# Potential and Limitations of Designing a Deep Learning Model for Discovering New Archaeological Sites: A Case with the Mesopotamian Floodplain

Marco Roccetti, Luca Casini, Giovanni Delnevo, Valentina Orrù, Nicolò Marchetti

Alma Mater Studiourum - University of Bologna
{marco.roccetti, luca.casini7, Giovanni.delnevo2, valentina.orru5, nicolo.marchetti}@unibo.it

## ABSTRACT

We have tried to provide an answer to the question whether a collection of satellite images, with notable archaeological sites, is informative enough to instruct a deep learning model that discovers new archaeological sites, well before archaeologists venture out in the field. Convolutional neural networks and satellite images in the visible light range were employed to detect sites in the Iraqi region of ***Qadisyah***. The preliminary results we achieved are interesting yet not still fully convincing. The AUC value we got is near 70%, while more interesting findings have come from the idea to map the numerical predictions into *heat-maps*, revealing the regions where a site can lie. Several motivations can explain this controversial output. Not least is the fact that our model was instructed to learn archaeological sites of a very different form and size.

## KEYWORDS

Smart data · Deep learning · Archaeology · Qadisiyah

**ACM Reference format:**

M. Roccetti et al. 2020. Potential and Limitations of Designing a Deep Learning Model to Discover New Archaeological Sites: A Case with the Mesopotamian Foodplain. In *Proceedings of GOODTECHS'20. ACM, Antwerp, Belgium, 6 pages*


*GoodTechs '20, September 14–16, 2020, Antwerp, Belgium*

*ACM ISBN 978-1-4503-7559-7/20/09…$15.00*
*https://doi.org/10.1145/3411170.3411254*

## 1 Introduction

Deep Learning is one of the most exciting technological innovations of the 21st century, changing the landscape when it comes to things machines can or cannot do. If there is a field that has reaped the fruits of deep neural networks, it sure is computer vision [1]. Convolutional Neural Networks (CNNs) have been behind most, if not all, the leaps forward in all applications related to vision. Vision, in all its *flavors*, like classification or detection, is notoriously a task as difficult for computers as it is easy for humans [2]. The biggest problem is that there is no easy way to specify what a machine has to look for, and also the search space quickly becomes enormous as the image size grows. Neural networks were able to overcome this limit and in the last few years it turned out that a computer can even see better than a human in many cases [3, 4]. In other situations, neural networks still can help greatly by compensating for human error or by speeding up a time-consuming procedure [5, 6]. This article will describe one of such scenarios and will discuss about the limitations that can affect this technology under specific circumstances. The situation is as follows. Since a long time, archaeologists around the world have been using remote sensing technology to investigate possible points of interest, before venturing out in the field. This activity is incredibly time consuming given the size of the areas they usually have to look at, and it also leaves significant room for human error. One idea is that of using a neural network to assist archaeologists, by highlighting possible sites of interest on a map, without moving from their desk at home. Such an idea has been developed in the context of a research project at the University of Bologna in the Iraqi region of *Qadisiyah*. The first results, presented here, are not yet good enough to consider the problem closed, nonetheless they open up interesting research directions and stimulate reflections. While, on one side, we face the challenging task of automating a new image-recognition process, on the other one, we should never forget that the attempt to automate a task that involves images of this specific kind leaves open the problem of the interpretation of the obtained results, whose meaning is not simply limited to the sphere of pure perception, as it is bound to the comprehension of a specific context that only an expert archaeologist can have [7]. This paper is structured as follows: in Section 2 we describe the technical problem we have tried to solve. Section 3 discusses the design choices we made to assemble both the dataset and the deep learning model we used. Section 4 presents the results we obtained, while Section 5 and Section 6, respectively, illustrate how heatmaps and contextual information can be put to good use to ameliorating the final output. Section 7 concludes the paper.

## 2 Archaeology and remote sensing

By remote sensing, we refer to all the techniques and technology that involves using satellite or aerial imagery to gather insight into or monitor some points of interest. This can range from keeping track of forests' growth or of the movement and size of glaciers, to even identifying pools in a neighbourhood, or looking for archaeological sites, like in our case [8-11]. It must be noted how remote sensing is not limited to the visible light spectrum, as satellites can have a broad range of different sensors on-board, that may highlight different properties of the target, like temperature or chemical composition, for example. The data can even go beyond 2-dimensional imagery: for instance, *lidar* is becoming more and more popular, and is able to measure distances with very high precision (in the order of millimetres), giving the possibility of creating extremely detailed digital terrain models, once they are mounted on drones.

When it comes to the field of archaeology, using remote sensing is very important as it allows one to scour the area s/he wants to investigate in order to pinpoint the most interesting sites to visit, before wasting resources and time directly on that location [12], and also may offer the opportunity of finding new, unidentified sites, that are only evident when observed from above. Automatizing this procedure is a kind of the holy grail of *digital archaeology*, and has been tried in the past with numerous other computer vision techniques (mostly based on geometrical pattern matching), yet with controversial results [13]. On the other side, deep learning offers a great potential, but it has its own complications, mainly due to the limited size of the images it can accept as its input; hence rendering the correspondent datasets quite *unwieldly*. Other crucial factors are bound to the nature of the searched objects: simple and consistent shapes, like burial mounds, can be easily identified, while varying shapes are much more difficult to be recognized. Also, the availability of elevation data, and 3D maps, can give an edge in spotting particular structures, that a simple 2D photo could not [14, 15]. Similarly, different channels in the image, beside visible light, could also be of great help.

## 3 A Deep Learning model for the *Qadis* Project

We have tried to automatize the remote sensing phase for the QADIS project of the University of Bologna. Archaeologists are working in the *Qadisiyah* region in Iraq, where an area of 1830 km$^2$ is being surveyed mainly, in search for a characteristic type of archaeological site, called Tell [16].

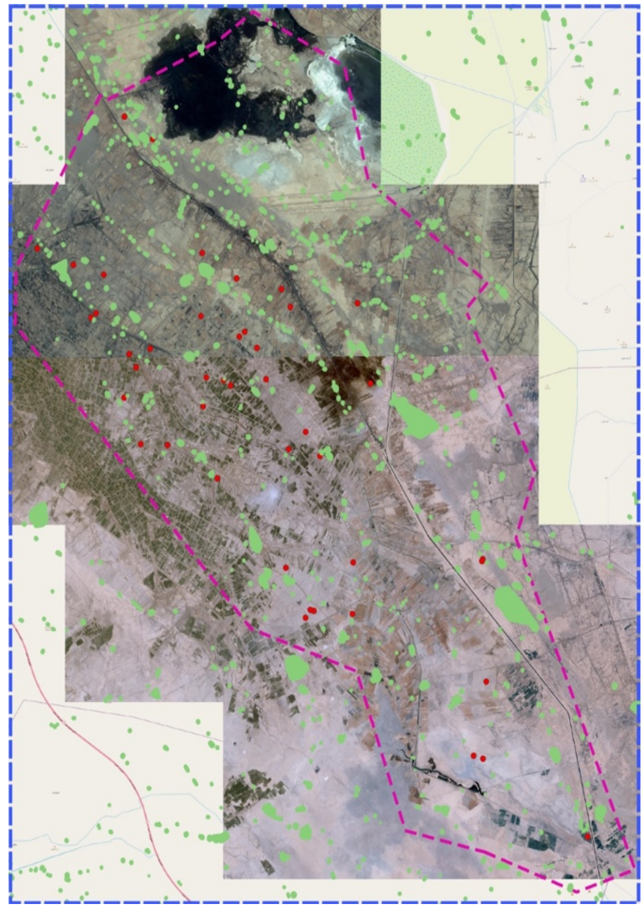

**Figure 1: The research area in central Mesopotamia. The purple dashed line circumscribes the QADIS survey area. The blue dotted line the extent considered in this paper. Surveyed archaeological sites are in green, while ground-truthed false sites from remote sensing in red.**

Those kind of sites can be found all across the Near East, and it is worth mentioning that other automatic techniques for detection have already been tried, albeit without engaging deep learning [17]. Our final goal was to speed up the remote sensing phase, by having an AI-based model providing reliable suggestions to an expert archaeologist. In particular, we considered a homogeneous geomorphological landscape, that of the central Mesopotamian floodplain, in order to present the model with a more consistent set of variables (which remain high). In particular, as shown in Figure 1, 166 polygons were drawn by the archaeologists, representing the extent of the archaeological region. Of these, 145 were confirmed upon visit, and 21 were found not to be actual archaeological sites. In addition, further 415 *archaeological* sites were previously surveyed by Adams (1091) [18]; yet they were not utilized in our study. Starting from the satellite photos and the shapes we were provided with, we first defined a rectangle to inscribe the Qadis area. This rectangle was then divided into **tiles**, corresponding to images of 299×299 pixels, preserving their native resolution and roughly corresponding to an area of almost 180x180 meters. Not to lose any information, we also used a kind of intermediate tile, between two consecutive tiles, essentially obtained shifting the window by half a tile, when we extracted the images. All this resulted into a set of approximately 300k tiles. To be considered is the fact that while the number of tiles in this dataset is huge, it is also extremely skewed in favor of tiles representing non-sites: in fact, the number of tiles that could represent true sites is just 3280. To be noticed, again, is the fact that these 3280 refer to tiles that can also have *almost a null* intersection with a true site (that is, they overlap just for a very small portion with a true site); hence making the validity of the information contained in that tile arguably significant.

As shown in Figure 2, to alleviate this problem, we selected tiles as representative of true sites, just in these two cases: i) they overlapped with true sites for a geographical extension of at least 10% of their area, ii) they overlapped with true sites for an extent of at least 30%. In the first case, we got some 2211 tiles, while in the second case we got some 1597 tiles; out of the total amount of 3280 tiles. These tiles represented the positive examples on the basis of which we constructed our datasets for training our deep learning model. To the positive cases, we added some negative cases (non-sites) to get two well-balanced datasets, respectively constructed of 4422 and 3194 examples. Those were randomly selected from tiles that had no intersection with the sites as suggested by the archaeologists.

Nonetheless, be it 4422 or 3168, a few thousand tiles (images) are not enough to train a classifier from scratch, and so we considered the technique termed **Transfer Learning**, as a viable strategy to adopt in our context. The basic idea was to use, as a basis for our training activity, a deep learning model that has already learnt how to classify images, and then to specialize it to deal with our tiles [20, 21].

To this aim, we exploited the *Google's Inception V3* model, made available for use as a pre-trained tool in the *Keras* library [22]. We used *Inception* as a feature extractor, with its weights (learned from the *ImageNet* dataset) frozen, by removing the classifying head and then replacing it with the new neural network to be trained. The model was trained for 10 epochs, as increasing the number of epochs yielded no significant difference when we came to the results. Not only, also class weights were employed (0.3 for non-site and 0.7 for sites, respectively), in an effort to help the model avoid mistakes on the positive cases.

Finally, in the hope of helping the model learn a more general representation, as well as to avoid overfitting, we also resorted to a data augmentation procedure that randomly applied a set of general geometrical transformations (like flipping, zoom and shear, for example) to all the images on a given batch.

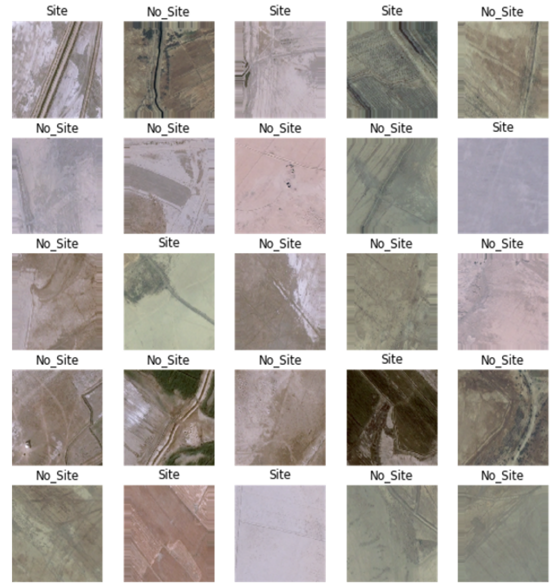


**Figure 2: An excerpt from the dataset. Tiles labeled as Site overlap with the dig sites shape provided by the archaeologists.**

Specifically, we resorted to the data augmentation procedure provided by the *Keras* platform, and termed *ImageDataLoader*.

## 4 Preliminary Results

Following the aforementioned procedures, we trained four models, two with a perfect balance between the quantity of sites and non-sites, and two with a ratio 1:2 of the number of sites against non-sites, respectively:

- Model/Experiment 1 - balanced dataset (1:1 site/non-site ratio) and the 10% overlapping rule;
- Model/Experiment 2 - unbalanced dataset (1:2 ratio) and the 10% overlapping rule;
- Model/Experiment 3 - balanced dataset (1:1 ratio) and the 30% overlapping rule;
- Model/Experiment 4 - unbalanced dataset (1:2 ratio) and the 30% overlapping rule.

The AUC accuracy prediction scores for the hold-out test set (20% of the total) are shown in Figure 3, and are all in the neighborhood of 60%; Experiment 2 achieves the lowest at 55%, while Experiment 3 is the best one, with a 63%. These poor performances were somewhat expected and further confirms the difficulty of this task.

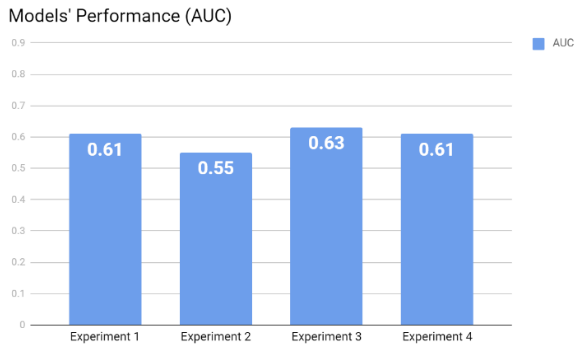


**Figure 3: Experimental Results on the 20% hold-out set.**

## 5 Using prediction heatmaps

Even if the general AUC metrics were not well promising, we decided to take the output from Model 3 and to overlay them to the geographical site map in order to print a pictorial impression of how the predictions came distributed over a real map. In spite of the numerical results, in fact, our intent was to check if (at least visually) the predictions returned by the model could still point the user in the right direction, by highlighting a particular spot on the map.

The following pictures (Figures 4, 5, 6, 7) show an example area from the map and were produced by overlaying the predictions returned by Model 3 on the QADIS area map, using the software QGIS. In particular, Figure 4 represents our starting point. The tiles in yellow should be predicted as sites, while those in blue as non-sites. Figure 5 should resemble a classical *heatmap*, where the more yellowish are the tiles, the higher is the associated probability they should represent (a part of) a true site, as predicted by the model (low probability zones become transparent because of the overlay mechanism).

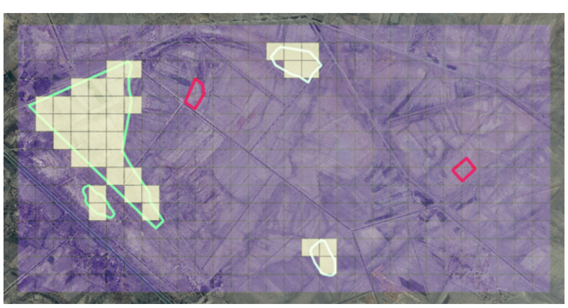

**Figure 4: Ground Truth example for 30% overlapping rule. Blue tiles are no-sites while yellow are sites.**

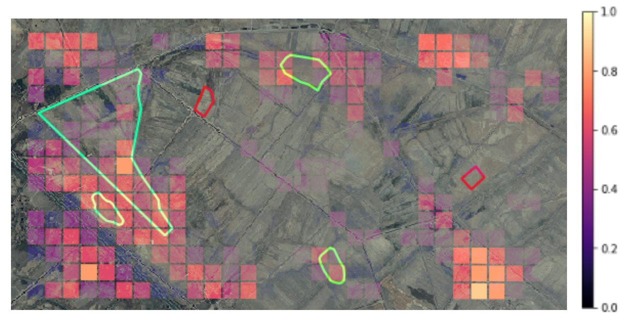


**Figure 5: Predictions example overlaid to the *Qadis* satellite map. High probability tiles tend to yellow, low probability tiles tend to black. The overlay technique in QGIS makes the latter transparent.**

While this result is still not convincing at all, we tried to follow a different line of reasoning for the interpretations of the predictions our model returned. The new idea we developed was to extend the prediction map to incorporate blocks of 5×5 tiles, over which the prediction probability was computed as equal to the maximum value of the contained tiles.

With this new approach (Figure 6), the big archaeological site on the left is discovered, while the small sites on the right are still off our radar, and also two false positive areas, unfortunately, emerge. To notice is the fact that we adopted a very simple *max* function to reason with our heatmap. Less naïve functions and filters could hopefully provide more accurate interpretations.

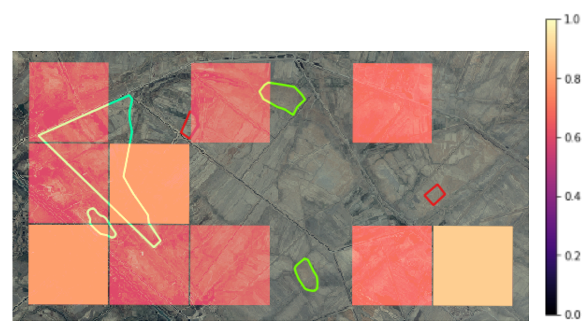


**Figure 6: The same prediction map of Fig.5 reduced to blocks of 5×5 tiles.**

Nonetheless, to check the robustness of our method, we tried to shift to the left all the scene, in order to verify if the predictions stay unchanged. Results are shown in Figure 7 and are controversial. On one side, it is confirmed that the big archaeological site is somewhat individuated (and the two little ones on the right stay off the radar), nonetheless more areas representing false positives emerge on the left of the scene, thus confirming that the approach is not still stable.

## 6 The role of the context and data augmentation

When a deep learning based model fails to provide accurate predictions, a classical method to recover from that failure is that of extending it to be trained to learn the surrounding context [25].

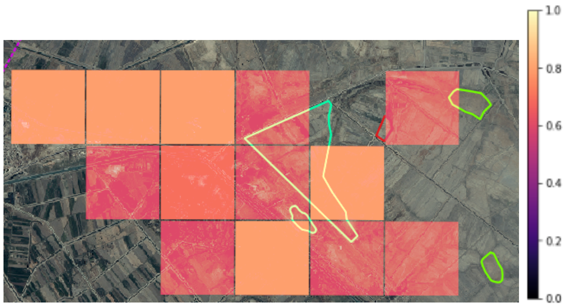


**Figure 7: Shifting the prediction map of Fig.6 to the right.**

With this in mind, we tried to extend our deep learning model with a secondary input to codify the larger area surrounding the tile we wanted to be learnt. The neural scheme is sketched in Figure 8 below, and the idea is that this additional image should somehow help encoding the contextual information around that tile in the hope this could clarify if it is a site or not.

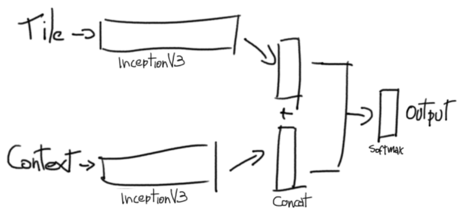


**Figure 8: A neural model with a second branch for contextual information.**

Summarizing, for each training tile we assembled a new image, composed of the 8 surrounding tiles around it, thus resulting in a 3x3 squared picture. This picture was then resized to a 299x299 tile, in order to feed it to our *Inceptionv3* model (the same used to analyze the primary tile). As seen from Figure 3, we deployed a new custom model composed of two parallel *Inceptionv3* branches, one for the tile, one for its 3x3 context, whose outputs were concatenated and fed to a dense layer for final prediction. The model was trained using the same hyper-parameters of the Model/Experiment 3. Summing up: 30% overlapping; 1:1 ratio, 10 epochs; .3 / .7 class weights. After training, the testing activity was performed again, and a **65%** AUC-ROC value was achieved, with just a moderate improvement in the prediction performances. Further to the idea of exploiting the context, inspired by similar research initiatives [26], a final improvement to our neural model was achieved by carrying out a more *intense* activity of *data augmentation*. Simply told, having a large dataset is crucial for any and each deep learning model as it is possible to improve its performance by augmenting the data we have at hand. To this aim, deep learning frameworks usually have built-in automatic data augmentation utilities; yet the use of those augmentation packages should be handled with care, since one of the risks is the transformation of the original bunch of data into an inconsistent dataset. Following this idea, at the end of our experiments, we also tried to exploit the data augmentation procedures made available in our development platform, yielding eight different tiles, after transformations, from each tile under consideration. This resulted into a quantity of tiles, representing positive examples (i.e., true sites), approximately equal to **12776**. At that point, to work with a balanced training dataset (where the number of positive examples equals the quantity of negative ones), we further added 12776 tiles representing non-sites, and then we retrained our model. The result we got, in terms of AUC_ROC on each single tile, was around the value of **71%**. While one could acknowledge this as an important improvement, again we point to the fact that an increased performance in the classification of single tiles, does not correspond automatically to a better performance in recognizing true archaeological sites in their entirety, owing to the motivation that our tiles are, on the map, just small portions of larger archaeological sites. In that, being our case very different (and more complex, as well) from that analyzed in [26], where instead a picture of an entire Vikings' grave squarely fits into a single tile.

## 7 Conclusions

We devised a deep learning model able to automatically spot probable archaeological sites, using satellite imagery from Mesopotamia, without the need that archaeologists venture out in the field. We divided the investigated area into relative small tiles (239x239 pixels), considering as positive examples those that overlapped already known sites, whose shapes the archaeologist had put in evidence. Using a pre-trained model from the *Keras* library as a feature extractor, we designed a new neural model with the intention to classify those tiles. The performances of the models, measured with the AUC-ROC metric, were approximately in the range 60-65%, with the best model topping at 70%. We used one of these models to produce a prediction heatmap to be overlaid to the actual satellite imagery in a GIS software. This way, the numerical predictions has given a general idea on where the sites could be, yet a considerable number of false positives emerge that could hobble the potential usefulness of our model. Hence, we are left with two main possible research roads, open in front of us to ameliorate the results: i) enhancing the present model, ii) changing the strategy. Following i), the present model could be enhanced with more sophisticated neural architectures and pre- and post- processing techniques that could lead to: a higher accuracy in the predictions in terms of an improved precision for the site class, as well as a reduced quantity of false positives. One idea could be using two parallel and separate networks, the former that looks at the tile and the latter that looks at all its neighbors (ideally a 3x3 square), in order to grasp the context of that tile better than we did; the two branches would then be concatenated to give the final prediction, at the end.

Another idea could be to use an encoder-decoder architecture to learn how to reconstruct a great number of tiles, without caring about the label, and then use the learned latent representation to train a classifier without the need for transfer learning. This approach may work better as the convolutional network would learn only from images related to the domain of archaeology, instead of using *Inception*. The second road is that of following a human-in-the-loop approach to create a system that gracefully leverages the capabilities of convolutional networks and the knowledge of human experts. These experts would be involved at all stages: by helping both in the selection of the most appropriate exemplar tiles from the dataset and in finding suitable ways to augment the data or to obtain more (e.g., using the new *FloodPlains Project* that refers to an area which is 40 times as large as Qadis - http://www.orientgis.net/); lastly, by reconsidering the prediction heatmaps, and suggesting how to make them more useful. Nonetheless, we can conclude that our study has achieved preliminary results that are still not fully convincing. The AUC-ROC did not surpass 70%, while more interesting findings came from the idea to map the numerical predictions into heat-maps. Among the motivations is the fact that our model was instructed to learn archaeological sites of very different forms. Our belief is that this lack of *uniformity* of what to be learnt is the cause of this partial failure [26, 27, 28].

## ACKNOWLEDGEMENTS

Thanks to Francesco Lamura for contributing to the experiments.

## REFERENCES

[1] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. 2018. Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience* 2018:1–13.

[2] Robert Geirhos, David H. J. Janssen, Heiko H. Schütt, Jonas Rauber, Matthias Bethge, and Felix A. Wichmann. 2018. Comparing deep neural networks against humans: object recognition when the signal gets weaker. *arXiv:1706.06969 [cs, q-bio, stat]* (December 2018).

[3] Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542, 7639:115–118.

[4] Scott Mayer McKinney, Marcin Sieniek, Varun Godbole, Jonathan Godwin, Natasha Antropova, Hutan Ashrafian, Trevor Back, Mary Chesus, Greg C. Corrado, Ara Darzi, Mozziyar Etemadi, Florencia Garcia-Vicente, Fiona J. Gilbert, Mark Halling-Brown, Demis Hassabis, Sunny Jansen, Alan Karthikesalingam, Christopher J. Kelly, Dominic King, Joseph R. Ledsam, David Melnick, Hormuz Mostofi, Lily Peng, Joshua Jay Reicher, Bernardino Romera-Paredes, Richard Sidebottom, Mustafa Suleyman, Daniel Tse, Kenneth C. Young, Jeffrey De Fauw, and Shravya Shetty. 2020. International evaluation of an AI system for breast cancer screening. *Nature* 577, 7788:89–94.

[5] Marco Roccetti, Giovanni Delnevo, Luca Casini, and Giuseppe Cappiello. 2019. Is bigger always better? A controversial journey to the center of machine learning design, with uses and misuses of big data for predicting water meter failures. *J Big Data* 6, 1, 70.

[6] Gualandi M. L., Scopigno R., Wolf L., Richards J., Garrigos J. B. I., Heinzelmann M., Hervas M. A., Vila L. and Zallocco M. 2016. ArchAIDE-Archaeological Automatic Interpretation and Documentation of Ceramics. *EUROGRAPHICS Workshop on Graphics and Cultural Heritage*. 203–206.

[7] A. Traviglia, D. Cowley, K. Lambers. 2016. Finding common ground: human and computer vision in archaeological prospection. *Newsletter of the Aerial Archaeology Research Group*. 53:11-24.

[8] Masoud Mahdianpari, Bahram Salehi, Mohammad Rezaee, Fariba Mohammadimanesh, and Yun Zhang. 2018. Very Deep Convolutional Neural Networks for Complex Land Cover Mapping Using Multispectral Remote Sensing Imagery. *Remote Sensing* 10, 7, 1119.

[9] Jean-Daniel Sylvain, Guillaume Drolet and Nicolas Brown. 2019. Mapping dead forest cover using a deep convolutional neural network and digital aerial photography. *J. of Photogrammetry and Remote Sensing*.156:14-26

[10] John C. Bliss, Thomas M. Bonnicksen, and Thomas H. Mace. 1980. Computer-aided classification of forest cover types from small scale aerial photography. *Environmental Management* 4, 6:499–510.

[11] Zsolt Domozi and Andras Molnar. 2019. Surveying private pools in suburban areas with neural network based on drone photos. *IEEE EUROCON 2019 -18th International Conference on Smart Technologies*, 1–6.

[12] Carrie Hritz. 2010. Tracing Settlement Patterns and Channel Systems in Southern Mesopotamia Using Remote Sensing. *Journal of Field Archaeology* 35, 2:184–203.

[13] Sam Redfern. A PC-Based System for Computer Assisted Archaeological Interpretation of Aerial Photographs. 1997. *Archaeology in the Age of the Internet. CAA97. Computer Applications and Quantitative Methods in Archaeology. Proceedings of the 25th Anniversary Conference*.

[14] Jean-Pierre Toumazet, Franck Vautier, Erwan Roussel, and Bertrand Dousteyssier. 2017. Automatic detection of complex archaeological grazing structures using airborne laser scanning data. *Journal of Archaeological Science: Reports* 12:569–579.

[15] Enrique Cerrillo-Cuenca. 2017. An approach to the automatic surveying of prehistoric barrows through LiDAR. *Quaternary International* 435: 135–145.

[16] Nicolò Marchetti, Abbas Al-Hussainy, Giacomo Benati, Giampaolo Luglio, Giulia Scazzosi, Marco Valeri, and Federico Zaina. 2019. The Rise of Urbanized Landscapes in Mesopotamia: The QADIS Integrated Survey Results and the Interpretation of Multi-Layered Historical Landscapes. *Zeitschrift für Assyriologie und vorderasiatische Archäologie* 109, 2:214–237.

[17] Bjoern H Menze, Simone Muhl, and Andrew G Sherratt. 2007. Virtual Survey on North Mesopotamian Tell Sites by Means of Satellite Remote Sensing. In *Broadening Horizons Multidisciplinary Approaches to Landscape Study. Cambridge Scholars Pub*. 17.

[18] Robert McC Adams. 1981. *Heartland of cities: surveys of ancient settlement and land use on the central floodplain of the Euphrates*. University of Chicago Press, Chicago.

[19] Luca Casini, Marco Roccetti, Giovanni Delnevo, Nicolò Marchetti, and Valentina Orrù. 2020. The Barrier of Meaning in Archaeological Data Science. *SCI-FI 2020: The 2020 International Science Fiction Prototyping Conference*, 61–65.

[20] Xiaobing Han, Yanfei Zhong, Liqin Cao, and Liangpei Zhang. 2017. Pre-Trained AlexNet Architecture with Pyramid Pooling and Supervision for High Spatial Resolution Remote Sensing Image Scene Classification. *Remote Sensing* 9, 8, 848.

[21] Dimitrios Marmanis, Mihai Datcu, Thomas Esch, and Uwe Stilla. 2016. Deep Learning Earth Observation Classification Using ImageNet Pretrained Networks. *IEEE Geoscience and Remote Sensing Letters* 13, 1:105–109.

[22] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. *IEEE conference on computer vision and pattern recognition*, 2818–2826.

[23] Jun Ding, Bo Chen, Hongwei Liu, and Mengyuan Huang. 2016. Convolutional Neural Network with Data Augmentation for SAR Target Recognition. *IEEE Geoscience and Remote Sensing Letters* 13, 3:364–368.

[24] Luis Perez and Jason Wang. 2017. The Effectiveness of Data Augmentation in Image Classification using Deep Learning. *arXiv:1712.04621 [cs]* (December 2017).

[25] Giovanni Delnevo, Marco Roccetti amd Silvia Mirri. 2019. Intelligent and good machines? The role of domain and context codification. Mobile Networks and Applications, Elsevier.

[26] Jorge Lazo. 2019. Detection of Archaeological Sites from Aerial Imagery using Deep Learning, Master Thesis. Lund Univesity Lybrary.

[27] Marco Roccetti, Giovanni Delnevo, Luca Casini and Paola Salomoni. 2020. A Cautionary Tale for Machine Learning Design: why we Still Need Human-Assisted Big Data Analysis. Mobile Neworks and Applications. Elsevier.

[28] Claudio E Palazzi, Stefano Ferretti, Stefano Cacciaguerra and Marco Roccetti. 2004. On maintaining interactivity in event delivery synchronization for mirrored game architectures. IEEE GLOBECOM Conference, IEEE.