Noninvasive low-cost method to identify armadillos' burrows: A machine learning approach

THIAGO F. RODRIGUES,[1],[2] *Applied Ecology Program, Luiz de Queiroz College of Agriculture, University of São Paulo, Av. Pádua Dias 11, Piracicaba, SP 13418-900, Brazil*

KEILLER NOGUEIRA, *Data Science Research Group, Computing Science and Mathematics Division, University of Stirling, Scotland, FK9 4LA, UK*

ADRIANO G. CHIARELLO, *Department of Biology, Faculty of Philosophy, Sciences and Languages of Ribeirão Preto, University of São Paulo, Av. Bandeirantes 3900, Ribeirão Preto, SP 14040-901, Brazil*

ABSTRACT Having accurate information about population parameters of armadillos (Mammalia, Cingulata) is essential for the conservation and management of this taxon, most species of which remain poorly studied. We investigated whether we could accurately identify 4 armadillo species (*Euphractus sexcinctus*, *Dasypus novemcinctus*, *Cabassous tatouay*, and *Cabassous unicinctus*) based on their burrow morphometry. We first selected published studies that reported measurements of width, height, and angle of the burrows used by the 4 species of aradillos. Then, using such data we simulate burrow measurements for each of the 4 species of armadillos, we created predictive models through supervised machine learning that were capable of correctly identifying the species of armadillos based on their burrows' morphometry. By using classification algorithms such as Random Forest, K-Nearest Neighbor, Support Vector Machine, Naive Bayes, and Decision Tree C5.0, we achieved the overall accuracy for the classification task by about 71%, including an overall Kappa index by about 61%. *Euphractus sexcinctus* was the most difficult species to discriminate and classify (approximately 68% of accuracy), whereas *C. unicinctus* was the easiest to discriminate (approximately 93% of accuracy). We found that it was possible to identify similar-sized armadillos based on the

This is the peer reviewed version of the following article: Rodrigues, T.F., Nogueira, K. and Chiarello, A.G. (2021), Noninvasive Low-cost Method to Identify Armadillos' Burrows: A Machine Learning Approach. *Wildlife Society Bulletin*, 45: 396-401, which has been published in final form at https://doi.org/10.1002/wsb.1222. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for self-archiving.

measurements of their burrows described in the literature. Finally, we developed an R function (armadilloID) that automatically identified the 4 species of armadillos using burrow morphology. As the data we used represented all studies that reported the morphometry of burrows for the 4 species of armadillos, we can generalize that our function can predict armadillo species beyonour data.

KEY WORDS burrow, conservation, *Cabassous unicinctus*, *Cabassous tatouay, Dasypus novemcinctus*, *Euphractus sexcinctus*, mammal, noninvasive method, Xenarthra

Estimating population parameters for wildlife is one of the primary interests of scientists and conservationists because of its decisive influence on wildlife management and conservation. However, estimating such parameters (e.g., density, abundancy, occurrence) is not straightforward, particularly for wide-ranging, low-density, elusive, and unstudied species, most of which are threatened (Schipper et al. 2008, Desbiez et al. 2018). For example, according to the IUCN Red List of Threatened Species, 10 out of the 20 extant armadillo species had unknown population trend information, whilst the other 6 were in decline; 5 species were categorized as Data Deficient, 5 as Near Threatened, and 2 as Vulnerable (IUCN 2019). Because it is not practical to employ tagging methods to derive armadillo population estimates due to logistical difficulties, high costs, and small numbers of possible captures (Loughry and McDonough 2013, Desbiez et al. 2018), less than 20% of armadillo studies were based on fieldwork conducted on wild populations (Superina et al. 2014). Alternatively, noninvasive methods such as camera trapping have been successfully used to access occupancy for armadillo species (Zimbres et al. 2013, Rodrigues and Chiarello 2018). However, camera-survey methods require a large number of cameras and high associated costs (e.g., Rodrigues and Chiarello 2018), cameras are vulnerable to animal damage, adverse weather, and theft or vandalism, and such methods are likely produce a small number of detections, particularly for rarer species (Maccarini et al. 2015, Desbiez et al. 2018).

Although armadillos are observed in nature above ground, they dig burrows for shelter (housing, raising offspring), protection (to hide from predators and to buffer against environmental temperatures), and feeding (foraging burrows) (McNab 1980, Eisenberg and Redford 1999, Desbiez et al. 2018). Burrows created by armadillos have specific shapes and sizes (Carter and Encarnaçao 1983, Abba et al. 2005, Trovati 2015, Attias et al. 2016, Desbiez et al. 2018), which are typically influenced by anatomical and morphological differences among species (e.g., Carter and Encarnaçao 1983, Attias et al. 2016). Several previous studies have used armadillo burrows to estimate population parameters including habitat use and density and to assess activity and behavior (Zimmerman 1990, McDonough et al. 2000, Abba et al. 2005, 2007, Desbiez et al. 2018). However, identifying armadillos based on their burrows is challenging (McDonough et al. 2000, Arteaga and Venticinque 2010). For instance, Carter and Encarnação (1983) fitted 4 species of armadillos (*Cabassous tatouay*, *Cabassous unicinctus*, *Euphractus sexcinctus*, and *Priodontes maximus*) with radio transmitters and found that the shape of the burrow entrances differed between the species. Not surprisingly, more records on burrow measurements are available for the species with the widest geographic range, the wide-ranging *Dasypus novemcinctus* (Zimmerman 1990, McDonough et al. 2000, Platt et al. 2004, Sawyer et al. 2012). In addition to *D. novemcinctus*, several additional species of armadillos were also monitored and had measurements of their burrows reported (Medri 2008, Attias et al. 2016, Desbiez et al. 2018). However, the relationship between the reported burrow measurements and species identification remains tenuous (e.g., Arteaga and Venticinque 2010). Few studies reported burrow measurements for different armadillo species (e.g., Carter and Encarnação 1983), and each reported measure can be different among the studies or is not always associated with their corresponding estimate of precision (variance or standard deviation), further complicating the identification process

(Carter and Encarnação 1983, McDonough et al. 2000, Medri 2008, Arteaga and Venticinque 2010, Trovati 2015).

Machine learning (ML) enables computers to solve tasks analyzing complex patterns without being explicitly programmed to solve those tasks (Sen 2018). State-of-the-art methods teach machines via supervised learning (i.e., by showing them correct pairs of inputs and outputs from labeled data), unsupervised learning (i.e., finding hidden information or structure from unlabeled data), and semi-supervised learning (i.e., a combination of supervised and unsupervised ML technique) (Sen 2018). Machine learning algorithms and models aim to maximize predictability based on data and have demonstrated high accuracy in predicting ecological patterns (Olden et al. 2008, Crisci et al. 2012, Thessen 2016). Machine learning models have been applied increasingly in ecology, including in studies of species distribution modeling (Elith et al. 2006, Phillips et al. 2006), species diversity (Olden et al. 2008), and distribution (Elith and Leathwick 2009), and represent a potential for improving species identification methods (Norouzzadeh et al. 2018).

Considering that *C. tatouay*, *C. unicinctus*, *D. novemcinctus*, and *E. sexcinctus* have overlapping ranges and inhabit regions affected by anthropogenic disturbances (Vivo et al. 2011, Egeskog et al. 2014, Trovati 2015), improving identification methods might help answer questions about habitat preferences and the role of anthropogenic threats for each species of armadillo, among several others. Our objective was to examine whether it was possible to identify similar sized armadillos based on their burrows using ML. Assuming that armadillo burrows have been correctly identified in published studies, we aimed to use simulating burrow measurements (width, height, and angle) for the 4 species of armadillos (*C. tatouay*, *C. unicinctus*, *D. novemcinctus*, and *E. sexcinctus*), based on data found in the literature, to train data and then classify it using supervised ML.

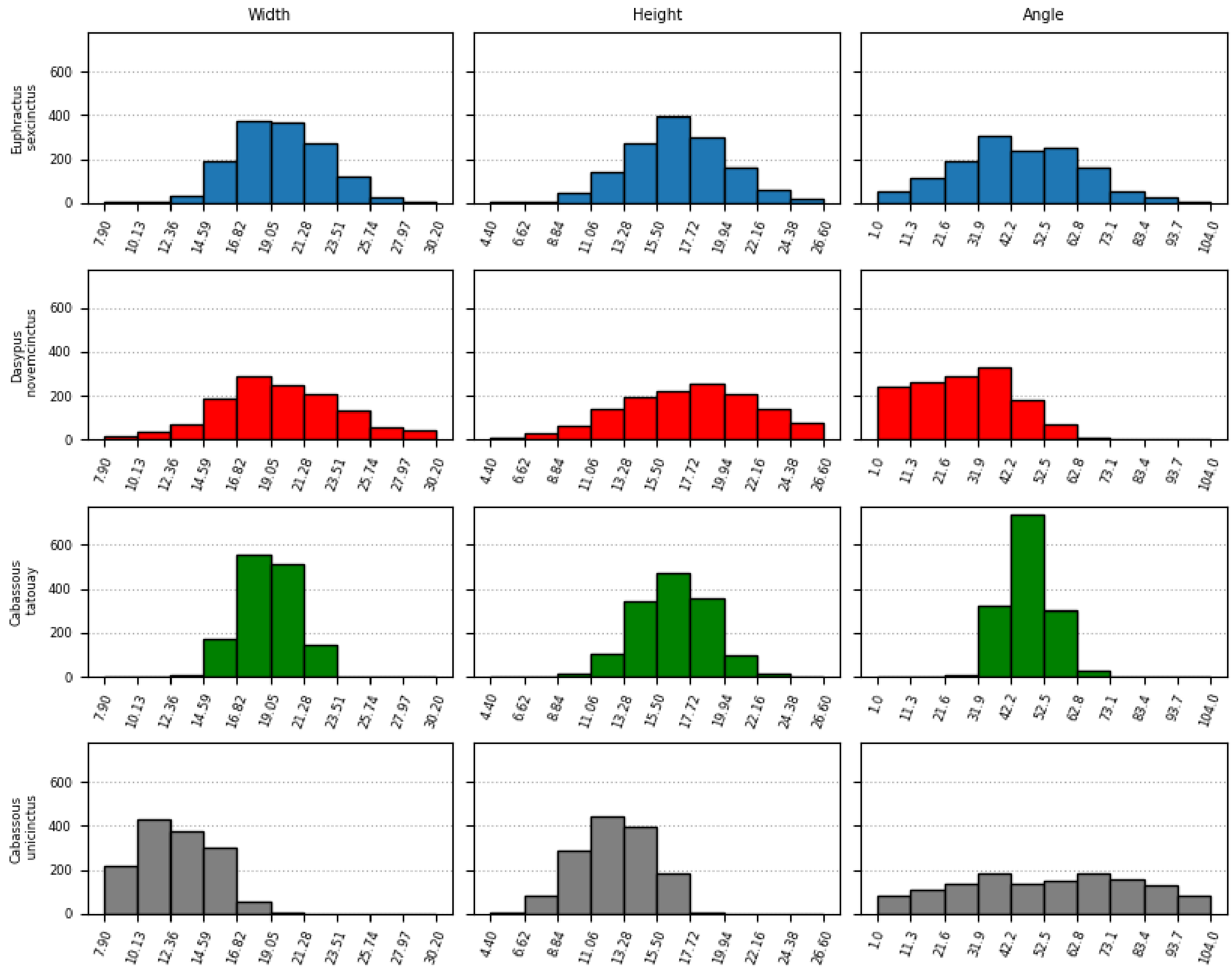
METHODS

Data Simulating Process

Considering that, to the best of our knowledge, there is no publicly available dataset for armadillo classification (i.e., there is no dataset that can be used to train a ML-based model), we decided to synthetically create one. This dataset was constructed based on 9 studies that reported morphometric measures of burrows for 4 species of armadillos (Table 1). From these works, we extracted and used the mean and standard deviation values to generate synthetic data for 3 measures: width, height, and angle. We choose these measurements because, they are the only measurements consistently present on the works related to morphometric measurements of armadillo burrows (Table 1), and they can provide relevant information that may assist in the discrimination of different burrows and the species of armadillo (Medri 2008, Sawyer et al. 2012, Attias et al. 2016, Desbiez et al. 2018).

Considering this lack of information, we created 2,000 synthetic samples (i.e., measurements of width, height, and angle) for each burrow class, based on all studies available for each species. For instance, we used 5 studies investigating *D. novemcinctus* burrows (Table 1). Therefore, in order to generate 2,000 samples for this class, we synthetically created 400 instances based on the (width, height, and angle) measurements from each one of the 5 studies. We repeated the same procedure for the other 3 species of armadillos, resulting in a final data set of 8,000 synthetic burrows. All this synthetic data were generated using the Normal family of distribution (rnorm function). For reported studies lacking corresponding standard deviation, we used the highest value found from studies that reported it for the same species (Table 1). Furthermore, for *C. tatouay* we used the Poisson family of distribution (rpois function) to generate data for angle since no standard deviation had been reported for this species thus far.

Although a considerable amount of data is created with this process, for each species we randomly selected and further exploited only 5,600 (70%) samples (Fig. 1), an important process to reduce bias of the data. The remaining 30% of samples was discarded from further analyses.



Machine Learning (ML) Model Development

The main goal of supervised ML algorithms is to build models capable of learning patterns from the data and then use this information to correctly classify unseen patterns. Correspondingly, our main goal was to create a ML model capable of learning burrows patterns (width, height, and angle) to correctly classify unseen burrow data (and consequently, armadillo species), using the synthetic data. It is important to highlight that this exploited data were scaled and centered, a common process performed in ML models (Mulaik 2009, Lantz 2019).

Since there is no single ML algorithm that fits all data (Tsai et al. 2009), we decided to assess 5 different ML-based techniques for the specified problem: Random Forest (rf; Breiman 2001), K-

Nearest Neighbors (knn; Hechenbichler et al. 2004), Support Vector Machine Radial (svmRadial; Scholkopf et al. 1997), Naïve Bayes Classification (nb; Rish 2001), and C5.0 Decision Tree (C5.0; Freund and Mason, 1999). Furthermore, since each one of those models has its own set of hyper-parameters (i.e., parameters that impact the learning process and, consequently, the outcome), we used the Grid-search method to search for the best set of hyper-parameters for each approach. This approach trains the same model several times varying the hyper-parameters (according to a pre-defined set of values) and then selects the model with the best performance for further analysis and investigations (Bergstra and Bengio 2012). All algorithms were implemented using the CARET package (Kuhn 2020) in the R program (R Core Team 2018).

To train and evaluate each method, we split the generated data into training (75%, i.e., 4,200 samples) and validation (25%, i.e., 1,400 instances) sets. The former is used to train the model, i.e., to make the model learn the patterns, whereas the latter is used to assess the model's performance. Observe that the model does not learn using the validation set, which is only used during the evaluation. By doing this, not only we avoid biasing the model, but we assess the model's performance in a scenario similar to the real world, i.e., a scenario in which unseen data (usually obtained in the fieldwork) is classified by the trained model. Aside from splitting the data, we also performed a 2-fold cross-validation repeated 3 times to minimize initial overfitting (Kohavi 1995). As the goal of ML algorithms is developing predictive models, we quantified the model performance using a confusion matrix and the Kappa index, highlighting accuracy, 95% CI, and other statistics by class (Table 2; Code File S1, available online in Supporting Information).

RESULTS

In general, all assessed methods had a very similar performance with approximately 70% of the overall accuracy and 61% of Kappa index accuracy (Table 2). As expected, the evaluated methods produced very similar results for all classes. Aide from this, it is interesting to observe that class *E*.

sexcinctus was the most difficult one to discriminate and classify (approximately 68% of accuracy), whereas class *C. unicinctus* is the easiest one to discriminate (approximately 93% of accuracy) (Table 3).

Overall, the results have shown that ML techniques are capable of identifying similar-body-sized armadillos based on their burrows using only 3 measurements (width, height, and angle), i.e., these measurements can be used to classify the burrows and, consequently, armadillo species.

DISCUSSION

We demonstrated that it is indeed possible to identify similar-body-sized armadillos based on their burrows. Supervised ML is an appropriate method able to deal with the complexity of the data by enabling the identification of armadillo burrows. Unlike traditional identification methods, ML models successfully found patterns and accurately matched them with the validation set. In such a way, it might be a useful tool that will help scientists to correctly identify armadillo burrows, providing a noninvasive, low-cost method to estimate population (i.e., relative abundance) or species (i.e., occupancy) parameters.

We found that the accuracy of ML models varied among our chosen species but overall was about 71%. Given this level of overall accuracy, the most appropriate use of the armadilloID function will be helping scientists and managers identifying the four species of armadillos considering altogether all labels from the 5 ML predictions. The use of complementary clues, when available, such as initial visual classification of the burrow shape (Trovati 2015) and presence of tracks or other local features (Sawyer et al. 2012, Desbiez et al. 2018) is therefore advisable. For instance, *E. sexcinctus* typically constructs burrows with an inverted U-shapedentrance, whereas the burrows of *C. unicinctus* have an almost perfectly round shape in an almost vertical angle (Carter and Encarnaçao 1983, Trovati 2015, Desbiez et al. 2018). Indeed, *C. unicinctus* showed higher accuracy of classification of their burrows. Moreover, *D. novemcinctus* is the only species that may have as many as

5 entrances into a single den (Carter and Encarnaçao 1983), and rotting leaves are often found near the entrances of burrows, especially after rains or floods (Talmage and Buchanan 1954). Therefore, using qualitative and local information, together with the morphometric measures, will surely guarantee higher levels of accuracy in correctly identifying armadillo species based on their burrows.

Typically, burrows are classified as either active or inactive during fieldwork (Sawyer et al. 2012). An active burrow has compacted forest litter, fresh excavations, or tracks at the entrance, whereas an inactive burrow typically has spider webs or debris in the entrance (Sawyer et al. 2012) and an eroded shape. Those differences are evident and easily identified in the field, although discriminating them does require some degree of experience by the observer. Although not tested, considering only measurements from active burrows may also increase the accuracy of armadillo identification from burrow measurements. We also highlighted the use of the 3 measures (width, height, and angle of the burrow's entrance) when collecting field data to identify armadillo species using the supervised ML methods. Apart from being easy to collect, they are the only measurements consistently present in the few studies reporting morphometric measurements of armadillo burrows. Burrow width and height provide more precise information than the diameter, as was also pointed out by Carter and Encarnação (1983).

Because of the armadillo's fossorial or semi-fossorial lifestyle (McBee and Baker 1982, Redford and Wetzel 1985, Hayssen 2014, Desbiez et al. 2018), searching for their burrows represents the most effective, low-cost sampling method for estimating population parameters. Such a sampling method might considerably increase our knowledge about armadillos, as some of them (e.g., *C.tatouay*, and *C.unicinctus*) spend most of their time underground (Hayssen 2014, Desbiez et al. 2018). We found that the use of novel technologies (machine learning) improved the usefulness of a noninvasive method, especially when dealing with low-density, elusive, and poorly known species such as the armadillos (Abba and Superina 2010, Desbiez et al. 2018).

Considering that the 4 armadillo species we examined inhabit a region affected by anthropogenic disturbances (Vivo et al. 2011, Egeskog et al. 2014, Trovati 2015), it is necessary to improve identification methods based on their burrows to consider the role of anthropogenic threats for each species. We still don't know much about anthropogenic effects on armadillos' populations (Abba and Superina 2010, Superina et al. 2014; Rodrigues et al. 2020). Correctly identifying armadillo burrows may increase our knowledge of species-specific habitat use, density, activity, behavior (McDonough et al. 2000, Abba et al. 2005, 2015, Arteaga and Venticinque 2008, Desbiez et al. 2018), and even ecosystem services such as bioturbation (Sawyer et al. 2012), and ecosystem-engineering (Desbiez and Kluyber 2013).

MANAGEMENT IMPLICATIONS

The deficit of information on armadillo populations in tropical ecosystems is partially due to the lack of cost-effective methodologies allowing managers to obtain data that will eventually lead to the development of appropriate management strategies. The supervised ML and the armadilloID function considered in this study indicate the potential that both have to identify similar-body-sized armadillos based on their burrows. The function presented here automates the burrow identification analysis for *E. sexcinctus*, *D. novemcinctus, C. tatouay*, and *C. unicinctus*, allowing scientists to use a noninvasive, low-cost method to study those armadillo species. Our method provides new insight towards preserving the old sampling methods (the mot cost-effective) while using new technologies such as ML to enable estimating population parameters of armadillo species. The function has the potentiality of expanding its options to embrace more armadillo species and more statistical models from other R packages. We encourage R programmers and ecologists to modify the code to satisfy their needs and expand the usage of armadilloID. Nevertheless, decision making about burrow classification should be made by the scientist itself, using the supervised ML together with as much qualitative field observation as possible, to more accurately identify species of armadillo based on burrow morphology.

ACKNOWLEDGMENTS

The authors would sincerely like to thank R. Applegate (Associate Editor), A. Knipps (Editorial Assistant), and 2 reviewers, Dr. Andras Zsom and one anonymous, for their invaluable comments that greatly improved the manuscript. The authors thank the Coordination for the Improvement of Higher Education (CAPES), the Brazilian Science Council (CNPq grant number 303101/2017-2), and the São Paulo Research Foundation (FAPESP grant number 2016/19106-1) for research fellowships and grants.

LITERATURE CITED

Abba, A. M., D. E. Udrizar-Sauthier, and S. F. Vizcaíno. 2005. Distribution and use of burrows and tunnels of *Chaetophractus villosus* (Mammalia, Xenarthra) in the eastern Argentinean pampas. Acta Theriologica 50:115–124.

Abba A. M., M. H. Cassini, and S. F. Vizcaíno. 2007. Effects of land use on the distribution of threes species of armadillos (Mammalia, Dasypodidae) in the pampas, Argentina. Journal of Mammalogy 88:502–507.

Abba, A. M., and M. Superina. 2010. The 2009/2010 Armadillo Red List Assessment. Edentata 11:135–184.

Abba, A. M., E. Zufiaurre, M. Codesido, and D. N. Bilenca. 2015. Burrowing activity by armadillos in agroecosystems of central Argentina: biogeography, land use, and rainfall effects. Agriculture, Ecosystems & Environment 200:54–61.

Agakov F., Bonilla, E., Cavazos, J., Franke, B., Fursin, G., O'Boyle, M. F. P., Thomson, J., Toussaint, M. and C. K. I. Williams. 2006. Using machine learning to focus iterative optimization.

International Symposium on Code Generation and Optimization, New York, NY, USA: 1–11.

Arteaga, M. C., and E. M. Venticinque. 2008. Influence of topography on the location and density of armadillo burrows (Dasypodidae: Xenarthra) in the central Amazon, Brazil. Mammalian Biology-Zeitschrift für Säugetierkunde 73:262–266.

Arteaga, M. C., and E. M. Venticinque. 2010. Cuevas de armadillos (Cingulata: Dasypodidae) en la Amazonía Central: son útiles para identificar especies? Edentata 11:29–34.

Attias, N., F. R. Miranda, L. M. M. Sena, W. M. Tomas, and G. M. Mourão. 2016. Yes, they can! Three-banded armadillos *Tolypeutes* sp. (Cingulata: Dasypodidae) dig their own burrows. Zoologia (Curitiba) 33: e20160035.

Bergstra, J., and Y. Bengio. 2012. Random search for hyper-parameter optimization. The Journal of Machine Learning Research 13.1:281–305.

Breiman, L. 2001. Random forests. Machine learning 45: 5–32.

Carter, T. S., and C. D. Encarnação. 1983. Characteristics and use of burrows by four species of armadillos in Brazil. Journal of Mammalogy 64:103–108.

Crisci, C., B. Ghattas, and G. Perera. 2012. A review of supervised machine learning algorithms and their applications to ecological data. Ecological Modelling 240:113–122.

Desbiez, A. L. J., and D. Kluyber. 2013. The role of giant armadillos (*Priodontes maximus*) as physical ecosystem engineers. Biotropica 45:537–540.

Desbiez, A. L. J., G. F. Massocato, D. Kluyber, and R. C. F. Santos. 2018. Unraveling the cryptic life of the southern naked-tailed armadillo, *Cabassous unicinctus squamicaudis* (Lund, 1845), in a Neotropical wetland: Home range, activity pattern, burrow use and reproductive behaviour.

Mammalian Biology 91:95–103.

Egeskog, A., F. Freitas, G. Berndes, G. Sparovek, and S. Wirsenius. 2014. Greenhouse gas balances and land use changes associated with the planned expansion (to 2020) of the sugarcane ethanol industry in Sao Paulo, Brazil. Biomass and Bioenergy 63:280–290.

Eisenberg, J. F., and K. H. Redford. 1999. Mammals of the Neotropics. The University of Chicago Press, Chicago.

Elith, J., C. H. Graham, R. P. Anderson, M. Dudík, S. Ferrier, A. Guisan, R. J. Hijmans, F. Huettmann, J. R. Leathwick, A. Lehmann et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129–151.

Elith, J., and J. R. Leathwick. 2009. Species distribution models: ecological explanation and prediction across space and time. Annual Review of Ecology, Evolution, and Systematics 40.

Freund, Y. and L. Mason. 1999. The alternating decision tree learning algorithm. Icml 99: 124–133.

Hayssen, V. 2014. *Cabassous tatouay* (cingulata: Dasypodidae). Mammalian Species 46:28–32.

Hechenbichler, K. and K. Schliep. 2004. Weighted k-nearest-neighbor techniques and ordinal classification. Sonderforschungsbereich 386: 1–17.

International Union for Conservation of Nature [IUCN]. 2019. The IUCN Red List of Threatened Species. Version 2019-1. <http://www.iucnredlist.org.>. Accessed 28 Feb 2020.

Kohavi, R. 1995. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence 14: 1137–1145.

Kuhn, M. 2020. caret: Classification and Regression Training. R package version 6.0-86.

https://CRAN.R-project.org/package=caret

Loughry, W. J., and C. M. McDonough. 2013. Beyond natural history: some thoughts about research priorities in the study of xenarthrans. Edentata 14:9–15.

Maccarini, T. B., N. Attias, Í. M. Medri, J. Marinho-Filho, and G. Mourão. 2015. Temperature influences the activity patterns of armadillo species in a large neotropical wetland. Mammal Research 60:403–409.

McBee, K., and R. J. Baker. 1982. *Dasypus novemcinctus*. Mammalian Species 162:1–9.

McDonough, C. M., M. A. Delaney, P. Q. Le, M. S. Blackmore, and W. J. Loughry. 2000. Burrow characteristics and habitat associations of armadillos in Brazil and the United States of America. 48:109–120.

McNab, B. K. 1980. Energetics and the limits to a temperate distribution in armadillos. Journal of Mammalogy 61:606–627.

Medri, Í. M. 2008. Ecologia e história natural do tatu-peba, *Euphractus sexcinctus* (Linnaeus, 1758), no pantanal da Nhecolândia, Mato Grosso do Sul. Dissertation, Universidade de Brasilia, Brasília, Brazil.

Mulaik, S. A. 2009. Foundations of factor analysis. CRC press, Boca Raton, USA.

Norouzzadeh, M. S., A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune. 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proceedings of the National Academy of Sciences 115:5716–5725.

Olden, J. D., J. J. Lawler, and N. L. Poff. 2008. Machine learning methods without tears: a primer for ecologists. The Quarterly Review of Biology 83:171–193.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231–259.

Platt, S. G., T. R. Rainwater, and S. W. Brewer. 2004. Aspects of the burrowing ecology of nine-banded armadillos in northern Belize. Mammalian Biology 69:217–224.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.r-project.org/>.

Redford, K. H., and R. M. Wetzel. 1985. *Euphractus sexcinctus*. Mammalian Species 252:1–4.

Rish, I. 2001. An empirical study of the naive Bayes classifier. International Joint Conference on Artificial Intelligence 3: 41–46.

Rodrigues, T. F., and A. G. Chiarello. 2018. Native forests within and outside protected areas are key for nine-banded armadillo (*Dasypus novemcinctus*) occupancy in agricultural landscapes. Agriculture, Ecosystems & Environment 266:133–141.

Rodrigues, T. F., A. M. B. Mantellatto, M. Superina, and A. G. Chiarello. 2020. Ecosystem services provided by armadillos. Biological Reviews 95:1–21.

Sawyer, C. F., D. C. Brinkman, V. D. Walker, T. D. Covington, and E. A. Stienstraw. 2012. The zoogeomorphic characteristics of burrows and burrowing by nine-banded armadillos (*Dasypus novemcinctus*). Geomorphology 157–158:122–130.

Schipper, J., J. S. Chanson, F. Chiozza, N. A. Cox, M. Hoffmann, V. Katariya, J. Lamoreux, A. S. L. Rodrigues, S. N. Stuart, H. J. Temple et al. 2008. The status of the world's land and marine mammals: diversity, threat, and knowledge. Science 322:225–230.

Scholkopf, B., Sung, K. K., Burges, C. J., Girosi, F., Niyogi, P., Poggio, T. and V. Vapnik. 1997.

Comparing support vector machines with Gaussian kernels to radial basis function classifiers. IEEE Transactions on Signal Processing 45: 2758–2765.

Sen, S. 2018. Machine learning and IoT: a biological perspective. CRC Press, Boca Raton, USA.

Superina, M., N. Pagnutti, and A. M. Abba. 2014. What do we know about armadillos? An analysis of four centuries of knowledge about a group of South American mammals, with emphasis on their conservation. Mammal Review 44:69–80.

Talmage, R. V, and G. D. Buchanan. 1954. The armadillo (*Dasypus novemcinctus*): a review of its natural history, ecology, anatomy and reproductive physiology. The Rice Institute Monograph in Biology 41:1–135.

Thessen, A. 2016. Adoption of machine learning techniques in ecology and earth science. One Ecosystem 1:e8621.

Trovati, R. G. 2015. Differentiation and characterization of burrows of two species of armadillos in the Brazilian Cerrado. Revista Chilena de Historia Natural 88:19.

Tsai, C. F., Y. F. Hsu, C. Y. Lin, and W. Y. Lin. 2009. Intrusion detection by machine learning: A review. Expert Systems with Applications 36:11994–12000.

Xu, Y. and R. Goodacre. 2018. On splitting training and validation set: A comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning. Journal of Analysis and Testing 2.3: 249–262.

Zimbres, B., M. M. Furtado, A. T. A., L. Silveira, R. Sollmann, R. B. Machado, and J. Marinho-filho. 2013. The impact of habitat fragmentation on the ecology of xenarthrans (Mammalia) in the Brazilian Cerrado. Landscape Ecology 28: 259–269.

Zimmerman, J. W. 1990. Burrow characteristics of the nine-banded armadillo, *Dasypus novemcinctus*. The Southwestern Naturalist 35:226–227.

Table 1. Species, measurements, and references used to generate the data of burrows. N = number of studies.

Species	N	Measurement	Reference[a]
C. tatouay	2	width, height, angle	Carter and Encarnação (1983), Anacleto (2006)[a]
C. unicinctus	4	width, height, angle	Carter and Encarnação (1983), Anacleto (2006)[a], Trovati (2009), Desbiez et al. (2018)
D. novemcinctus	5	width, height, angle	Zimmerman (1990), McDonough et al. (2000)[a] – EUA, Platt et al. (2004), Anacleto (2006)[a], Sawyer et al. (2012)
E. sexcinctus	4	width, height, angle	Carter and Encarnação (1983), Anacleto, (2006)[a], Medri (2008)[a], Trovati (2009)

[a]angle not reported

Table 2. Performance of the machine-learning (ML) algorithms over the validation set.

ML algorithms	Accuracy	95% CI	Kappa
Random Forest	0.69	0.67 to 0.72	0.59
k-Nearest Neighbors	0.71	0.68 to 0.73	0.61
Support Vector Machine	0.72	0.70 to 0.74	0.63
Naïve Bayes	0.69	0.67 to 0.72	0.59
C5.0 Decision Tree	0.72	0.69 to 0.74	0.62

Table 3. Machine-learning (ML) algorithms accuracy over the validation set by the species of armadillos.

ML algorithms	*E. sexcinctus*	*D. novemcinctus*	*C. tatouay*	*C. unicinctus*
Random Forest	0.68	0.75	0.84	0.92
k-Nearest Neighbors	0.67	0.76	0.86	0.93
Support Vector Machine	0.68	0.77	0.88	0.93
Naïve Bayes	0.65	0.76	0.87	0.90
C5.0 Decision Tree	0.68	0.78	0.86	0.93

Summary for online Table of Contents:

1. Our findings advance on the use of novel technologies (machine learning) enabling the use of a non-invasive method (burrows) for dealing with low-density, elusive, and not well-known species such as the armadillos.

2. A non-invasive method for estimating population parameters of armadillo species will surely guarantee higher efforts towards armadillo management and conservation.