Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

The Digital Geophysical Eye: A Magnetic Signature Database and Physics-Guided AI Framework for Archaeological Magnetic Interpretation

Abir Jrad

abirgeo@gmail.com

Water Research and Technologies Centre Borj Cedria: Centre de Recherche et des Technologies des Eaux de Borj Cedria https://orcid.org/0000-0002-8718-1216

Yoann Quesnel

CEREGE: Centre Europeen de Recherche et d'Enseignement des Geosciences de l'Environnement

Pierre Rochette

CEREGE: Centre Europeen de Recherche et d'Enseignement des Geosciences de l'Environnement

Chokri Jallouli

University of Tunis El Manar Faculty of Sciences of Tunis: Universite de Tunis El Manar Faculte des Sciences de Tunis

Pierre Ethienne Mathé

CEREGE: Centre Europeen de Recherche et d'Enseignement des Geosciences de l'Environnement

Research Article

Keywords: Digital Geophysical Eye, Magnetic Signature Database, Forward modelling, Convolutional Neural Network, Explainable Artificial Intelligence

Posted Date: July 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10207513/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

The Digital Geophysical Eye: A Magnetic Signature Database and Physics-Guided AI Framework for Archaeological Magnetic Interpretation

Abir Jrad[a], Yoann Quesnel[b], Pierre Rochette[b], Chokri Jallouli[c,d], Pierre Ethienne Mathé[b]

[a] *Georesources Laboratory LGR, Water Research and Technologies Centre, BP 273, 8020, Soliman, Tunisia*

[b] Aix-Marseille University, CEREGE, CNRS, IRD, UM34, CEREGE, Aix-en-Provence, France.

[c] Geological Department, College of Science, King Saud University, P.O Box. 2455, Riyadh, 11451, KSA.

[d] University of Tunis El Manar, Faculté des Sciences de Tunis, 2092 Tunis, Tunisia.

Corresponding author : *abir.jrad@certe.rnrt.tn*

Abstract

Magnetic prospection is one of the best non invasive archaeological investigation techniques. Despite the progress in instrumentation and data processing, the interpretation of archaeological magnetic anomalies is still largely based on expert knowledge and therefore not easily reproducible and standardized. This study proposes a methodological framework for formalizing and structuring archaeological magnetic expertise, named 'Digital Geophysical Eye', in a formal, reproducible and transferable interpretation process.

The proposed framework is based on a Magnetic Signature Database (MSD) which is structured by the dominant physical origin, not the archaeological function, of the magnetic signatures. The first implementation is based on one representative magnetic survey of the excavation and validated in the field, magnetic susceptibility measurements and physics-based forward modelling. Elementary source geometries that are suitable for archaeological structures were modeled as rectangular prisms, cylinders, irregular cylinders, discretized hollow cylinders, or spheres, with parameters derived from the archaeological observations taken at sites in Tunisia and France. Five representative archaeological classes were established: limestone walls, funerary incineration structures, palaeohearths, ceramic kilns and granite balls.

The physically constrained synthetic database consisted of 6,000 synthetic magnetic anomalies from five archaeological classes. A CNN model trained with these data resulted in an overall classification accuracy of 90.7% on the synthetic benchmark dataset generated from the validated magnetic signatures. Limestone walls and granite balls yielded the highest performance (F1-scores of 1.00 and 0.997, respectively), while the main ambiguities were among the combustion-related archaeological structures, such as funerary incineration structures, palaeohearths and ceramic kilns, whose thermoremanent magnetic signatures partially overlap. The feasibility of the proposed methodology was also confirmed by feature-

map analysis, which gradually revealed meaningful spatial features of the magnetic anomalies.

This work does not propose a fully operational AI system, but rather shows the potential of combining archaeological observations, laboratory measurements, forward modelling and artificial intelligence in a single, physics-based knowledge system. The Digital Geophysical Eye represents a proof of concept to build the methodological foundations for future applications of Explainable AI. Future Digital Geophysical Eye systems will be more robust and more generalizable as more and larger collections of excavation validated magnetic surveys from a variety of archaeological contexts are progressively added, and an international Magnetic Signature Library for archaeological magnetic interpretation will be developed.

Keywords: Digital Geophysical Eye; Magnetic Signature Database; Forward modelling; Convolutional Neural Network; Explainable Artificial Intelligence.

1. Introduction

Magnetic prospection is one of the most widely used non-invasive geophysical techniques in archaeological investigations due to its rapid acquisition, high spatial resolution and the ability to detect buried structures, without excavation (Aitken, 1958; Clark, 1996; Gaffney & Gater, 2003; Quesnel et al., 2011). Magnetic susceptibility, thermoremanent magnetization and induced magnetization can detect archaeological features like walls, kilns, hearths, furnaces, pits, ditches and funerary structures, and can often identify complete archaeological layouts prior to excavation (Linford, 2004; Aspinall et al., 2008; Neubauer, 2004). The sensitivity of magnetometers, positioning systems and data processing have made significant advances over the last 30 years, so that the quality and spatial resolution of archaeological magnetic surveys has improved significantly, enabling more complex archaeological landscapes to be surveyed efficiently.

Even with all these technological advances, the interpretation of magnetic anomalies is still largely reliant on human expertise. In contrast to the data acquisition process, which is very standardised, the interpretation of archaeological features is still largely dependent on the experience of the geophysicist, who uses geometric, polarity and amplitude data, along with archaeological background knowledge and knowledge of magnetic properties, to make inferences about the nature of the buried features (Clark, 1996; Gaffney & Gater, 2003; Aspinall et al., 2008). This interpretative process is not a formalized procedure and is usually developed over a number of years of field investigations, comparing with excavation work and experience. For this reason, two experienced interpreters may come up with two different archaeological interpretations for the same magnetic dataset, especially in complex archaeological environments.

Many studies have described the magnetic signatures of individual archaeological structures such as prehistoric hearths (Linford, 2004; Jrad et al., 2014), kilns (Jordanova et al., 2001),

funerary structures (Gómez-Paccard et al., 2012) and architectural remains (Neubauer, 2004). Overall, however, these observations have been isolated case studies. They have made significant strides in the interpretation of archaeological features, but they have not often been summarized into a coherent, transferable framework of knowledge that can be used to systematically classify the archaeological magnetic anomalies.

Artificial intelligence has significantly transformed image analysis in geosciences over the last decade. Machine learning and deep-learning methods have been successfully used in geosciences for remote sensing, seismic interpretation, geological mapping and potential-field analysis (LeCun et al., 2015; Goodfellow et al., 2016; Bergen et al., 2019; Travassos et al., 2021). AI has also been used to identify archaeological sites automatically using satellite imagery (Caspari & Crespo, 2019; Trier et al., 2019). More recently, convolutional neural networks (CNNs) have shown impressive results in the recognition of more complex spatial patterns directly from the images, without the need to define manually the features (Krizhevsky et al., 2012; LeCun et al., 2015). In the realm of archaeological geophysics, the recent efforts of AI technologies have centered on anomaly detection, image segmentation, object identification, or automated classification of geophysical and remote sensing data (e.g., Danese et al., 2025; Küçükdemirci et al., 2022). These strategies have shown the potential of deep learning for archaeological prospection but are mostly based on image-based learning and are less often formalized in geophysical reasoning for the expert interpretation.

Despite these developments, there is still limited use of AI in archaeological magnetic prospection. While a number of existing studies have concentrated on image segmentation and anomaly detection or geophysical inversion, there has been relatively little work on the formalization of the process of expert reasoning that underlies the archaeological interpretation of magnetic anomalies. AI algorithms are usually trained on existing datasets, and they do not explicitly include the geophysical knowledge that has been gained from decades of archaeological research. This has left the field of archaeological knowledge and AI largely untouched.

The main hypothesis of the present study is that magnetic descriptors can be formalized and made structured, and then learned by artificial intelligence, from the experience of an experienced geophysical interpretation. This knowledge, known as the Geophysical Eye, is the cognitive process by which geophysicists are able to identify characteristic archaeological magnetic signatures, by analysing the anomaly geometry, polarity, spatial organisation, magnetic origin and archaeological context. This tacit knowledge can be codified into quantitative terms, thus creating the basis for a repeatable computational tool, which is the Digital Geophysical Eye.

The concept was tested through the organisation of archaeological observations, gathered in Tunisia and France, in a Magnetic Signature Database, involving both excavation, validated magnetic surveys and physically constrained forward modelling, performed in the laboratory. The first version of this database is based on five representative archaeological classes. Instead of relying only on measured magnetic maps, the proposed framework combines complementary sources of archaeological information, such as field magnetic surveys,

magnetic susceptibility measurements and archaeological documentation, while maintaining the physical meaning of each magnetic signature, resulting in a sufficient number of training data for machine learning applications.

A summary of the overall methodological framework is presented in Figure 1. The archaeological observations are first translated into quantitative magnetic descriptors and are stored in the Magnetic Signature Database. These descriptors are then used to create physically constrained synthetic anomalies for training a Convolutional Neural Network as a proof of concept. The goal is not to create a universal archaeological classifier, but rather to assess if the characteristic signatures traditionally identified by experts in the field of geophysics can be developed by image-based artificial intelligence without losing the archaeological and geophysical value.

The goals of this study are then fourfold: (1) to develop a structured Magnetic Signature Database of archaeological magnetic signatures; (2) to introduce the concept of the Digital Geophysical Eye as a reproducible interpretation framework; (3) to test the feasibility of automatic recognition of representative archaeological magnetic signatures through the use of a Convolutional Neural Network; and (4) to set the methodological foundations for an international reference database of excavation, validated archaeological magnetic signatures for future explainable AI in archaeological prospection.

2. The Digital Geophysical Eye: Formalizing Archaeological Magnetic Expertise

Archaeological magnetic survey is essentially a spatial pattern recognition process. The automated signal processing typically used to enhance or filter magnetic data is not the only type of information needed for archaeological interpretation, which also involves integrating a variety of other data such as the geometry of the anomalies, their polarity, their amplitude and wavelength, their spatial organisation, the geological context and the archaeological knowledge (Clark, 1996; Gaffney & Gater, 2003; Aspinall et al., 2008). As the geophysical survey results are compared with excavation results over time, the experienced geophysicist will gradually develop the ability to identify the magnetic signature of the archaeological structure.

This acquired knowledge is called the Geophysical Eye in the current study. It is not an interpretation rule but rather a cognitive process that involves simultaneously evaluating several magnetic descriptors to determine the most probable archaeological origin of an anomaly. But in practice, a good interpreter is not likely to be able to locate a buried structure based on the amplitude of the magnetic signal. The interpretation, on the other hand, is derived from the interaction of a number of complementary features, such as the geometry, polarity, size, spatial continuity and magnetic strength of the anomaly, and the magnetic characteristics of the archaeological material.

This interpretation is well known in archaeological geophysics and is based on physical evidence. Archaeological features, such as fired structures including kilns, furnaces, palaeohearths and cremation features, typically have strong magnetic anomalies with high thermoremanent magnetic component acquired during firing (Jordanova et al., 2001; Linford,

2004). Magnetic susceptibility contrasts of the masonry, including limestone walls, are normally identified by the magnetic anomalies, which will have different geometry and amplitude to the combustion related magnetic anomalies (Neubauer, 2004). Likewise, lithological artefacts such as granitic objects show magnetic responses that are largely determined by their intrinsic magnetic susceptibility. Archaeological items with different functions can have different magnetic signatures even if they are buried in the same geologic context, because of these differences in magnetic origin.

Induced Magnetization (Eq.1):

$$Mi=\chi H_0 \qquad \text{(Eq.1)}$$

- Mi : induced magnetization
- χ : magnetic susceptibility
- H_0 : Earth's magnetic field

Total Magnetization (Eq.2)

$$M=Mi+Mr \quad \text{(Eq.2)}$$

- Mr : thermoremanent magnetization.

The magnetic anomaly observed is either induced or thermoremanent or a mixture of both, depending on the archaeological material. These are the interpretative principles commonly applied in archaeological work but they are still largely subjective and based on personal experience. Therefore, archaeological magnetic interpretation is hard to standardise, reproduce and transfer from one operator to another. The same magnetic data can thus be interpreted in different ways by different interpreters, especially when the data is collected in a complex archaeological field, where many sources of anomalies are present.

The goal of the present study is to formalize this kind of expert reasoning in a repeatable computational model, which is called the Digital Geophysical Eye. The Digital Geophysical Eye is not a replacement for geophysical skills, but a clear description of them in terms of measurable magnetic descriptors that can be organized, compared and eventually learned by artificial intelligence. In this context, the geophysical interpretation is described as a hierarchical process, in which the observations made in the field are increasingly translated into magnetic signatures in the form of magnetic quantities, which are then utilized for machine learning applications.

Figure 1 shows the conceptual process of the Digital Geophysical Eye. The methodology starts with the archaeological evidence derived from magnetic survey and magnetic measurements carried out in the laboratory, and from archaeological documentation, both of which are backed up by the results of the excavation. The observations are then converted to quantitative magnetic descriptors that describe the geometry, polarity, amplitude, wavelength and the dominant magnetic origin of the anomaly. The descriptors are stored in a Magnetic Signature Database that forms the knowledge base of the Digital Geophysical Eye. A representative set of synthetic magnetic anomalies are then generated using physically constrained forward modelling, while maintaining the physical properties of the

archaeological observations. Lastly, the magnetic signatures are standardized and used to train a Convolutional Neural Network (CNN) that can recognize typical archaeomagnetic patterns.

Excavation provided archaeological magnetic data is very limited and only one field example was available for each archaeological class. Each representative example was then used to create a large number of physically constrained synthetic anomalies by changing the burial depth, geometry, orientation, magnetic intensity, and background noise of each anomaly within realistic archaeological constraints, without any augmentation of the image.

This workflow shows that artificial intelligence is not the main goal of the methodology, but rather the last phase. The uniqueness of the Digital Geophysical Eye is the explicit formalization of the archaeological magnetic expertise in a structured and reproducible framework that connects the archaeological observations, the magnetic measurements, forward modelling and machine learning. The CNN is thus a proof of concept that the knowledge traditionally learned from the expert's interpretation can be translated into quantitative descriptors that can be used for automatic recognition without losing the archaeological and geophysical value.

The Digital Geophysical Eye must be considered a methodological framework that is constantly evolving rather than a classification algorithm. The Magnetic Signature Database will continue to grow as new excavation, validated magnetic surveys are added, adding to the knowledge base and strengthening future AI models.

3. Magnetic Signature Database

The Digital Geophysical Eye is based on an organized representation of the archaeological magnetic signatures that can be used to convert expert interpretation to quantitative information. To achieve this objective, a preliminary Magnetic Signature Database (MSD) was developed. This database classifies archaeological features according to the physical mechanisms controlling their magnetic response, instead of an archaeological classification only.

The proposed database brings together several sources of archaeological and geophysical evidence, which is different from conventional image databases that are used for machine learning. These include excavation validated magnetic surveys, laboratory measurements of magnetic susceptibility and natural remanent magnetization, archaeological documentation and physically constrained forward modelling. The combination of these complementary sources makes it possible to maintain a direct link between the magnetic signatures and measurable physical properties, and to retain the magnetic interpretation.

To create the first version of the Magnetic Signature Database, five representative archaeological classes were selected: limestone walls (archeological site in Hergla, Tunisa), funerary incineration structures (necropolis of Richeaume XIII, France), palaeohearths (Lazaret Cave in Nice, France), ceramic kilns (Cordouls in Puylaurens, France) and granite balls (Carthage Museum in Tunisia). The classes were selected because they are different magnetic response classes that are frequently encountered in archaeological prospection and

because there was a reliable archaeological validation for each class. Excavation validated magnetic surveys of limestone walls (Jrad et al., 2024), incineration structures (Quesnel et al., 2011) and palaeohearths (Jrad et al., 2014) are complemented by the kiln class, which is based on the magnetic properties of fired ceramic materials and archaeological observations (Boissinot, 2010). The granite ball was not detected as a field magnetic anomaly, as is typical for the other archaeological classes. Instead, its magnetic signature was obtained by measuring magnetic susceptibility and forward magnetic modelling using the geometry of the object (Bhattacharyya, 1964; Blakely, 1995). In this way, physically realistic magnetic anomalies were generated to represent granitic archaeological artefacts.

Instead of qualitative description of the archaeological objects, each class is defined by a common set of quantitative magnetic descriptors (Eq.3).

$$S=(A,P,\lambda,G,D,M) \quad \text{(Eq.3)}$$

- A = amplitude
- P = polarity
- λ = wavelength
- G = geometry
- D = burial depth
- M = dominant magnetic origin

The descriptors are comprised of continuous variables (amplitude, wavelength, and burial depth) and categorical variables (dominant magnetic origin, and anomaly geometry), which enables expert knowledge of the archaeological information to be captured in a structured form that can then be incorporated into a database for subsequent machine learning applications.

A quantitative descriptor vector (Eq.4) is thus used to represent each archaeological magnetic signature.

$$Si=(Ai,Pi,\lambda i,Gi,Di,Mi) \quad \text{(Eq.4)}$$

These descriptors include the dominant magnetic origin (induced or thermoremanent magnetization), anomaly polarity, geometry, wavelength, spatial extent, relative amplitude and geophysical interpretation. All these parameters are used together to make the traditional diagnostic criteria assessed by an experienced geophysicist when interpreting magnetic data.

The resulting classification is summarized in Figure 2, which forms the nucleus of the Digital Geophysical Eye. The intent of this classification is not to supplant the magnetic interpretation process, but to codify the reasoning that has been undertaken by the expert. The Magnetic Signature Database allows for the expression of archaeological magnetic signatures in a standardized manner, resulting in a reproducible framework that can be expanded in a systematic way as new observations are made from new excavations.

The magnetic signatures in the database were then physically constrained forward modeled into standard magnetic images to make them suitable for machine learning applications. The

relationship between archaeological evidence and physical characterization, is illustrated in Figure 2. Synthetic anomalies (Figure 3) were obtained from magnetic surveys and forward modelling all the classes except the granite ball. As a result, all synthetic anomalies in the Magnetic Signature Database are linked to known archaeological data and physical characteristics. These latter retain the main diagnostic features of the archaeological class to which they belong, but with controlled variations in orientation, size, burial depth and magnetic intensity. These differences enhance the training set with additional variation, but do not change the physical behavior of the archaeological sources.

The Magnetic Signature Database is thus much more than a training set for a CNN. It is the knowledge structure of the Digital Geophysical Eye, which provides a standardised way of structuring the archaeological magnetic knowledge into quantitative descriptors that are reproducible, extensible and applicable to artificial intelligence. The database can gradually transform into an international reference library for explainable AI and transfer of geophysical knowledge between archaeological projects and geographical areas as more archaeological classes and magnetic surveys are added, validated by excavation.

4. From the Magnetic Signature Database to AI Assisted Recognition

4.1. Generation of the synthetic magnetic dataset

The Magnetic Signature Database detailed in the previous section is the knowledge base of the Digital Geophysical Eye. The number of existing magnetic anomalies, however, is too small at this time to train a deep learning model, as they are only available from excavation sites. A synthetic magnetic dataset was created using physically constrained forward modelling to overcome this limitation and maintain the physical meaning of the magnetic signatures.

Class specific forward modelling was used to generate representative magnetic anomalies based upon the archaeological evidence summarised in Figure 2. The identification of a single standard model for all archaeological features was avoided, each class being represented by a geometrical body based on the morphology and predominant magnetic response. Long rectangular prisms were used to represent limestone walls, shallow cylindrical bodies to represent funerary incineration structures and palaeohearths, hollow cylindrical (ring shaped) bodies to represent ceramic kilns and spherical bodies to represent granite balls. The magnetic properties that were given to each of the models were restricted by field magnetic surveys, magnetic susceptibility measurements, and archaeological observations and excavation data.

To reproduce the natural variability encountered in archaeological magnetic prospection, the forward models were generated using controlled variations in source dimensions, burial depth, orientation, magnetization intensity and background noise. This approach produced a physically consistent synthetic dataset that preserves the characteristic magnetic behaviour of each archaeological class while providing sufficient variability for robust CNN training. The synthetic images are therefore firmly grounded in documented archaeological evidence and measurable physical properties, thus providing a physically meaningful archaeological magnetic signature rather than a numerical simulation (Eq.5).

$$\Delta T = f(G, M, z, H_0) \quad \text{(Eq.5)}$$

- G = source geometry
- M = magnetization
- z = burial depth
- H_0 = geomagnetic field.

The aim of this procedure is not to capture all possible archaeological scenarios but to create a standardised image database that will record the key diagnostic features identified by the more experienced geophysicists. This way, the Digital Geophysical Eye can represent the archaeological knowledge in a way that is compatible with machine learning, while preserving the connection between the training images and the actual geophysical observations.

The magnetic response is thus a function of the geometry of the source as well as the magnetic properties (Eq.6).

$$S = f(G, M, z, \theta, \eta) \quad \text{(Eq.6)}$$

- θ : orientation
- η : background noise.

4.2. CNN proof of concept

In order to test the possibility of automatic recognition of the characteristic magnetic signatures stored in the Magnetic Signature Database, a Convolutional Neural Network (CNN) was developed as a proof of concept. CNNs are one of the most successful deep-learning architectures for image analysis due to their ability to extract hierarchical spatial features directly from images without manually defined descriptors (Krizhevsky et al., 2012; LeCun et al., 2015; Goodfellow et al., 2016).

The goal of the CNN is not to replace geophysical interpretation but to assess the discriminatory information in the magnetic descriptors formalized by the Digital Geophysical Eye that is sufficient for automatic classification. This meant that the network architecture was kept relatively simple, with the aim of studying the feasibility of the proposed methodology, rather than optimizing the performance of the deep learning.

These anomalies were then transformed into standardized grayscale images of the same size and randomly split into training, validation and testing sets. To achieve consistent numerical ranges between samples while preserving the relative magnetic contrasts of each archaeological class, the following procedures were used:

CNN is composed of a series of convolutional layers that learn spatial patterns and max pooling layers that gradually downsample the image while retaining the most informative aspects (Eq.7)

$$F(i,j) = \sum_m \sum_n I(i-m, j-n)K(m,n) \quad \text{(Eq.7)}$$

- I = input image
- K = convolution kernel
- F = feature map.

These features are then flattened and finally linked to fully connected layers with the final Softmax classification layer that classifies each anomaly into one of the five archaeological classes in the Magnetic Signature Database.

The Adam optimizer and categorical cross entropy loss function were used for training. Early stopping was employed to prevent overfitting while retaining the model which had the lowest validation loss.

The trained CNN was assessed based on overall classification accuracy. This measures how many samples were correctly classified out of all the test instances. Accuracy (Eq.8) is defined as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad \text{(Eq.8)}$$

- TP : True Positive
- TN : True Negative
- FP : False Positive
- FN : False Negative

Since the dataset has various archaeological classes, we calculated more class-specific evaluation metrics to get a clearer picture of how well the classification is performing. Precision (Eq.9), Recall (Eq.10) and the F1-score (Eq.11) are defined as follows:

- $Precision = \frac{TP}{TP+FP}$ (Eq.9)
- $Recall = \frac{TP}{TP+FN}$ (Eq.10)
- $F1 = 2\frac{Precision*Recall}{Precision+Recall}$ (Eq.11)

These evaluation metrics offer additional insights into how classifiers perform. They measure the trustworthiness of positive predictions (Precision), the effectiveness in correctly identifying samples from each class (Recall), and their harmonic balance (F1-score).

The methodology starts from the archaeological evidence that has been added to the Magnetic Signature Database, then it gradually creates synthetic magnetic anomalies, preprocesses the images, trains the convolutional neural network and finally predicts the archaeological class corresponding to each magnetic signature. In this process, Artificial Intelligence is not the goal but rather the last step in the Digital Geophysical Eye.

The trained CNN is then tested on an independent testing set to verify its recognition capability of representative archaeological magnetic signatures from the physically constrained modelling framework. The corresponding training behaviour, classification

performance and learned spatial representations are presented in the following Results section.

The framework was implemented in Python version 3 using TensorFlow, Keras, NumPy, Pandas, Scikit-learn and Matplotlib. We used Google Colab with an NVIDIA T4 GPU to train and evaluate the model. The images were resized to 64 x 64 grayscale pixels and normalized to a range of 0 to 1. The dataset had 6000 images, which we randomly split into 70 % for training 15 % for validation and 15 % for testing. The CNN had four blocks with 3 x 3 kernels then batch normalization max pooling, global average pooling, a layer with 128 neurons and 30% dropout and finally a softmax output layer, with five classes. We trained the model using the Adam optimizer with a learning rate of 5×10^{-4}. The model used cross-entropy loss, a batch size of 32 and stopped early based on validation loss.

5. Results

5.1. Construction of the Magnetic Signature Database

The methodology described in Sections 3 and 4 led to the first archaeological magnetic prospection dedicated Magnetic Signature Database (MSD). Five representative archaeological classes are included in the database, backed up by additional evidence, such as magnetic survey validated by excavation and laboratory magnetic measurements and physically constrained forward modelling.

Figure 2 shows the diversity of magnetic sources considered in this study, and the archaeological evidence used in the construction of the database. Excavation validated magnetic surveys are used to represent limestone walls; funerary incineration structures and palaeohearths are represented by a combination of field magnetic observations and excavation data and laboratory measurements on fired ceramics for the ceramic kiln. Archaeological artefacts with magnetic susceptibility measurements and physically constrained forward modelling are used to represent the granite ball class.

For each archaeological class, quantitative descriptors including anomaly amplitude, polarity, wavelength, geometry, orientation and burial depth were extracted from excavation, validated magnetic surveys and used to parameterize the corresponding forward models. These descriptors are the knowledge base of the Digital Geophysical Eye and the information needed to create representative synthetic magnetic anomalies for machine learning applications.

5.2. Generation of Synthetic Magnetic Anomalies

For each archaeological class, representative synthetic magnetic anomalies were created with physically constrained forward modelling using the Magnetic Signature Database (Figure 3).

The synthetic anomalies of each class retain the main diagnostic features found during the geophysical interpretation, such as the geometry of the anomaly, its polarity, the extent of the area and relative magnetic intensity. The variations in burial depth, orientation, size and background noise were controlled to simulate the natural variability that would be found

during archaeological prospection, but with the same physical properties of the magnetic responses.

The synthetic dataset is thus an extension of the archaeological observations and not a separate numerical simulation. All the anomalies generated are traceable back to geophysical measurements and physical properties.

5.3. CNN training performance

The synthetic magnetic dataset was then used to train the convolutional neural network (CNN) described in Section 4. The plots for the evolution of the training and validation accuracy (8) are given in Figure 4a and the loss functions are given in Figure 4b.

The two accuracy curves show stable convergence during training, suggesting that the chosen architecture is capable of capturing the spatial structure of the five archaeological classes in the synthetic data, which was generated under controlled conditions. Similarly, both loss curves show a gradual decline followed by a steady level, indicating a gradual optimization process without notable differences between the training and validation sets.

During the first few epochs, training errors merged quickly, from around 20% to over 80% in terms of validation accuracy, which suggests that the network quickly learned the main spatial features of the synthetic magnetic signatures. The minimum validation loss was achieved at epoch 14 and further training did not yield any notable improvement. Early stopping thus kept the model at this optimum, thus not allowing unnecessary further training and avoiding overfitting.

The early stopping automatically selected the model with the lowest validation loss, without unnecessary iterations and keeping the network which has the best validation performance.

5.4. Classification performance

The physically constrained synthetic magnetic signature database was used to generate an independent testing dataset to assess the classification capability of the proposed Digital Geophysical Eye.

The CNN was able to achieve an overall classification accuracy of 90.7% on the independent testing dataset (Table 1). The performances were different depending on how similar the archaeological classes were. The limestone walls and granite balls were almost perfectly identified with F1-scores of 1.000 and 0.997, respectively. These classes have unique induced magnetic signatures and well defined geometries, and are easily distinguished from combustion associated structures. The primary classification difficulties were with funerary incineration structures, palaeohearths and ceramic kilns. The confusion matrix (Figure 5) shows that funerary incineration structures were occasionally classified as ceramic kilns (14 cases) or palaeohearths (13 cases), whereas ceramic kilns were mainly confused with funerary incineration structures (23 cases). The errors are physically consistent because the three classes are all dominated by thermoremanent magnetization and have small positive anomalies with similar spatial properties.

These ambiguities are also frequently found in manual archaeological interpretation, in which the identification of different fired structures often requires further archaeological and geological data.

These results illustrate that the magnetic descriptors defined in the Magnetic Signature Database provide a sufficient amount of information to enable automatic recognition of representative archaeological magnetic signatures. The current mapping of magnetic anomaly morphology to archaeological classes without losing their physical meaning is a proof of concept, but the classifier was able to learn meaningful relationships.

5.5. Learned magnetic representations

The feature maps generated from the first convolutional layer (Figure 6) give direct insights of the information learnt by CNN. Each filter highlights different magnetic properties of the data, not the original anomaly. The elongated geometry and sharp polarity transition are the main features of the limestone wall that are highlighted by several of the filters, while others emphasize the anomaly boundaries and orientation. The filters gradually isolate the compact positive core, the surrounding negative halo and the radial intensity gradients in the combustion related structures (funerary incineration structure, palaeohearth and ceramic kiln). While the thermoremanent magnetic response of these classes is similar, the differences in the size and spatial distribution of the anomalies are captured by different filters, which is why the CNN is able to classify these classes with good, but not perfect, results. The granite ball yields the simplest feature maps, which are characterized by a small scaled symmetric response of an isolated induced magnetic source.

These observations show that the CNN is not a memory device for single magnetic pictures. Rather, it gradually breaks down each anomaly into its constituent spatial descriptors, similar to those that are used tacitly by skilled geophysicists in the visual interpretation process.

6. Discussion

6.1. From expert interpretation to a reproducible digital framework

The identification of characteristic anomaly patterns by experienced geophysicists and the correlation of these patterns with buried archaeological features have been the main factors in the success of archaeological magnetic prospection (Clark, 1996; Gaffney & Gater, 2003; Aspinall et al., 2008). This interpretation process has been successfully used over the last few decades, but is still largely based on individual experience and has not been codified into a clear, repeatable, and explicit process.

The main contribution of the present study is therefore not the use of a convolutional neural network, but rather the formalization of the archaeological magnetic expertise into the Digital Geophysical Eye. The expert interpretation is gradually being replaced by a structured workflow that links the archaeological observation to quantitative magnetic descriptors, physically constrained forward modelling and artificial intelligence as illustrated in Figure 1. This method transforms the cognitive process that has been acquired over many years of

archaeological practice into a repeatable computational process, without losing its concrete foundation.

The development of the Magnetic Signature Database is one of the major components of this formalisation. Instead of the archaeological descriptions being stored in the database, each archaeological class is arranged according to the dominant magnetic origin, the diagnostic spatial characteristics and the validation method. This classification is a direct reflection of the reasoning process that most of the more experienced archaeological geophysicists would undertake implicitly, and provides a framework that can be progressively expanded with further case studies from the excavations.

The Magnetic Signature Database is not a typical image database. Rather it is a physics-based knowledge base, which integrates a number of sources of geophysical information, including magnetic surveys confirmed by excavation, magnetic susceptibility measurements, material magnetic properties, archaeological observations and forward modelling.

The reported classification results apply only to the controlled synthetic dataset. They show that the Magnetic Signature Database contains enough diagnostic information to distinguish the selected archaeological classes. They do not represent the expected performance on real archaeological surveys.

6.2. The Magnetic Signature Database is physically consistent.

One of the big challenges in applying deep learning to archaeological geophysics is the lack of magnetic data from archaeological excavations. Archaeological magnetic surveys are typically not large enough to provide the large labelled datasets required for training deep neural networks as are used in conventional computer vision applications. Thus, numerous published studies of AI are based on image augmentation or purely numerical simulations, without any direct ties to archaeological observations.

The methodology proposed here tries to overcome this limitation by merging three complementary sources of information: excavation and validated magnetic surveys, laboratory magnetic measurements and physically constrained forward modelling (Figure 2). Each synthetic anomaly in the training set is then directly associated to the archaeological evidence, and therefore the physical behaviour of the original magnetic sources in the images generated is preserved. However, the synthetic models are idealized archaeological scenarios. Geological background variations, overlapping anomalies, acquisition artefacts, interpolation effects, heterogeneous soil magnetisation and partially preserved archaeological remains are possible sources of real magnetic surveys. Thus the synthetic data in this paper should be considered as a suitable controlled example for methodological validation and not as a complete representation of the magnetic complexity of a archaeological site. The difference between these is important for archaeological prospection as magnetic anomalies are associated with different physical mechanisms. The combustion related structures (Jordanova et al., 2001; Linford, 2004) are the ones that are dominated by thermoremanent magnetization, gained during the combustion of the structures (palaeohearths, ceramic kilns, funerary incineration structures). In contrast, induced magnetization due to magnetic susceptibility

contrasts with surrounding sediments is the most important control on architectural remains and granitic artefacts. This classification of the archaeological classes based on the physical mechanisms, offers a more solid foundation for machine learning applications and assists in interpretation of the machine learning classifications.

One of the major features of this new approach is that the synthetic data set does not attempt to produce all of the possible magnetic anomalies that can be found in archaeological features. Instead, it is the typical physical reaction of each archaeological class, some of which are intentional geometric, orientational, burial depth, magnetic intensity and noise variations. This means that the CNN is learning general physical properties, instead of specific archaeological examples, and will serve as a foundation for further development of the Magnetic Signature Database.

6.3. Artificial Intelligence as an Extension of the Geophysical Eye

The CNN designed in this study was not optimized for classification purposes, but rather it was designed as a proof of concept. The primary objective was to see whether it was possible to automatically learn the archaeological magnetic descriptors established in the Digital Geophysical Eye without losing their physical meaning.

Figure 4 shows the training curves, which showed that the proposed architecture successfully learned the reproducible spatial information from the synthetic magnetic dataset and converged stably during training. Likewise, the confusion matrix (Figure 5) reveals that the CNN was able to recognize limestone walls and granite balls with nearly 100% accuracy and the overall accuracy was 90.7%. The errors that were left were mainly between the combustion related archaeological structures, which suggest that the classifier retained the physical relationships between the classes of the archaeological structures rather than making random predictions. The lower performance observed for combustion related archaeological structures is due to the fact that the thermoremanent magnetic signatures of these structures are very similar in nature and not a limitation of the conceptual framework.

The feature maps obtained from the first convolutional layer (Figure 6) give some clues about the internal dynamics of the network. The first filters are used to identify basic spatial properties like polarity changes, anomaly edges, local gradients and compact magnetic sources. The simple representations are then progressively connected to other simple representations, resulting in more complex spatial patterns that are increasingly similar to the diagnostic criteria that are typically used by professional archaeophysical interpreters to carry out magnetic surveys.

The Digital Geophysical Eye should not therefore be considered as an autonomous interpretation system which can replace the geophysicist. Instead, AI is a computer-based extension of expert reasoning, in which it encodes common magnetic signatures as repeatable quantitative descriptors. The final archaeological interpretation is thus kept under the control of the experts and can be supported by a transparent and physically consistent decision making process.

6.4. Perspectives and limitations

The present study demonstrates that the archaeological magnetic knowledge can be codified into an artificial intelligence approach inspired by physics. The classification performance achieved for some of the archaeological classes in this initial implementation is still moderate, but the results clearly illustrate that representative magnetic signatures can be learned from physically constrained synthetic data based on the archaeological observations made during excavation.

The current implementation's biggest drawback is the limited size of the Magnetic Signature Database. At present, only five representative archaeological classes have been selected from the case studies, and carefully chosen for the database. This approach ensures full physical traceability from archaeological observations to magnetic measurements and forward modelling, but cannot yet cover the full range of archaeological magnetic signatures that can be found in various geological and archaeological contexts.

Likewise, the synthetic training data set was created using physically constrained forward models based on known archaeological observations (Figure 3). The number of magnetic surveys acquired by the excavators that have been validated should increase in the future versions of the Digital Geophysical Eye, as should the number of surveys conducted under different geological conditions and acquisition parameters and archaeological contexts, while maintaining the physical meaning of the magnetic signatures. This increase would help improve the robustness and usability of future AI models.

Further developments will also involve more advanced AI methods like Gradient weighted Class Activation Mapping (Grad-CAM) (Selvaraju et al., 2017), Vision Transformers (Dosovitskiy et al., 2021) and attention-based architectures that will allow geophysicists to identify the most important parts of the magnetic anomaly for the final classification. Meanwhile, other geophysical measurements, such as vertical magnetic gradients, multi-height magnetic readings, reduction to the pole transformations (if applicable), and other physical measurements may be useful to better distinguish between archaeologically similar magnetic features.

Because there are few magnetic datasets that have been confirmed by excavation, the field observations that are available have been used to build and check the Magnetic Signature Database, instead of being used to independently test the CNN. Future work will progressively add newly excavated sites to build an independent external validation dataset.

Lastly, the Digital Geophysical Eye aims to become a global Magnetic Signature Library of excavation proven archaeological magnetic survey data from a range of geological sites, from the proof of concept described here. This joint database would slowly enhance the knowledge transfer in archaeology and the development of future explainable AI systems, which would be dedicated to archaeological magnetic interpretation.

7. Conclusion

This study presents the Digital Geophysical Eye, a new methodological approach in which the expertise of an geophysicist is formalized into a physics-driven and explainable computational process, that can be reproduced. Although significant progress has been made in magnetic instrumentation, in survey methodology and in the processing of magnetic data, much of the archaeological magnetic interpretation remains in the hands of individual geophysicists. The proposed framework overcomes this limitation by encoding expert reasoning in standardized quantitative magnetic descriptors that are directly related to measurable physical properties.

The key innovation of the proposed approach is the development of the Magnetic Signature Database that integrates the results of magnetic surveys with the results of magnetic measurements in the laboratory and the results of physically constrained forward modelling in a single knowledge base. The database is different from the traditional image databases in that it retains the physical location of the magnetic anomalies found in the archaeological site and also offers standardised descriptors that are suitable for use in AI applications.

The proof of concept convolutional neural network showed that representative archaeological magnetic signatures can be automatically learned from this physics-guided database in controlled modelling conditions. The results also demonstrated that classes with characteristic magnetic responses, such as limestone walls and granite artefacts, were identified with high reliability, while the main ambiguities were between archaeological structures with similar thermoremanent magnetic characteristics, related to combustion. The observations made are entirely consistent with the physical behaviour of archaeological magnetic sources, and with the problem which is frequently faced in an expert interpretation.

The key contribution of this work is not the development of a specific deep learning architecture, but rather the provision of a reproducible methodology that combines archaeological observations, magnetic measurements, forward modelling with artificial intelligence in a coherent framework. Within this framework, the CNN's importance lies mainly in the proof of concept obtained by the Digital Geophysical Eye, and it does not present the most significant contribution of our study.

The present implementation is made on purpose as a first proof of concept. Future excavations and archaeological case studies will be added to the Magnetic Signature Database, and new magnetic descriptors and additional geologic contexts will be added to the database to further expand the collection. These developments will enhance the ability to discriminate between magnetically similar archaeological structures and keep the direct link between artificial intelligence and physically interpretable archaeological evidence.

The Digital Geophysical Eye should thus be considered as the methodological basis of an International Magnetic Signature Library for Archaeological Geophysics in the future. This framework could enhance the reproducibility, transparency and transferability of archaeological magnetic interpretation, and offer a solid foundation for future explainable artificial intelligence in archaeological prospection by progressively integrating validated archaeological magnetic signatures from various research groups. The CNN results show that the proposed magnetic descriptors have enough information for the reliable automatic recognition under controlled modelling conditions. More importantly, the classification errors

that were observed are physically meaningful, and are almost exclusively between classes of archaeological material with similar thermoremanent magnetic properties. This indicates that the Digital Geophysical Eye is able to capture the physical reasoning used by the geophysicist and offers a repeatable framework that could be used by future explainable AI.

And even if the Digital Geophysical Eye was developed for archaeological magnetic prospection, but it coul be a general model of knowledge formalization in applied geophysics. The same approach could be taken for other geophysical techniques such as gravity, electrical and electromagnetic surveys, where measurable physical properties can be used to define method specific signature databases and forward modelling. Similarly, the method could be extended to other geoscience problems, such as magnetic anomaly recognition in mineral exploration, geological mapping, or environmental geophysics, where expert interpretation is still significantly reliant on experience, and where explainable AI is useful, with physically constrained knowledge bases.

Acknowledgments

The authors would like to express their sincere thanks to the archaeologists and site directors who made the magnetic surveys and the archaeological test of the detected magnetic anomalies possible. We would like to thank Professor Taher Ghalia for the Hergla archaeological site (Tunisia), Dr Florence Mocci for the Richeaume archaeological site (France), Professor Samir Khatib for the Lazaret archaeological site (France) and Professor Samir Aounallah and Dr Moez Achour for the archaeological sites of Carthage (Tunisia). Their cooperation, which also involved giving access to the results of the excavations and to the archaeological interpretation of the magnetic anomalies, was crucial in the development and validation of the Magnetic Signature Database described in this study.

Conflict of Interest statement: The authors declare that they have no conflict of interest.

Data availability statement: Data will be available on request.

References

Aitken, M. J. (1958). Magnetic prospecting in archaeology. Antiquity, 32(126), 270–271.

Aspinall, A., Gaffney, C., & Schmidt, A. (2008). Magnetometry for archaeologists. Alta Mira Press.

Bhattacharyya, B. K. (1964). Magnetic anomalies due to prism-shaped bodies with arbitrary polarization. Geophysics, 29, 517–531.

Bergen, K. J., Johnson, P. A., de Hoop, M. V., & Beroza, G. C. (2019). Machine learning for data-driven discovery in solid Earth geoscience. Science, 363(6433), eaau0323. https://doi.org/10.1126/science.aau0323

Blakely, R. J. (1995). Potential Theory in Gravity and Magnetic Applications. Cambridge University Press.

Boissinot, P. avec la collaboration du Club Archéologique du Puylaurentais (2010) Cordouls (Puylaurens, Tarn): Campagne 2010, Sondages programmés. SRA Occitanie.

Clark, A. (1996). Seeing beneath the soil: Prospecting methods in archaeology (2nd ed.). Routledge.

Caspari, G., & Crespo, P. (2019). Convolutional neural networks for archaeological site detection: Finding "princely" tombs. Journal of Archaeological Science, 110, 104998. https://doi.org/10.1016/j.jas.2019.104998

Danese, M., et al. (2025). AI methods for enhancing and recognizing archaeological features in heterogeneous geophysical datasets. Scientific Reports. https://pmc.ncbi.nlm.nih.gov/articles/PMC12460655/

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations (ICLR).

Gaffney, C., & Gater, J. (2003). Revealing the buried past: Geophysics for archaeologists. Tempus Publishing.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Jordanova, N., Kovacheva, M., & Kostadinova, M. (2001). Archaeomagnetic investigations of archaeological structures subjected to heating. Physics and Chemistry of the Earth, Part A: Solid Earth and Geodesy, 26(6–8), 555–562.

Jrad, A., Quesnel, Y., Rochette, P., Jallouli, C., Khatib, S., Boukbida, H. and Demory, F. (2014), Magnetic Investigations of Buried Palaeohearths Inside a Palaeolithic Cave (Lazaret, Nice, France). Archaeol. Prospect., 21: 87-101. https://doi.org/10.1002/arp.1469

Jrad, A., Collins-Elliott, S.A., Akerraz, A., Ben Romdhane, H. (2024). The State of Archaeo-geophysics in the Maghreb: Case Studies from Tunisia and Morocco. In: Cuenca-Garcia, C., Asăndulesei, A., Lowe, K.M. (eds) World Archaeo-Geophysics. One World Archaeology. Springer, Cham. https://doi.org/10.1007/978-3-031-57900-4_14

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

Küçükdemirci, M., et al. (2022). GPR Data Processing and Interpretation Based on Artificial Intelligence: A Review. Remote Sensing, 14, 3377. https://www.mdpi.com/2072-4292/14/14/3377

Le Cun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. https://doi.org/10.1038/nature14539

Linford, N. (2004). Magnetic ghosts: Mineral magnetic measurements on Roman and Anglo-Saxon graves. Archaeological Prospection, 11(3), 167–180.

Neubauer, W. (2004). GIS in archaeological prospection. In H. Kamermans & K. Wansleeben (Eds.), The future of archaeological prospection (pp. 219–226). Archaeological Studies Leiden University.

Quesnel, Y., Jrad, A., Mocci, F., Gattacceca, J., Mathé, P.-.-E., Parisot, J.-.-C., Hermitte, D., Dumas, V., Dussouillez, P., Walsh, K., Miramont, C., Bonnet, S. and Uehara, M. (2011), Geophysical Signatures of a Roman and Early Medieval Necropolis. Archaeol. Prospect., 18: 105-115. https://doi.org/10.1002/arp.411

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-CAM: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 618–626.

Travassos XL, Avila SL, Ida N (2021), "Artificial Neural Networks and Machine Learning techniques applied to Ground Penetrating Radar: A review". Applied Computing and Informatics, Vol. 17 No. 2 pp. 296–308, doi: https://doi.org/10.1016/j.aci.2018.10.001

Trier ØD, Cowley DC, Waldeland AU. Using deep neural networks on airborne laser scanning data: Results from a case study of semi-automatic mapping of archaeological topography on Arran, Scotland. Archaeological Prospection. 2019;26:165–175. https://doi.org/10.1002/arp.1731

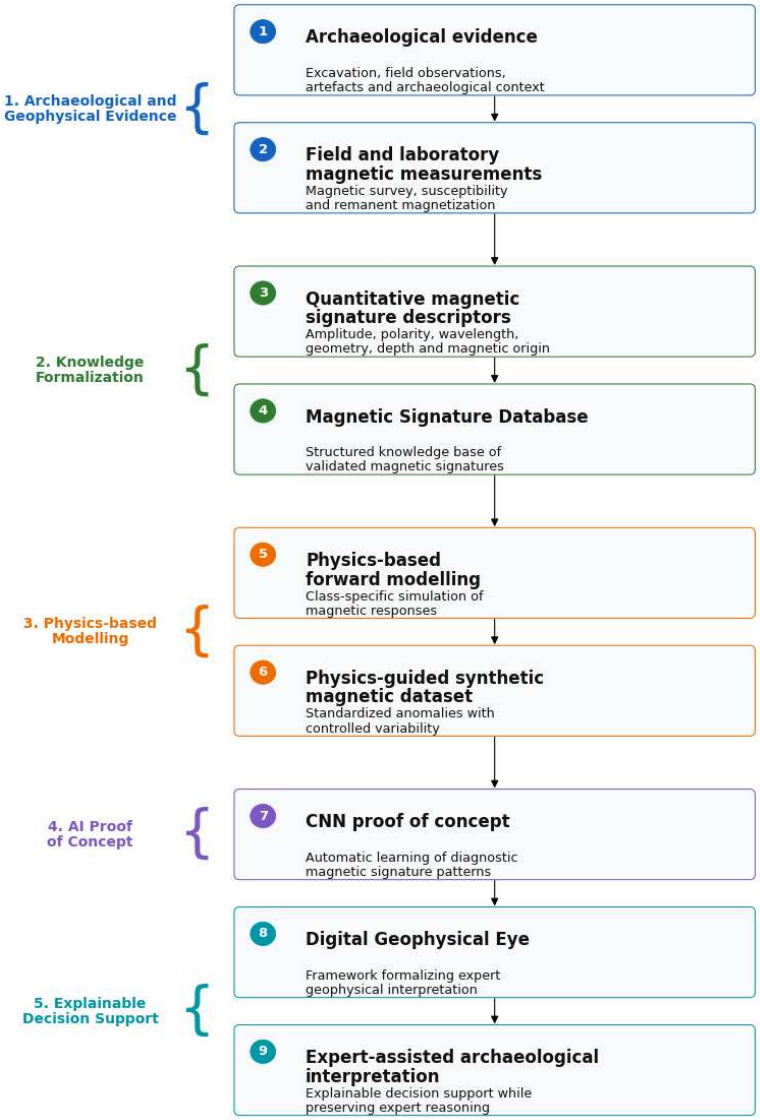


Figure 1. Conceptual framework of the Digital Geophysical Eye

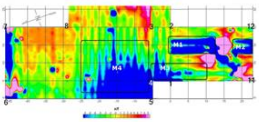
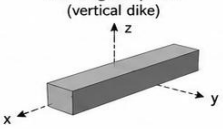
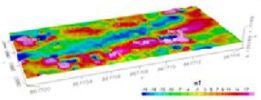
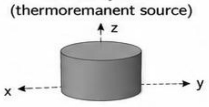
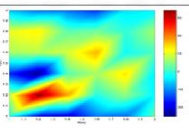
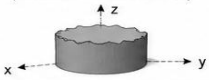
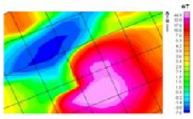
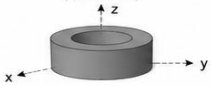
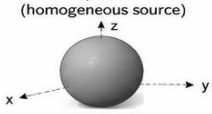
Archaeological Class	Excavation Evidence (photograph)	Representative Field Magnetic Anomaly (nT)	Dominant Magnetic Source	Forward Model Geometry (idealized shape)
(a) Limestone Wall			Induced magnetization (susceptibility contrast)	Rectangular prism (vertical dike) 
(b) Funerary Incineration			Thermoremanent magnetization	Vertical cylinder (thermoremanent source) 
(c) Palaeohearth			Thermoremanent magnetization	Irregular cylinder (thermoremanent source) 
(d) Ceramic Kiln			Strong thermoremanent magnetization	Hollow cylinder (kiln walls) 
(e) Granite Ball		Magnetic susceptibility $\sim 92.1 \times 10^{-5}$ SI (induced magnetization)	High magnetic susceptibility (induced magnetization)	Sphere (homogeneous source) 

Figure 2. Representative archaeological magnetic signatures forming the first version of the Magnetic Signature Database (MSD): (a) Limestone wall in the archaeological site of Hergla (Tunisia); (b) Funerary incinerations in the necropolis of Richeaume XIII; (c) Plaeohearth in the Lazaret Cave in Nice (France); (d) Ceramic Kiln in the Cordouls in Puylaurens (France) and (e) Granite ball (catapult) in the museum of Carthage (Tunisia).

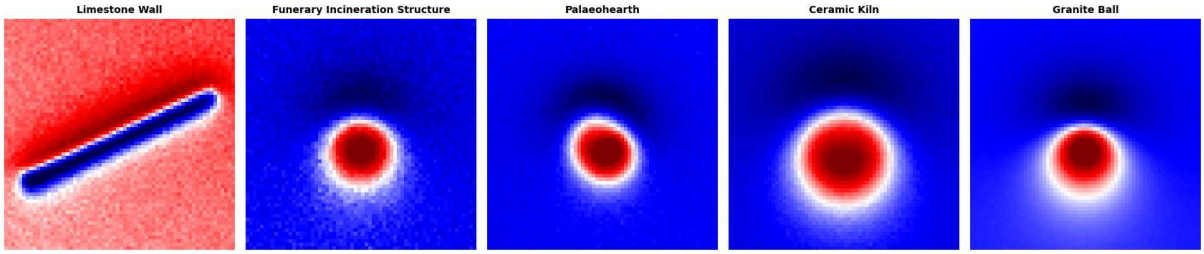
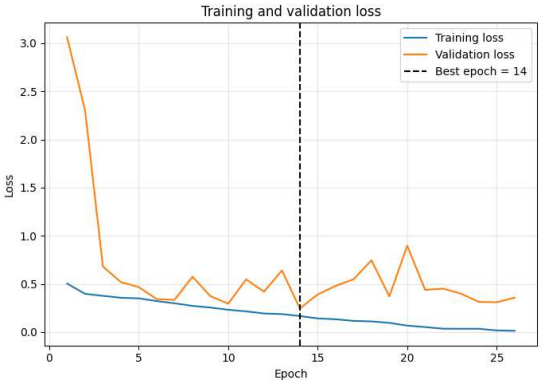


Figure 3. Class-specific forward modelling.

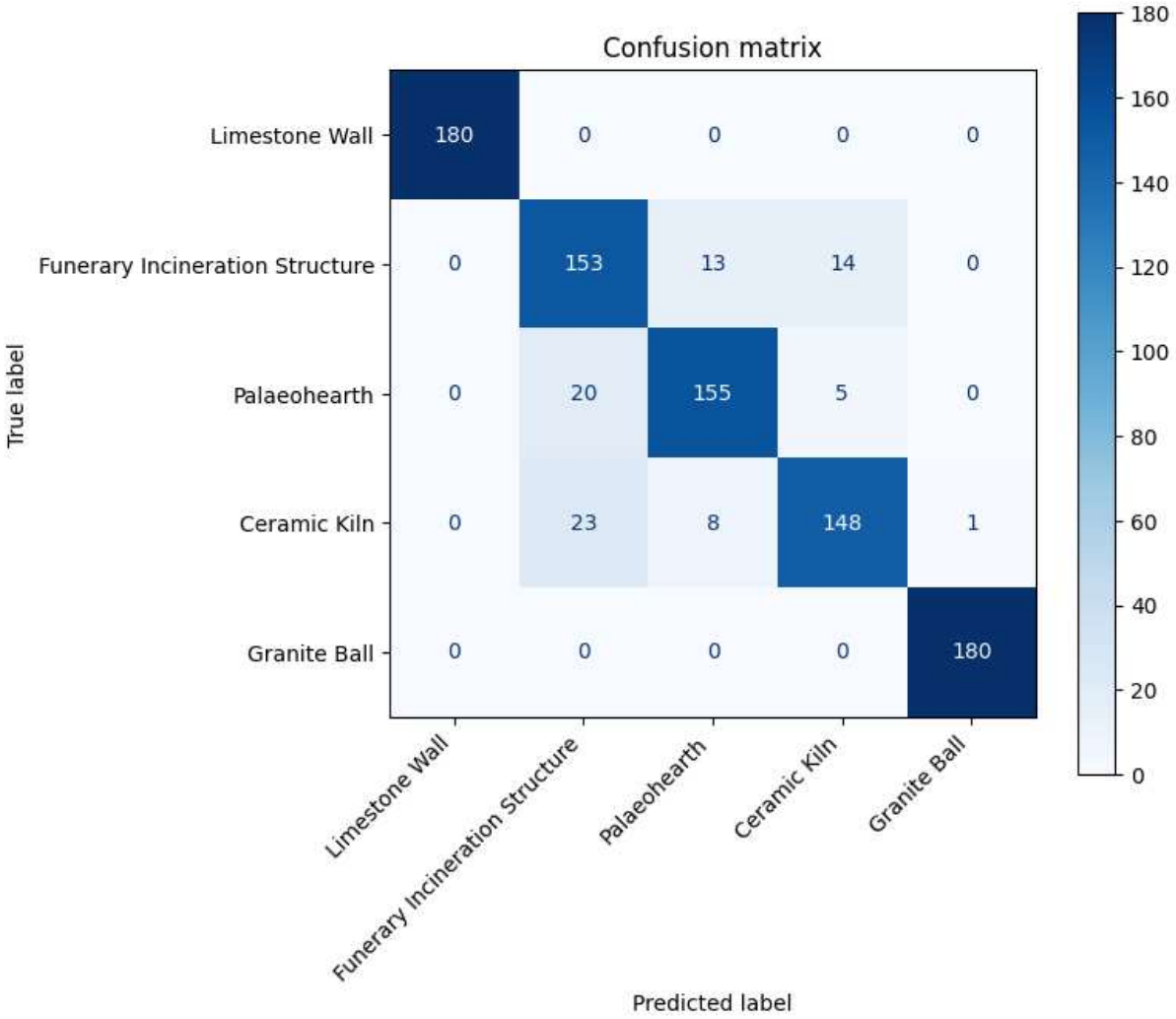
(a)



(b)



771 Figure 4. Training curves



772

773 **Figure 5.** Confusion matrix

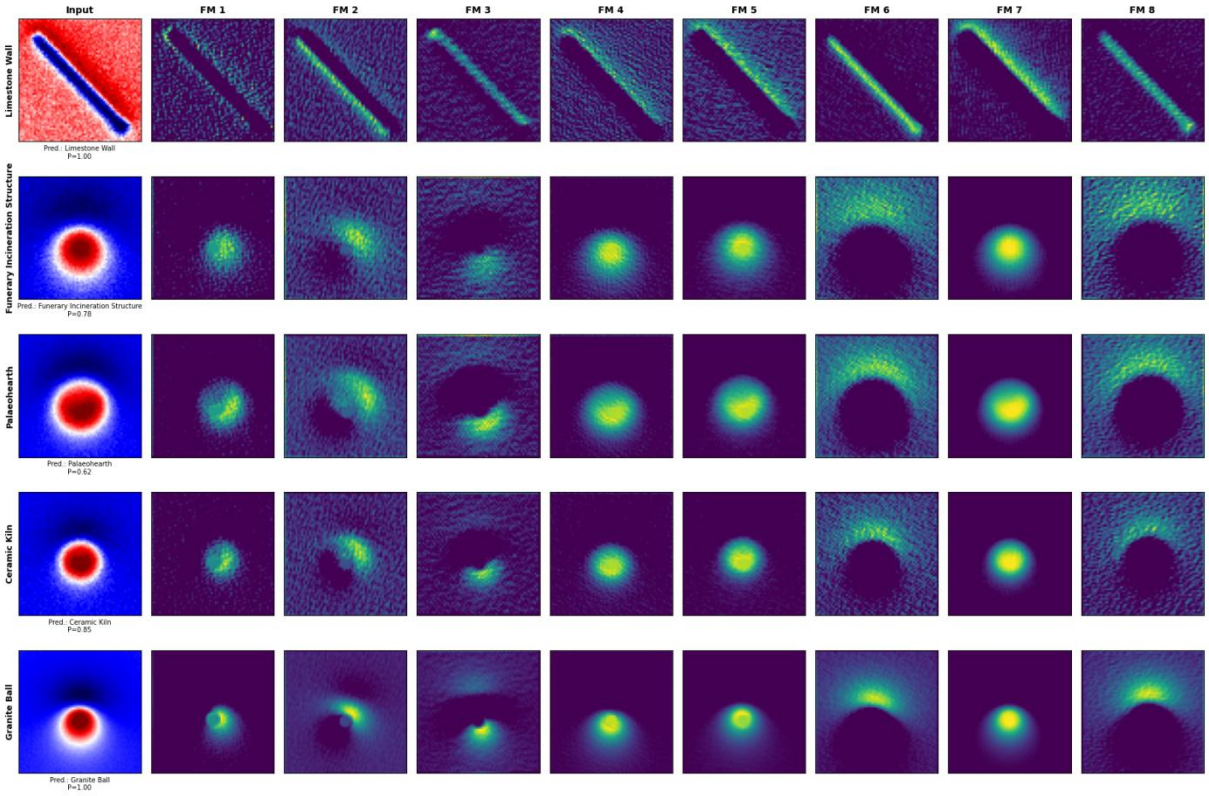


Figure 6. Feature maps extracted from the first convolutional layer of the CNN. The input magnetic anomaly (left) is progressively decomposed into multiple feature maps (FM1–FM8)

Table 1. Classification performance of the Digital Geophysical Eye for representative archaeological magnetic signature classes. Precision, recall and F1-score were computed using the evaluation metrics defined in Eqs. (9), (10) and (11).

Archaeological Class	Precision	Recall	F1-score
Limestone Wall	1.000000	1.000000	1.000000
Funerary Incineration Structure	0.780612	0.850000	0.813830
Palaeohearth	0.880682	0.861111	0.870787
Ceramic Kiln	0.886228	0.822222	0.853026
Granite Ball	0.994475	1.000000	0.997230