



LJMU Research Online

Lallensack, JN, Romilio, A and Falkingham, PL

A machine learning approach for the discrimination of theropod and ornithischian dinosaur tracks

<http://researchonline.ljmu.ac.uk/id/eprint/18087/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

Lallensack, JN, Romilio, A and Falkingham, PL (2022) A machine learning approach for the discrimination of theropod and ornithischian dinosaur tracks. Journal of the Royal Society Interface, 19. ISSN 1742-5662

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>

Pre-print required by UK regulations. This is an unformatted manuscript. The final, formatted, version is available at <https://royalsocietypublishing.org/doi/10.1098/rsif.2022.0588>

A machine learning approach for the discrimination of theropod and ornithischian dinosaur tracks

Jens N. Lallensack^{1*}, Anthony Romilio², and Peter L. Falkingham¹

1: School of Biological & Environmental Sciences, Liverpool John Moores University, James Parsons Building, Bryon Street, Liverpool, L3 3AF, UK

2: School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, Australia

*corresponding author, email: jens.lallensack@gmail.com

ORCID:

Jens N. Lallensack: 0000-0003-4104-5232

Anthony Romilio: 0000-0002-1218-3567

Peter L. Falkingham: 0000-0002-1218-3567

Abstract

Fossil tracks are important palaeobiological data sources. The quantitative analysis of their shape, however, has been hampered by their high variability and lack of discrete margins and landmarks. We here present the first approach using deep convolutional neural networks (DCNNs) to study fossil tracks, overcoming the limitations of previous statistical approaches. We employ a DCNN to discriminate between theropod and ornithischian dinosaur tracks based on a total of 1372 outline silhouettes. The DCNN consistently outperformed human experts on an independent test set. We also used the DCNN to classify tracks of a large tridactyl trackmaker from Lark Quarry, Australia, the identity of which has been subject to intense debate. The presented approach can only be considered a first step towards the wider application of machine learning in fossil track research, which is not limited to classification problems. Current limitations, such as the subjectivity and information loss inherent of interpretive outlines, may be overcome in the future by training neural networks on 3D models directly, though this will require an increased uptake in digitization among workers in the field.

Keywords: Computer vision, artificial intelligence, fossil footprints, subjectivity, objective quantitative analysis

Introduction

Distinguishing between tridactyl (three-toed) dinosaur tracks of the herbivorous ornithischians and the predominantly carnivorous theropods is a complex and long-standing problem [1–9]. Broadly, ornithischian tracks are expected to be wider and more symmetric than theropod tracks, with digit impression III less projecting beyond digit impressions II and IV, and with digit impressions being broader, more splayed apart, and terminating in blunt hoof marks instead of sharp claw marks. However, any of these characteristics can be found in both groups, and which are the most important depends on the particular track type in question. Moratalla et al. [1] presented a quantitative approach to discriminate these groups, albeit limited to larger theropod and ornithopod tracks. Limitations of this approach include the small sample size, issues with the measurement scheme, and omission of relevant shape characteristics [1,5,6]; despite this, the method has found wide application [3,9–12].

To overcome the limitations of previous statistical approaches, and to remove as much subjectivity as possible, we trained and then employed an artificial neural network to categorise outlines of tridactyl dinosaur footprints as theropod or ornithischian. Artificial neural networks are a type of non-linear model that can learn from data, and a principal component of machine learning and artificial intelligence (AI). Inspired by the structure of the human brain, such neural networks comprise interconnected nodes (or neurons), with each connection represented by a number (weight). The DCNN learns from its mistakes during multiple training iterations in which the weights are optimised to achieve an increasingly good fit to the data structure [13]. We here employ a deep convolutional neural network (DCNN) comprised of multiple layers of neurons (deep learning) [13].

We applied our model to an independent test set and compared its identifications with those of five human experts. We then used the neural network to assess outlines from a large trackway from Lark Quarry, Australia; a tracksite also known as the Dinosaur Stampede National Monument (Australian Natural Heritage List Place ID 105664). The identity of the largest trackmaker has been the subject of intense discussion, particularly in the past decade, namely as having been either a predatory theropod dinosaur or a herbivorous ornithischian [3,5,14–20]. The ambiguity of the footprint shapes and lack of agreement amongst track-workers make this trackway an ideal case study for our machine learning approach.

Materials and Methods

Standardised track silhouettes

Neural networks can work with a variety of input data, such as simple measurements, categorical data, 2D images, and even 3D models. For best performance, however, it is crucial to maximize relevant information while minimizing irrelevant information.

Moratalla et al. [1] present both a quantitative approach and a list of qualitative criteria to discriminate theropod and ornithischian tracks. Linear and angular measurements, as used by Moratalla et al. [1], may be selected to capture what are believed to be the most relevant shape features of a track. However, such measurements may not capture all relevant details, such as the shape of claw marks. Furthermore, such measurements can be difficult to apply consistently to a large number of tracks of varying shapes [21].

On the other extreme to linear measurements, 3D models capture a maximum of shape information [22], but are highly complex and include a lot of information irrelevant to the identification of the tracks (e.g., the presence of cracks or ripple marks; what some authors term ‘extramorphological’ features). Although neural networks are able to become invariant to such irrelevant information, they require a larger sample size to do so.

In order to minimize irrelevant information and maximize sample size, we use black-and-white silhouettes of interpretive track outlines. Using silhouettes as input data avoids the need for measuring particular anatomical features, allowing for inclusion of any track shape. Furthermore, silhouettes may capture most of the information that is considered relevant for the identification of tracks, including features such as digit terminations and the number and shape of phalangeal pads. However, such outlines have been repeatedly criticised because of their inherent subjectivity and simplification of complex 3D morphologies [e.g., 19,23].

Published outline drawings can be complex and may consist of multiple and/or discontinuous lines, dotted lines to indicate uncertainty, internal shadings, and other features. Such complexity does often bear the artistic signature of the individual ichnologist. By converging such drawings into simpler black-and-white silhouettes, the artistic signature is mostly removed, avoiding irrelevant detail that may deteriorate the performance of the model. Silhouettes may be represented as a continuous outline or as a set of separate outlines that may show individual pad impressions.

Conversion of published outlines into our standard silhouette format often required some degree of interpolation; e.g., dotted lines were treated as solid lines, and gaps in the outline were closed. We included tracks of functionally tridactyl ornithischian (basal ornithischians, ornithopods, and stegosaurs) and non-avian theropod dinosaurs spanning from the Upper Triassic to the Upper Cretaceous. This included tracks that show the impression of digit I in addition to digits II–IV, as long as this digit is unlikely to have played a major role in locomotion in the trackmaker. We did not include tracks interpreted to have been made by partially buoyant animals, nor did we include tracks that show substantial incompleteness, but we include cases in which the track walls have collapsed, as in penetrative tracks [24–26]. Some of the included tracks contain little relevant anatomical detail, or even misleading features, due to unfavourable substrate conditions, trackmaker behaviour, or postformational alteration of the tracks. By training the model on a sufficiently large sample size, this ambiguity is hoped to be reflected by a decreased confidence of the model when classifying such tracks.

The majority of outlines were collected from the literature, building on a dataset previously collected by Lallensack et al. [7]. Multiple outlines from single trackways were included where available, and, in some cases, different interpretations of the same track were included. This data is supplemented with outlines drawn by one of us (JNL) from 3D models that are published as part of this contribution or have been published elsewhere. The silhouettes are rotated upright (relative to digit impression III) before analysis and downscaled to 100x100 pixels, with the shape centred in the image. The model was made invariant to subtle variations in rotation and position (see below).

Our original data set consisted of 677 ornithischian and 959 theropodan examples (see supplemental data). This imbalance between the categories is caused by the much greater abundance of theropod tracks in the fossil record, even though attempts were made to include as many ornithischian tracks as possible. Models trained on the full imbalanced data set tended to perform well on large tracks, but appeared to be biased towards theropods when testing ambiguous small tracks similar to the ichnotaxon *Anomoepus*. We consequently applied undersampling (see, e.g., [27]) to arrive at a balanced data set by randomly removing theropodan examples until their number was equal to that of ornithischian examples.

Model training

The model training was done using the open-source machine learning library TensorFlow (www.tensorflow.org; version 2.9), which was controlled through the Keras interface (<https://keras.io>) using Python scripting.

The model architecture employed, and the parameters used, may drastically influence the performance of a model. Precise optimisation of models by finding the best parameters is often impractical given the number of possibilities that can be set. We tested separate model architectures of increasing complexity, different numbers of epochs (training iterations during which the model sees all data in the training set) and batch sizes (the number of silhouettes the model will see at the same time) on test sets that were randomly separated from the training set. We used loss rather than accuracy to select models [28] and discarded those that showed a problematic degree of overfitting. Following these tests, we selected a modification of the VGG16 architecture with four convolutional

layers with 32, 64, 128, and 128 neurons, respectively. Pooling layers were used in-between these layers, and a dense layer of 256 neurons was used for flattening. After each pooling layer and after the 256 neuron dense layer, we use a dropout of 0.2 [29]. We selected “Adams” as optimizer and a batch size of 30. We used early stopping to select the model version at the epoch where loss was minimal [30]. For our selected model, this mechanism led to a training period of 74 epochs, after which the validation loss was at 0.37 and the validation accuracy at 0.83.

We used online data augmentation in order to make the model more invariant against slight variations in rotation, position, and the exact configuration of pixels of the shape margins, all of which are not informative for classification purposes. This was achieved by random horizontal flipping; rotation by a random value between -30 and +30 degrees; and a slight random sift in x and y direction. These operations were performed on each image before each epoch – as our model was trained with 74 epochs, 74 versions per image were created.

For each tested silhouette, the neural network returns a numerical value that ranged between 0 and 1, to denote confidence of track affinities as ornithischian and theropodan, respectively. A value of 0.5 indicates an ambiguous outcome without any tendency towards either category. The neural network model, along with detailed step-by-step instructions, is provided in the supplemental data.

The final model was validated on a test set of 36 tracks that were collected by JNL from the literature (see supplemental data). Tracks were selected to obtain a varied sample that includes both relatively obvious candidates and challenging ones. In a few cases, the silhouettes do seem difficult or even impossible to assign to their correct label, as they show characteristics of both groups. The same test set was given to five researchers who all are established experts on tridactyl dinosaur tracks. These researchers were asked to identify each track either as ornithischian (0), theropod (1), or ambiguous (0.5).

Results and Discussion

Performance on test set and comparison with human performance

On the test set, the neural network outperformed human experts by a constant margin (Fig. 1, Tab. 1). 86% of tracks were correctly classified while 14% were incorrectly classified. If all neural network decisions above 0.4 and below 0.6 are treated as ambiguous (22% of tracks), the neural network classified 67% of tracks correctly and 11% incorrectly. The human experts, on average, classified 57% correctly, 20% incorrectly, and 24% as ambiguous. These results demonstrate the ability of big data applications to outperform human experience when trained on a sufficient sample size. Most of the test set examples are significantly different from any of those the model was trained on, demonstrating the ability of the model to classify new distinct shapes.

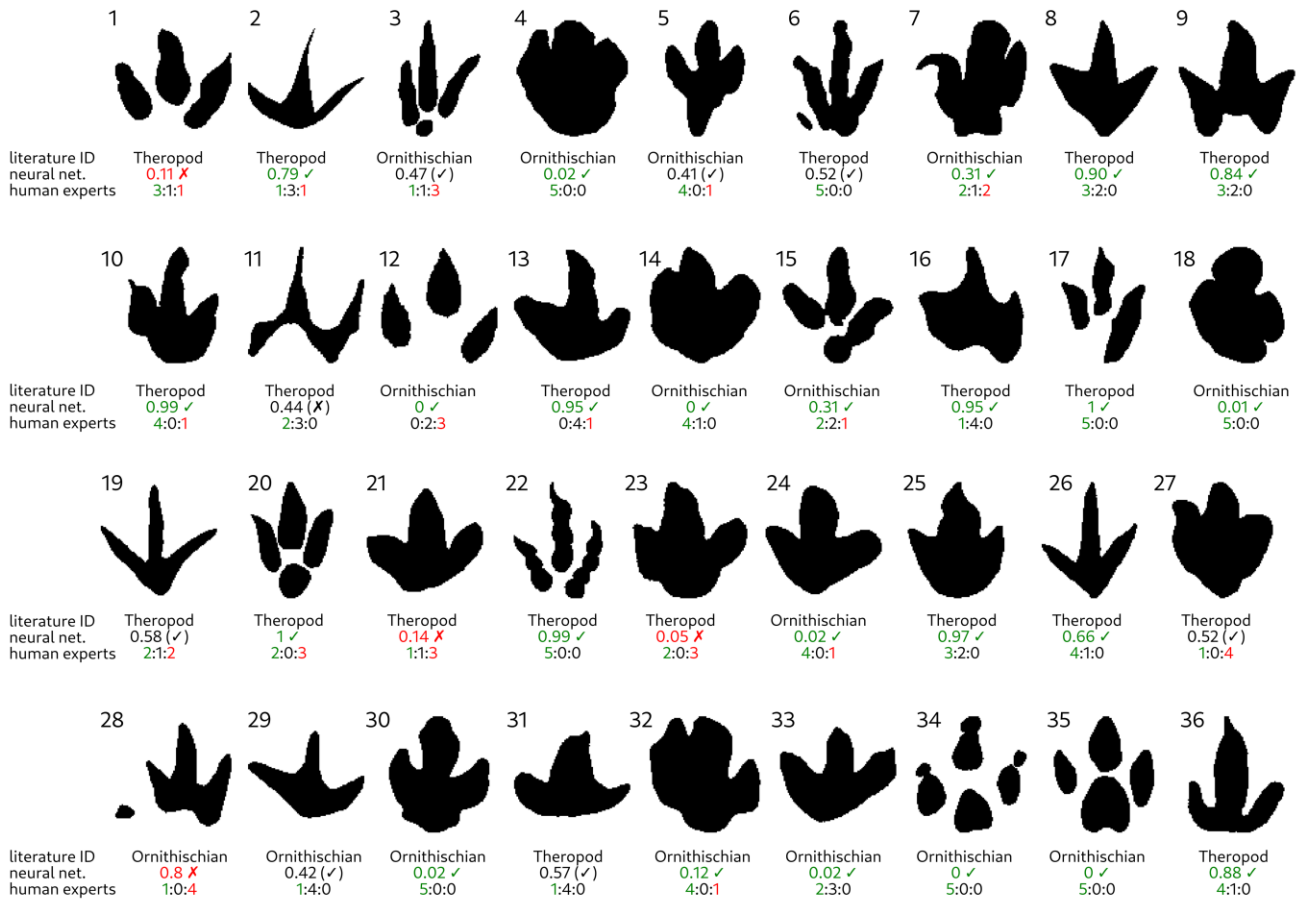


Fig. 1. The test set used to compare the performance of the neural network model with that of five human experts. The neural network returned values between 0 and 1, with values <0.5 indicating more ornithischian-like shapes and values >0.5 indicating theropod-like shapes. Values between 0.4 and 0.6 are here considered as “ambiguous”. Human experts marked each track as either “ornithischian”, “ambiguous”, or “theropodan”. The ratio between correct, ambiguous, and incorrect identifications of these five experts is shown for each track.

Identifier	% correct	% ambiguous	% incorrect
Expert 1	67	3	31
Expert 2	58	25	17
Expert 3	58	25	17
Expert 4	42	44	14
Expert 5	58	22	19
Neural Network (no ambiguous cases)	86	0	14
Neural Network (0.4–0.6 treated as ambiguous)	67	22	11

Table 1: Performance of the individual human experts and the neural network on the test set. Note

that high percentages of correct identifications are associated with high percentages of incorrect identifications, and vice versa.

Limitations and strengths

Our neural network was trained to classify shapes without any context information. When an ichnologist decides how to assign a track, such context information (e.g., size, stratigraphy, and shape of other tracks of the same trackway) plays an important role. While it is generally possible to provide a neural network such context information, this comes at a cost: The model will be more likely to overfit, or, in other words, may fail to make a correct assignment when the track in question was found outside of its expected context. This may lead to circular arguments: If, for example, information on size and geological age is provided to the model, large tridactyl tracks from the Upper Jurassic are more unlikely to be classified as ornithischians because no such examples are present in the training set. This may easily lead to false conclusions if the *a priori* assumption (no large ornithischians in the Upper Jurassic) is wrong. In this light, this apparent shortcoming of our neural network may therefore be its greatest strength: providing objective assessments of shape independent of context. Consequently, it has to be the job of the ichnologist to combine the neural network evaluation of the shape with all relevant context information to arrive at a meaningful interpretation of the track.

In the initial stages of the present project, the authors were undecided on whether to include size information. Size is undeniably important in some cases – e.g., a 50 cm long slender-toed track can confidently be assigned to a theropod trackmaker, even though the same shape at a much smaller scale might well be produced by an ornithischian trackmaker. On the other hand, we may argue that tracks that are unambiguously classifiable based on size do not require the judgement of our neural network to start with. In all ambiguous cases, evaluating shape independent of size can yet again be considered a strength, as it brackets off *a priori* assumptions about the possible size range of a trackmaker, which may be based on incomplete knowledge. This reasoning notwithstanding, we did collect size information along with all outlines of our dataset. Tests using an earlier model version that included scale information did not appear to produce better results than other models that did not include such information, and we consequently abstained from including this information. Another difficult decision is the identification of a track as belonging to the left or right foot. Tests using early model versions did not show clear improvements in performance when presented with uniform right shapes versus randomly mirrored shapes, and most tracks of unknown trackmaker identification cannot be unambiguously attributed to left or right to start with. Consequently, we decided to augment the data to randomly mirror the shapes, making the model invariant against left or right.

Another inevitable limitation is our reliance on previous identifications of tracks as either “theropod” or “ornithischian” to train the models. We assume that the vast majority of these *a priori* identifications are probably correct even when the shapes themselves are ambiguous, since the context usually offers valuable clues (most unambiguously the presence of manus impressions, which only occur in ornithischians). However, misidentifications cannot be ruled out especially for smaller tracks which are particularly ambiguous.

Finally, the greatest limitations of the presented approach are the interpretive outline silhouettes themselves. Advantages of using such silhouettes, namely their simplicity and wide availability, are

compromised by their inherent subjectivity [19,22]. This subjectivity is highlighted by Fig. 2, where outlines produced by different researchers may be different in shape to a problematical degree [23,31]. Besides adding substantial noise, this subjectivity may shift the neural network evaluation towards one or the other category, as is the case with track 3 (Fig. 2, outline of PLF and JNL vs that of AR). Furthermore, given the loss of potentially valuable information, the reduction of information content can be considered an advantage and a drawback at the same time. Although extramorphological features tend to be excluded from interpretive outline drawings to arrive at a meaningful hypothesis about the shape of the foot, this is not always possible, and the degree to which outlines are “idealised” varies greatly between separate ichnologists. Indeed, some authors [e.g., 32] have argued that ‘extramorphological’ features should be considered part of the track itself (e.g. displacement rims, collapse structures), and that these features may contain information about the trackmaker that should not be excluded.

As we authors experienced with the Lark Quarry trackway (see below), the production of meaningful outlines was not always possible, and identifying which features were related to the foot morphology, and which to sediment movement, and later erosion, was difficult. Given these drawbacks, our approach using interpretive outlines can only be considered the first step, and more objective and comprehensive data formats, such as 3D shapes, need to be considered for future applications of neural networks (see below).

Analysis of Lark Quarry tracks

Previous debate on the identification of the large tridactyl trackway Q1 from Lark Quarry was complicated by two points of disagreement: 1) which tracks best reflect the anatomy of the foot, and 2) how the outline of the individual track should be drawn. The interpretation of outlines of individual tracks has been even more controversial and has been argued to be significantly influenced by the researcher’s opinion about the trackmaker responsible [3,5,16–20,23].

We here consider all tracks of the Lark Quarry trackway except for tracks 7, 9, and 11, as these tracks lack an obvious tridactyl morphology. Outlines were drawn by all three of us based on digital models made from the archival cast material, before damage and reconstruction occurred to many of the tracks *in situ*. It should be noted that in previous work, PLF favoured a theropod affinity [19] and AR favoured an ornithopod affinity [3,16], while JNL considered the tracks to be ambiguous. In addition, we produced objective outlines of all tracks using the algorithm of Lallensack (2019), run with default settings. These objective outlines mimic the way humans interpret outlines but are by design devoid of *a priori* assumptions about the trackmaker.

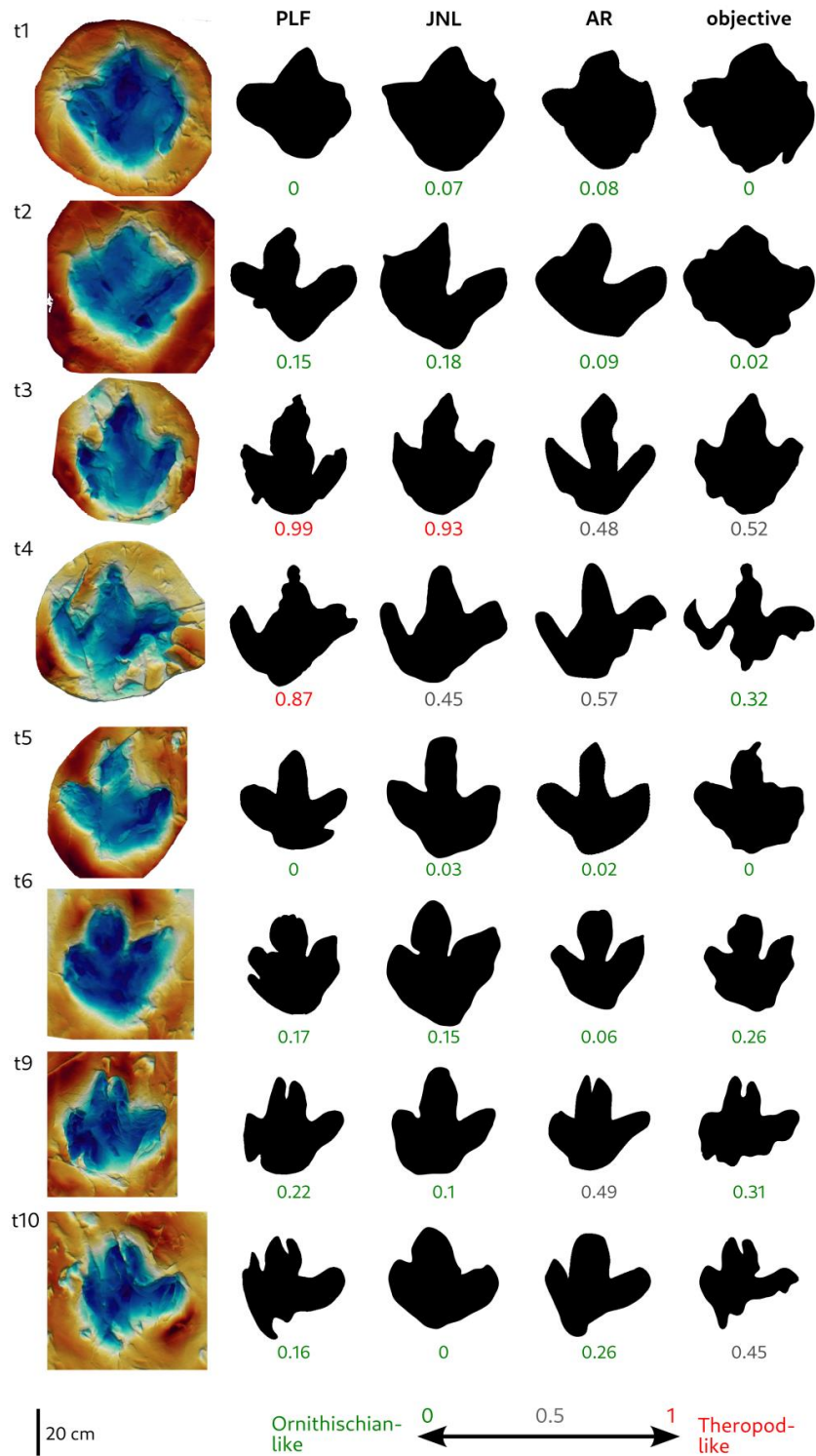


Fig. 2. Tracks 1–6, 9, and 10 of trackway LQ1 from Lark Quarry, Australia. Height maps of 3D models are shown together with respective interpretive outlines produced by the three authors (PLF, JNL, and AR) and the objective outline algorithm of [23]. Note the marked differences in shape and size between outlines, in particular of track 3. Scores returned by our neural network indicate that most tracks with the exception of track 3 are ornithischian-like.

For six of the eight tracks, the neural network tended to be in favour of an ornithischian affinity (Fig. 2). The mean score of PLF, who favoured a theropod affinity of the tracks, was slightly higher (0.32) than that of AR (0.26), JNL (0.24), and the objective outlines (0.24). More substantial differences were found for track 3, which Thulborn and Wade (1984) considered the best-preserved track: Outlines by PLF (0.99) and JNL (0.93) were found to be very theropod-like, while outlines by AR (0.48) and the objective outlines (0.52) were ambiguous. This indicates that although *a priori* assumptions about the responsible trackmaker may indeed influence the result in some cases, general trends seem to be robust (i.e., this track was the most theropod-like regardless of bias).

However, the mean of scores of all considered tracks is not necessarily an informative measure, as in many cases only one or few tracks will convey the anatomy of the foot accurately. We therefore continue by taking a closer look at those tracks that have been identified as the best preserved by previous workers: track 3, 5, and 6. Track 3 is clearly different in appearance from all other tracks, and is consistently recovered as the most theropod-like. Analysis with gradcam [33] of the outline of JNL of track 3 (Fig. 3A) suggests that the curvature of digit impression III and the shape of the heel are the important features that led to its classification as theropod. The curvature of digit III is also indicated by longitudinal ridges at the floor and wall of the track that might represent pull-up structures, indicating that this curvature is original and not the result of erosion [16].

Tracks 5 and 6, which are unambiguously ornithopod-like, are similar to some of the other, more fragmented tracks. A direct overlay of interpretive “maximum” outlines of tracks 3 and 5 (Fig. 3C) shows that the latter is distinctly wider with larger interdigital angles and a generally more robust appearance. Both tracks are, however, almost identical in length. If track 3 does reflect the foot anatomy more closely, it must follow that other tracks were broadened due to erosion or rotation of the foot. If, on the other hand, tracks 5 and 6 more closely match the foot anatomy, it follows that digits II and IV of track 3 are partially collapsed during or after track formation.

While morphologies of most tracks other than track 3 tend to suggest an ornithischian trackmaker, the long stride lengths and low pace angulation values of around 180° [16] are more typical for a theropod trackmaker. We may also be running into the limitations of using outlines for relatively deep tracks, where outlines created near the top of the track are broader and more rounded (and hence more ornithopodan) than outlines from deeper in the track, at the foot-sediment interface, as demonstrated by Falkingham [19]. This is confirmed by testing (subjectively selected) “maximum” and “minimum” contour lines of track 3 (Fig. 3B), where the maximum outline was ambiguous (score of 0.51) and the minimum outline very theropodan-like (score of 0.98). Future studies involving neural networks trained on 3D models of tracks and/or trackway parameters may be able to resolve this ambiguity, as we shall come to in our conclusions.

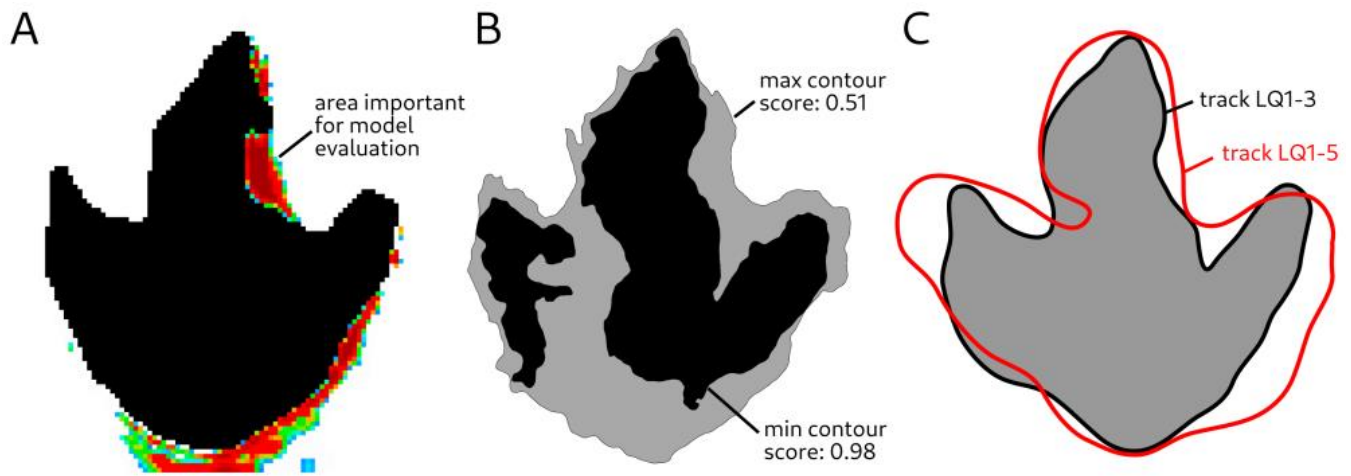


Fig. 3. Analysis of Lark Quarry track 3. A: Gradcam [33] map superimposed on the interpretive outline of JNL, highlighting the areas that were important for the classification of this shape as theropod. B: A (subjectively chosen) ‘minimum’ contour line superimposed on a ‘maximum’ contour line. The minimum contour line gives neural network scores close to 1, suggesting a theropod trackmaker, while the maximum contour gives an ambiguous score, suggesting that contours become more theropod-like with increasing depth. C: Interpretive outline of track 5 superimposed on that of track 3, showing marked differences in width but not in length.

Re-evaluation of other tracks and implications on trackmaker identification

Shapes attributed or similar to those of the ornithischian ichnogenus *Anomoepus* are often recovered as ambiguous by our model. *Anomoepus* tracks are typically identified based on the presence of manus impressions, inward rotation of the pes, and low relative projection of digit III beyond digits II and IV [34]. The ambiguity expressed by our model, however, suggests that *Anomoepus* cannot be reliably distinguished based on pes track shape alone when context data is not available. Such difficulties may be illustrated by an isolated track from the Hettangian of Poland that was described as a particularly large example of the ornithischian ichnogenus *Moyenisauropus* (= *Anomoepus*) [35]. Our neural network, however, suggests that a theropod affinity is more likely (0.72). A trackway of three tracks from the Lower Jurassic (Sinemurian) Razorback beds of Australia had been traditionally interpreted as theropodan in origin [36,37], while a more recent evaluation identified these tracks as ornithischian *Anomoepus* tracks [38]. Our neural network confirms the latter interpretations (scores of 0.42, 0.44, and 0.32 for the three tracks, respectively).

Large tridactyl tracks, such as those of trackway 1 from Lark Quarry, pose different problems. Large ornithopod tracks are primarily differentiated from large theropod tracks based on their overall wide shape and their broad and rounded toe impressions that lack claw traces. Similar ornithopod-like features can, however, occur due to unfavourable substrate properties and/or within-sediment rotation of the foot, while theropod-like features are more unlikely to result from such factors. Consequently, large tridactyl tracks may be more easily mis-identified as ornithopod than as theropod. Schulp & Al-Wosabi [11] attributed a large tridactyl trackway from the Upper Jurassic of Yemen to an ornithopod trackmaker, although its large size (track length >50 cm) is unusual for

ornithopods of that age. This was subsequently questioned by Piñuela et al. [8], who argued that features of *some* of the tracks are theropod-like. Our neural network indeed suggested an ornithopod affinity for most tracks but two (tracks 5 and 13), which were more theropod-like (0.59 and 0.71, respectively). As in the case of Lark Quarry, the identification of the trackway will ultimately depend on the question which of these tracks do more closely match the trackmaker's foot, which cannot be fully addressed without careful restudy of the tracks, or 3D models thereof.

Future research

The value of fossil tracks as palaeontological data sources cannot be overstated, yet their interpretation can be challenging. The shape of tracks is influenced by multiple factors, including anatomy, substrate properties, and behaviour, but also post-formational alteration [32,39]. Although tracks are organised into separate ichnotaxa based on shape features thought to reflect anatomy, these ichnotaxa in fact form a continuum of shapes rather than discrete classes [21]; such continua must be analysed using rigorous quantitative methods. Neural networks may overcome limitations of previous quantitative approaches and are able to adapt to the immense complexity of the data. The application of neural networks may go well beyond simple classification tasks as carried out here. By learning to distinguish relevant from irrelevant shape features, these methods could be used to constrain foot shape, movement, and sediment properties for a given track or trackway, and even to produce outline drawings and measurements from 3D models that capture the important shape features better than any current objective outline technique [19,23]. Neural networks, in principle, and as demonstrated here, have the ability to outperform human experts in any of these tasks.

Traditional ichnology has been based on interpretive outline drawings in order to document and communicate track morphologies. Such outline drawings are, however, deeply problematic for quantitative analysis because of 1) their subjectivity, 2) the loss of 3D information, and 3) the abstraction of 3D shapes to a set of margins which often do not exist in reality. While outlines drawn by a single well-versed researcher might be considered to be consistent, outlines of separate researchers are certainly not. Subjective decisions during drawing outlines include the precise position of the track margin [19,23], but also the degree of "improvement" to interpolate outlines, remove post-formational alterations, and to highlight anatomical features of interest. These subjective decisions can result in strikingly different outline tracings of a single track [e.g., 19,23]. Because such decisions may be influenced by *a priori* assumptions, in particular about the trackmaker taxon responsible, the danger of circular arguments when analysing these outlines is real. Despite this, there is still value in outline drawings for quantitative analysis, if carried out carefully, as demonstrated in the present approach. Using such drawings can only be the first step, and future work will soon need to switch to 3D models to train such neural networks.

There has been a growing movement toward collecting, and publishing, 3D data, particularly as part of a 'standard ichnological protocol' [22], utilizing methods such as laser scanning, and particularly photogrammetry [40–42]. Long-term digital preservation, however, requires not only the capture of sets of photographs suitable for photogrammetry, but also the storage of such data in curated repositories such as Figshare and Morphosource, or as supplemental data associated with published papers. If kept outside of such repositories, data is likely to be lost eventually after retirement of the researcher. Only if and when 3D data is more widely available, quantitative methods such as neural

networks will be able to fully unlock the potential of fossil tracks as palaeobiological data sources.

Acknowledgements

We wish to thank Diego Castanera, Alexander Wagensommer, and James O. Farlow for completing our test set as “human experts”. We also thank three anonymous reviewers for their valuable comments that improved the manuscript.

Funding

This work was funded by the German Research Foundation (DFG) under grant LA 4611/2-1 to Lallensack.

Supplemental Data

Supplemental data can be found in figshare (<http://dx.doi.org/doi:10.6084/m9.figshare.c.6272858>) and contains the neural network; the full data set used to train and test the neural network; associated metadata and references; and detailed instructions how to use the neural network to classify unknown tracks.

References

1. Moratalla JJ, Sanz JL, Jimenez S. 1988 Multivariate analysis on Lower Cretaceous dinosaur footprints: discrimination between ornithopods and theropods. *Geobios* **21**, 395–408.
2. Demathieu GR. 1990 Problems in discrimination of tridactyl dinosaur footprints, exemplified by the Hettangian trackways, the Causses, France. *Ichnos* **1**, 97–110.
3. Romilio A, Salisbury SW. 2011 A reassessment of large theropod dinosaur tracks from the mid-Cretaceous (late Albian–Cenomanian) Winton Formation of Lark Quarry, central-western Queensland, Australia: A case for mistaken identity. *Cretaceous Research* **32**, 135–142. (doi:10.1016/j.cretres.2010.11.003)
4. Castanera D, Pascual C, Razzolini NL, Vila B, Barco JL, Canudo JI. 2013 Discriminating between medium-sized tridactyl trackmakers: Tracking ornithopod tracks in the base of the Cretaceous (Berriasian, Spain). *PLoS ONE* **8**, e81830. (doi:10.1371/journal.pone.0081830)
5. Thulborn RA. 2013 Lark Quarry revisited: a critique of methods used to identify a large dinosaurian track-maker in the Winton Formation (Albian–Cenomanian), western Queensland, Australia. *Alcheringa: An Australasian Journal of Palaeontology* **37**, 312–330. (doi:10.1080/03115518.2013.748482)
6. Lallensack JN, van Heteren AH, Wings O. 2016 Geometric morphometric analysis of intratrackway variability: a case study on theropod and ornithopod dinosaur trackways from Münchehagen (Lower Cretaceous, Germany). *PeerJ* **4**, e2059.

7. Lallensack JN, Engler T, Barthel HJ. 2020 Shape variability in tridactyl dinosaur footprints: the significance of size and function. *Palaeontology* **63**, 203–228. (doi:10.1111/pala.12449)
8. Pinuela L, Garcia-Ramos JC, Romano M, Ruiz-Omenaca JI. 2016 First record of gregarious behavior in robust medium-sized Jurassic Ornithopods: evidence from the Kimmeridgian trackways of Asturias (N. Spain) and some general considerations on other medium-large ornithopod tracks in the Mesozoic record. *Ichnos* **23**, 298–311.
9. Figueiredo S, Dinis P, Belo J, Rosina P, Bachtsevanidou Strantzali I. 2017 A new record of a possible ornithopod footprint from the Lower Cretaceous of Cabo Espichel (Sesimbra, Portugal). *Bollettino della Società Paleontologica Italiana* **56**, 217–231.
10. Mateus O, Milàn J. 2008 Ichnological evidence for giant ornithopod dinosaurs in the Upper Jurassic Lourinha Formation, Portugal. *Oryctos* **8**, 47–52.
11. Schulp AS, Al-Wosabi M. 2012 Telling apart ornithopod and theropod trackways: A closer look at a large, Late Jurassic tridactyl dinosaur trackway at Serwah, Republic of Yemen. *Ichnos* **19**, 194–198. (doi:10.1080/10420940.2012.710672)
12. dePolo PE *et al.* 2020 Novel track morphotypes from new tracksites indicate increased Middle Jurassic dinosaur diversity on the Isle of Skye, Scotland. *PLOS ONE* **15**, e0229640. (doi:10.1371/journal.pone.0229640)
13. Rawat W, Wang Z. 2017 Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review. *Neural Computation* **29**, 2352–2449. (doi:10.1162/neco_a_00990)
14. Thulborn RA, Wade M. 1984 Dinosaur trackways in the Winton Formation (mid-Cretaceous) of Queensland. *Memoirs of the Queensland Museum* **21**, 413–517.
15. Thulborn RA, Wade M. 1989 A footprint as a history of movement. In *Dinosaur Tracks and Traces* (eds DD Gillette, GM Lockley), pp. 51–56. Cambridge, UK: Cambridge University Press.
16. Romilio A, Salisbury SW. 2014 Large dinosaurian tracks from the Upper Cretaceous (Cenomanian–Turonian) portion of the Winton Formation, Lark Quarry, central-western Queensland, Australia: 3D photogrammetric analysis renders the ‘stampede trigger’ scenario unlikely. *Cretaceous Research* **51**, 186–207. (doi:10.1016/j.cretres.2014.06.003)
17. Romilio A, Tucker RT, Salisbury SW. 2013 Reevaluation of the Lark Quarry dinosaur Tracksite (late Albian–Cenomanian Winton Formation, central-western Queensland, Australia): no longer a stampede? *Journal of Vertebrate Paleontology* **33**, 102–120. (doi:10.1080/02724634.2012.694591)
18. Thulborn RA. 2017 Behaviour of Dinosaurian Track-Makers in the Winton Formation (Cretaceous, Albian–Cenomanian) at Lark Quarry, Western Queensland, Australia: Running or Swimming? *Ichnos* **24**, 1–18. (doi:10.1080/10420940.2015.1129326)
19. Falkingham PL. 2016 Applying Objective Methods to Subjective Track Outlines. In *Dinosaur Tracks: The Next Steps* (eds PL Falkingham, D Marty, A Richter), pp. 72–81. Bloomington: Indiana

University Press.

20. White MA, Cook AG, Rumbold SJ. 2017 A methodology of theropod print replication utilising the pedal reconstruction of *Australovenator* and a simulated paleo-sediment. *PeerJ* **5**, e3427. (doi:10.7717/peerj.3427)
21. Lockley MG. 1998 Philosophical perspectives on theropod track morphology: blending qualities and quantities in the science of ichnology. *GAIA* **15**, 279–300.
22. Falkingham PL *et al.* 2018 A standard protocol for documenting modern and fossil ichnological data. *Palaeontology* **61**, 469–480. (doi:10.1111/pala.12373)
23. Lallensack JN. 2019 Automatic generation of objective footprint outlines. *PeerJ* **7**, e7203. (doi:10.7717/peerj.7203)
24. Gatesy SM, Falkingham P. 2020 Hitchcock's Leptodactyli, penetrative tracks, and dinosaur footprint diversity. *Journal of Vertebrate Paleontology* **e1781142**. (doi:10.1080/02724634.2020.1781142)
25. Falkingham PL, Turner ML, Gatesy SM. 2020 Constructing and testing hypotheses of dinosaur foot motions from fossil tracks using digitization and simulation. *Palaeontology* **63**, 865–880. (doi:10.1111/pala.12502)
26. Carvalho I de S, Cunha PP, Figueiredo SMD. 2022 Dinoturbation in Upper Jurassic siliciclastic levels at Cabo Mondego (Lusitanian Basin, Portugal): evidences in a fluvial-dominated deltaic succession. *Palaeoworld* **31**, 455–477. (doi:10.1016/j.palwor.2021.09.001)
27. Johnson JM, Khoshgoftaar TM. 2019 Survey on deep learning with class imbalance. *Journal of Big Data* **6**, 1–54.
28. Hernández-Orallo J, Flach P, Ferri Ramírez C. 2012 A unified view of performance metrics: Translating threshold choice into expected classification loss. *Journal of Machine Learning Research* **13**, 2813–2869.
29. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. 2014 Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* **15**, 1929–1958.
30. Prechelt L. 1998 Early Stopping - But When? In *Neural Networks: Tricks of the Trade* (eds GB Orr, K-R Müller), pp. 55–69. Berlin, Heidelberg: Springer. (doi:10.1007/3-540-49430-8_3)
31. Thulborn RA. 1990 *Dinosaur tracks*. London, New York: Chapman and Hall.
32. Gatesy SM, Falkingham PL. 2017 Neither bones nor feet: track morphological variation and 'preservation quality'. *Journal of Vertebrate Paleontology* **37**, e1314298.
33. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. 2020 Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Int J Comput Vis* **128**, 336–359.

(doi:10.1007/s11263-019-01228-7)

34. Lockley MG, Klein H, McHugh JB, Romilio A. 2021 Fruita's first fossil footprint exhibit: the discovery of forgotten specimens in an historic former museum building. *New Mexico Museum of Natural History & Science Bulletin – Fossil Record* **7** **82**, 219–226.
35. Gierliński G, Kowalski KZ. 2006 Footprint of a large, Early Jurassic ornithischian from the ancient sacred site of Kontrewers, Poland. *The Triassic-Jurassic Terrestrial Transition. New Mexico Museum of Natural History and Science Bulletin* **37**, 217–220.
36. Staines HRE. 1954 Dinosaur footprints at Mount Morgan. *Queensland Government Mining Journal* **55**, 483–485.
37. Bartholomai A. 1966 Fossil footprints in Queensland. *Australian Natural History* **15**, 147–150.
38. Romilio A. 2020 Additional notes on the Mount Morgan dinosaur tracks from the Lower Jurassic (Sinemurian) Razorback beds, Queensland, Australia. *Historical Biology* , 1–3.
39. Falkingham PL. 2014 Interpreting ecology and behaviour from the vertebrate fossil track record. *Journal of Zoology* **292**, 222–228.
40. Falkingham PL. 2012 Acquisition of high resolution three-dimensional models using free, open-source, photogrammetric software. *Palaeontologia Electronica* **15**, 1–15.
41. Matthews NA, Noble T, Breithaupt BH. 2016 Close-Range Photogrammetry for 3-D Ichnology: The Basics of Photogrammetric Ichnology. In *Dinosaur Tracks: The Next Steps* (eds PL Falkingham, D Marty, A Richter), pp. 28–55. Bloomington, Indiana: Indiana University Press.
42. Lallensack JN, Buchwitz M, Romilio A. 2020 Photogrammetry in ichnology: 3D model generation, visualisation, and data extraction. *Journal of Paleontological Techniques* **22**, 1–18.