbyshire	10	35	1508
10/11/2023	Chilterns AONB (various farms)	Buckinghamshire	6	36	431
13/11/2023	Chilterns AONB (various farms)	Buckinghamshire	7	39	431
01/12/2023	Chilterns AONB (various farms)	Buckinghamshire	5	17	184
12/12/2023	Kensington Gardens	London	23	47	1514
27/03/2024	Southwood Open Space	London	11	47	337
06/04/2024	Grove Farm Nature Reserve	London	17	43	1309
22/04/2024	Sefton Park	Liverpool	10	47	415
29/04/2024	Paradise Fields	London	3	18	174
22/05/2024	Low Oxque Farm	North Yorkshire	8	49	448

5.1 Earthworm Sampling Day Format and Guidance

Each event started with an introduction to the project by an earthworm specialist and an outline of the event. The earthworm specialist selected the locations and sampling methods for each sampling point, with the aim of maximising the species diversity collected from each sample site.

Volunteers were trained to locate and handle earthworms (Brown, Earthworm Recorders Handbook [Version 8], 2019). Juvenile specimens were returned to the habitats they were found within and adult specimens were photographed by both the earthworm specialist and the volunteers (using their own mobile devices) as per the **Earthworm Specimen Photography Protocol**.

For each sample point, adult earthworms were separated into individual white tubs with a label. Volunteers were instructed to photograph the label and then take multiple images of the earthworm within the tub without the label. The white tubs were lined up with each volunteer moving down the line to enable volunteers to easily keep track of which specimens they had each photographed.



Photo 5: Volunteer earthworm surveyors photographing specimens according to the earthworm specimen photography protocol at Woodhouse Farm (North Yorkshire. Image credit: Keiron Derek Brown.

Photographed specimens were collected for identification by an earthworm specialist at a later date. Each specimen was placed into a tube of preservative alongside the specimen label. Additional details of the biological record (such as date, recorder, grid reference, site name, habitat and sampling method) were recorded on an earthworm recording form.

Following each event, folders were created for each attendee on the Earthworm Images Google Drive and they were invited to submit their images (including the photographs of the labels) to their individual folders. Following the identification of the specimens, the images were processed by the earthworm specialist (using the label photos as separators for each specimen) and sorted into the relevant earthworm species folders.

In addition to specimens collected during the Earthworm Sampling Days, a further 67 specimens were collected for the project and an additional 624 images were generated as a result. This included 36 specimens collected by Aidan Keith and 31 specimens collected by Keiron Derek Brown.

5.2 Secondary Achievements of Earthworm Sampling Days

In addition to the primary of aim of generating a training image library, the Earthworm Sampling Day programme also benefited the National Earthworm Recording Scheme by generating 295 new species occurrence records, including 11 instances where species were recorded for the first time within a Watsonian vice county.





24

50

57

63

65

Buckinghamshire

Denbighshire

Derbyshire

South-west Yorkshire North-west Yorkshire

Aporrectodea nocturna Dendrobaena veneta Lumbricus festivus Octolasion cyaneum Octolasion lacteum

Lumbricus festivus

Aporrectodea caliginosa Aporrectodea rosea Octolasion cyaneum Bimastos eiseni Eisenia fetida/andrei

Furthermore, the author was also able to take the opportunity to photograph some earthworm specimens in natural environments and build up an image library of earthworm species that can be used in future identification training courses, earthworm identification resources and engagement activities (see **Figure 4**).



Figure 4: Images of live earthworms in natural environments taken by the author during Earthworm Sampling Days.

A number of earthworm specimens that were collected for the project were also submitted to both the regional earthworm collection at The World Museum (Liverpool) and the national collection at the Natural History Museum (London).

6 Training Image Library Summary

This section summarises the total number of images sourced for each species of earthworm that is known to occur in natural environments in the UK and Ireland.



650 Earthworm specimens

determined to

species level



Species of earthworm identified



Species of earthworm yet to be found and photographed



11,827 Images of

Images of earthworms added to the training library



Specimens in progress (comprising a total of 352 images)

Table 3: Summary of images uploaded to the training image library by species.

Species	Number of specimens	Number of images
Allolobophora chlorotica	68	1409
Aporrectodea caliginosa	147	2428
Aporrectodea cupulifera		
Aporrectodea icterica	18	322
Aporrectodea limicola	5	31
Aporrectodea longa	68	1442
Aporrectodea nocturna	7	108
Aporrectodea rosea	51	640
Bimastos eiseni	4	63
Bimastos rubidus	26	576
Dendrobaena attemsi		
Dendrobaena hortensis		
Dendrobaena octaedra	7	227
Dendrobaena pygmaea		
Dendrobaena veneta	10	170
Eisenia andrei/fetida agg	20	333
Eiseniella tetraedra	12	143
Helodrilus oculatus		
Kenleenus armadas		
Lumbricus castaneus	67	1136
Lumbricus festivus	10	227
Lumbricus friendi		
Lumbricus rubellus	77	1723
Lumbricus terrestris	15	240
Microscolex phosphoreus		
Murchieona muldali	152	101
Octolasion cyaneum	12	158
Octolasion lacteum	12	196
Satchellius mammalis	8	154
Sparganophilus tamesis		



Figure 5: Examples of earthworm images submitted to the image library.

7 Algorithm Development

Images were augmented to increase the size of the dataset, which involved duplicating and editing images (for example, by flipping axis and altering zoom).

The images within the training image library were divided into three groups:

- 1. **Training images** (7,141): These images are used to train the algorithm in earthworm species identification during each epoch (training step).
- 2. **Validation images** (2,437): These images are used after each epoch to validate the performance of the algorithm.
- 3. **Test images** (2,594): These images are used to test the performance of the algorithm at the end of the project.

The algorithm was **trained for 34 epochs (training steps)**, as this was the point at which the algorithm was no longer improving in terms of performance, and achieved an **overall accuracy of 42%** (see **Figure 6**).

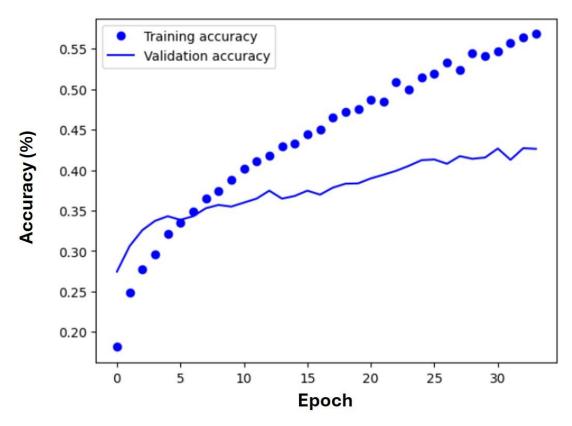


Figure 6: Training and validation accuracy with each successive epoch (training step).

Due to the training image library containing multiple images of each individual earthworm specimen, it is important these sets images of a single specimen to only be used for a single use (either training, testing or validation) and not for multiple uses as this would lead to a significant overestimation of the classifier's ability.

7.1 Algorithm performance by image

We can assess the performance of the algorithm by examining the features within images that the algorithm used to classify the image to a species.

Figure 7 demonstrates via heat maps of images how different areas within an image were used when classifying the image to a species identification. 'Hot' colours illustrate the areas of an image that were considered by the algorithm to be most important for classification.

We would expect the algorithm to be using the specific area where the earthworm is located within an image to make classifications. Although this was sometimes the case, we see numerous cases where the algorithm is using information from elsewhere in the image to classify the image.

This is likely to be significantly impacting the accuracy rate of the algorithm.

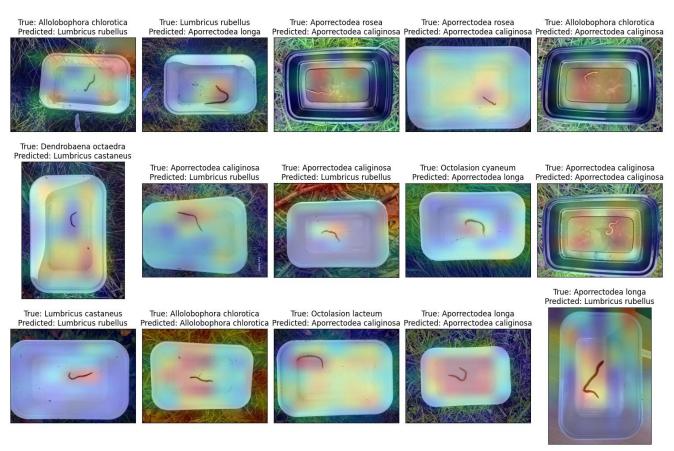


Figure 7: Heat maps of images showing how different areas within an image were used when classifying the image to a species identification. 'Hot' colours illustrate the areas of an image that were considered by the algorithm to be most important for classification.

Two potential solutions to address this issue are:

- 1. **Increasing the size of the training dataset** so that features such as the shape of the tub and environment around the tub become less meaningful.
- 2. **Masking out the background of the image** to remove parts of the image that we do not want the algorithm to use for classification (though it should be noted that this reduces the ability of the algorithm to use scale for classification).

7.2 Algorithm performance by species

Although the overall algorithm accuracy was 42%, the accuracy of the algorithm to accurately identify individual species was significantly variable and ranged from 0% to 69% depending on the species.

When the test results are plotted in a confusion matrix (see **Figure 8**) we can see where species were misidentified and which species the algorithm falsely classified them as. **An algorithm with a high accuracy rate would display a strong diagonal band,** and any significant values away from this diagonal band indicate common misclassifications.

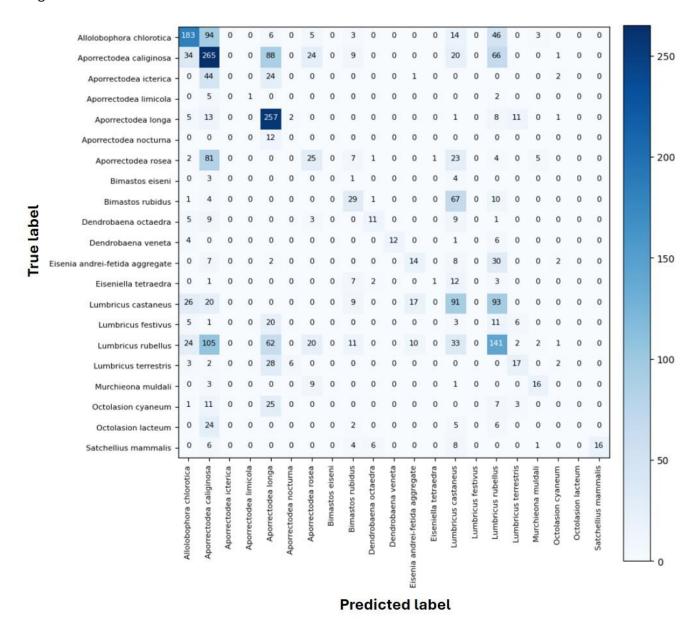


Figure 8: Confusion matrix for algorithm of earthworm classifications. Colour indicates the number of images in the test dataset.

Six species (Aporrectodea icterica, Aporrectodea nocturna, Bimastos eiseni, Lumbricus festivus, Octolasion cyaneum and Octolasion lacteum) were never classified correctly. Interestingly, these are not necessarily the species that the author (an experienced earthworm determiner) would have predicted to the algorithm to have struggled with correctly classifying. This suggests that issues with the image collection protocol and algorithm development could be impacting the performance algorithm and further investigation is recommended to establish if overall accuracy can be improved by improving the way in which image data is collected and used to train the algorithm.

We can also calculate various statistics to assess the performance of the algorithm for individual species to investigate how variable the accuracy is across different species.

The **precision** gives us the probability that when the algorithm suggests a species, it is correct. It is calculated using the following formula.

$$Precision = \frac{Number\ of\ true\ positive}{Numbr\ of\ true\ positives + Number\ of\ false\ positives}$$

The **recall** gives the probability that given an image of a particular species, the algorithm will get it right. It is calculated using the following formula.

$$Recall = \frac{Number\ of\ true\ positive}{Numbr\ of\ true\ positives + Number\ of\ false\ negatives}$$

The **f1-score** combines both precision and recall and symmetrically represents them via a harmonic mean to give us a simplified overall measure of accuracy for any given species. It is calculated using the following formula.

$$f1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

f1-scores can range from 0 to 1, with a score of 1 for any given species representing an algorithm that will always identify that species correctly. Conversely, a score of 0 for a given species represents an algorithm that is unable to ever identify a species correctly.

The precision, recall and f1 scores are presented alongside the image test set size in **Table 4**.

Table 4: Species level assessment of algorithm performance for earthworm species classification from images.

Species	Precision	Recall	f1-score	Test set size
Dendrobaena veneta	1	0.52	0.69	23
Aporrectodea longa	0.49	0.86	0.63	298
Allolobophora chlorotica	0.62	0.52	0.57	354
Murchieona muldali	0.59	0.55	0.57	29
Satchellius mammalis	1	0.39	0.56	41
Aporrectodea caliginosa	0.38	0.52	0.44	507
Dendrobaena octaedra	0.52	0.29	0.37	38
Lumbricus terrestris	0.44	0.29	0.35	58
Lumbricus castaneus	0.3	0.36	0.33	256
Lumbricus rubellus	0.32	0.34	0.33	411
Bimastos rubidus	0.35	0.26	0.30	112
Eisenia andrei/fetida agg	0.33	0.22	0.27	63
Aporrectodea limicola	1	0.12	0.22	8
Aporrectodea rosea	0.29	0.17	0.21	149
Eiseniella tetraedra	0.50	0.04	0.07	26
Aporrectodea icterica	0	0	0	71
Aporrectodea nocturna	0	0	0	12
Bimastos eiseni	0	0	0	8
Lumbricus festivus	0	0	0	46
Octolasion cyaneum	0	0	0	47
Octolasion lacteum	0	0	0	37

The size of the training dataset does not appear to have a strong effect on f1-scores. However, the size of the training dataset does impact specific misclassification rates as an imbalanced training dataset leads the algorithm to predict common species (i.e. those with a larger training dataset) when it is uncertain, enabling the algorithm to 'play the odds'. For example, the largest image dataset for a species was for *Aporrectodea caliginosa* and we can see that other species are commonly misclassified as this species within **Figure 8**.

8 Conclusions

This proof-of-concept study has demonstrated that:

- 1. It is possible to recruit volunteers to collect a large number of images of earthworms using a detailed protocol.
- 2. These images can be used as a training dataset by combining collection with laboratory identification.
- 3. The collection protocol should be refined to avoid the over-handling of specimens, keep the collection tray clean and avoid strong shadows.
- 4. The image set is strongly biased towards common species and more sampling effort will be required to balance the existing library.
- 5. Using a segmentation algorithm to mask out the earthworms prior to classification may improve the performance of the algorithm.
- 6. Species that bare proving too difficult for the algorithm to classify should be grouped into aggregates to improve the overall results.
- 7. Computer vision models show some promise but more development is needed for a usable solution.

9 Bibliography

- Brown, K. D. (2019). *Earthworm Recorders Handbook [Version 8]*. London: Earthworm Society of Britain. Retrieved from https://www.earthwormsoc.org.uk/handbook
- Brown, K. D. (2022). *Soil Pit Survey Form [Version 1]*. 2022: Earthworm Society of Britain. Retrieved from https://www.earthwormsoc.org.uk/SoilPitSurveyForm
- Brown, K. D. (2022). *Verification*. Retrieved from Earthworm Society of Britain: https://www.earthwormsoc.org.uk/verification
- Carpenter, D., Sherlock, E., Jones, D. T., Chiminoides, J., Writer, T., Neilson, R., . . . Eggleton, P. (2012). Mapping of earthworm distribution for the British Isles and Eire highlights the under-recording of an ecologically important group. *Biodiversity Conservation*, 21, 475-485. doi:10.1007/s10531-011-0194-x
- Eggleton, P., Inward, K., Smith, J., Jones, D. T., & Sherlock, E. (2009). A six year study of earthworm (Lumbricidae) populations in pasture woodland in southern England shows their response to soil temerature and soil moisture. *Soil Biology & Biochemistry, 41*, 1857-1865. doi:10.1016/j.soilbio.2009.06.007
- Jones, D. T., & Eggleton, P. (2014). *Earthworms in England: distribution, abundance and habitats (Natural England Commissioned Report NECR145)*. Natural England. Retrieved from http://publications.naturalengland.org.uk/publication/5174957155811328
- Keith, A. H., & Robinson, D. A. (2012). Earthworms as Natural Capital: Ecosystem Service Providers in Agricultural Soils. *Economology Journal*, *2*, 91-99. Retrieved from https://nora.nerc.ac.uk/id/eprint/15994
- Sherlock, E. (2018). *Key to the Earthworms of the UK and Ireland* (2nd ed.). Shrewsbury: Field Studies Council.
- Sims, R. W., & Gerard, B. M. (1999). *Earthworms. Synopses of the British Fauna (New Series)* (Vol. 39). London: Linnean Society of London.