Rock Art Classification Through Privacy-Guaranteed Ensemble Machine Learning

Ovidiu-Emilian Buțiu[1]

Artificial Intelligence and Distributed Computing
West University of Timișoara
Bulevardul Vasile Pârvan nr. 4, Timişoara Timiş România 300223
ovidiu.butiu01@e-uvt.ro

Abstract. The use of digital media provides important benefits in preserving, but also gaining an insight into cultural or natural heritage, such as rock art practices in the past and the present. With this digitized data, a constant need for assisting AI software in speeding up the monotonous parts of deciphering rock art is created. This paper reviews the current state and the existing problems in the field of rock arts, which is in the domain of digital heritage. The discovered main issues are the high variety and high volume of data, which goes unused due to various circumstances. We compare different approaches involving individual models like Faster R-CNN and ensemble frameworks containing variants of the architectures Resnet, Resnext, and Densenet with voting systems, like PATE. We experiment with the most promising find PATE, which offers a privacy-assured ensemble of models, each one of them trained using local data. The conclusion gained after the analysis of various papers and the implementation of the closest approach to fulfilling our needs, PATE, is under-performing with an average accuracy of around 10-15% We believe the main cause is the small and unbalanced dataset, which could be further artificially increased in future research through GANs and transformations.

Keywords: Rock Art · Classification · Resnext64 · Resnet18 · Densenet201 · Voting Mechanism · Ensemble ML · PATE · Faster R-CNN.

1 Introduction

Rock art is currently studied around the world, with almost every country providing research sites of the past. These sites can be of different context types, a few discovered examples being trophy rooms, and shelters used for initiation rituals and day-to-day activities [12]. They also vary in visual depiction due to cultural differences [5], outsider and environmental damage[12], or used tools, all of these mainly depending on the region. Many teams are looking into these shelters, trying to uncover their rich past, each with huge volumes of data to analyze and interpret.

Currently, the largest body of evidence of the beginnings of humanity's culture is represented by prehistoric rock art. It has provided a profound influence when it comes to the beliefs and cultural conventions of consequent societies up to now. As such, it is an important part of the collective memory of humanity and also a most significant and enduring insight into our cultural evolution.

The problem we look to solve is the classification of these rock art motifs using ML algorithms. For possible solutions, we not only look into individual model approaches, but also ensemble ones and ones that provide dataset privacy based on identified challenges described in the problem formulation.

This paper aims to provide a better understanding of the apparent obstacles to the task of classifying rock art. We believe that the main challenges stem from data heterogeneity and imbalance in the proposed solutions we've gathered from other research papers. This is shown in different papers as the rock art dataset they use is smaller, usually under 200 images of motifs, and tends to feature an imbalanced, yet varied amount of simple and complex symbols.

Through our exploratory research, the gained insight could stand as the foundation of various future improvements for both image classification of heterogeneous datasets, and for the development of AI software assisting in rock art classification. With the use of the PATE framework, there is the apparent possibility of also establishing a distributed ensemble of various privately trained models, which could be used cooperatively by researchers in the competitive domain of rock arts, where a lack of trust among peers is common, caused by fear of having ideas, or even personal unpublished work stolen, as an already published paper is considered the only way to prove one's ownership in the fight against plagiarism. This leads us to form the following research questions:

- What models used in image classification tasks obtain a satisfying performance for rock art?
- What are the encountered difficulties when dealing with rock art motifs?
- How much does using an ensemble of ML models instead of individual models impact performance?
- Would a mix of different models be more beneficial to fit the student model in our approach, or would a homogeneous ensemble provide better results?

For the remainder of the paper, we look deeper into the problem we are focusing on, the state-of-the-art concluded from related works, followed by describing our contribution. Afterwards, we describe possible ethical concerns. We follow with the experiments taken to answer the offered questions, prove the efficiency of our solution, and provide a thorough comparison with the prior researched similar approaches to our problem. Finally, we offer a conclusion reiterating the problem, alongside a short summary of our solution and its achieved results. A short list of future directions which we plan to pursue next is also provided at the end.

2 Problem Formulation

The main problem of rock art classification is having to go through this large amount of digital data with the tasks of identifying and finding the context for each apparent motif, which results in an overall very repetitive and tedious loop. The proposed solution for this classification problem is the use of ML algorithms. This approach saves a small amount of time for each identified motif, which could range from one or two to tens for every image. In turn, it would overall save a generous amount of time spent on identifying lots of data.

At the same time, the proposed solution must deal with the difficulties caused by data heterogeneity, as motifs can range from simple shapes to complex and varied depictions. This is noted by another similar approach [6], which experienced poor results when dealing with motifs depicting people, due to their heterogeneous nature.

This points us towards the use of a varied dataset alongside ensemble ML, which offers the possibility of using varied models together, each trained on a subset of data, which in our case would be the images split into countries of origin. This would also mean a better performance than simply using a generic ML algorithm. The problem with this approach is that it necessitates a privacy guarantee for the datasets used in training and testing, due to either the lack of

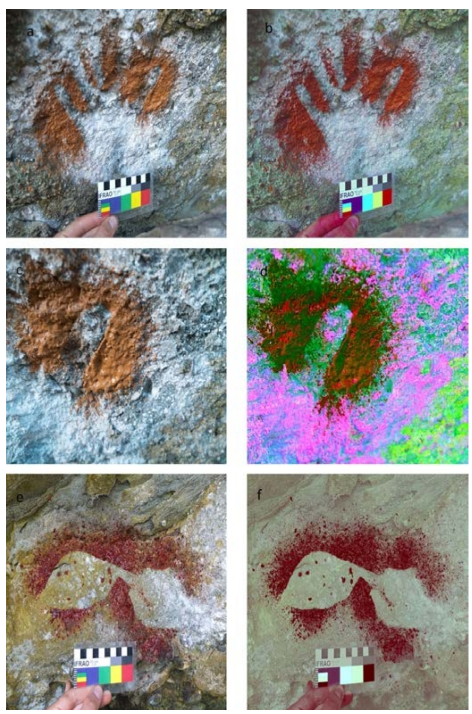

Fig. 1. Examples of various stencil types discovered at Apuranga site in 2018. *"Left down (a, c and e) before enhancement: a hand/palm, a thumb, and unknown leaf stencils, respectively, with the appearance of being recently made. Right down (b, d, and f) after picture enhancement using DStretch (yre, crgb, yre color filters, respectively) shows no older art (original photos: William Pleiber, Papuan Past Project)."* [12]

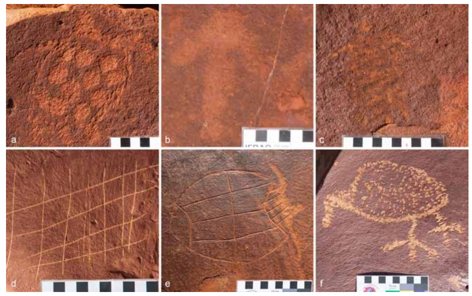

Fig. 2. Examples of various motifs, some harder to spot the shape of, created using different techniques: *"(a) pecked turtle showing two differently sized impact shapes ; (b) pounded bird track motif, probably produced with a rounded hammerstone; (c) abraded turtle motif; (d) scratched grid; (e) incised and abraded turtle; (f) pecked and gouged bird."* [5]

connections and trust between researchers, or legal or ethical constraints, which makes procuring and using data difficult.

To provide an ensemble ML approach with a privacy guarantee, which would allow the usage of private data in training and testing, we look towards Private Aggregation of Teacher Ensembles (PATE) [10]. This framework transfers knowledge from an ensemble of teacher models, trained on partitions of the data, to a student model. This protects the privacy of the training data used during learning and also offers the freedom of picking any suitable learning technique for data to use for the teachers.

At the same time, partitioning the data avoids overlapping, which produces a variety of models that independently predict labels. PATE also offers easy scalability, allowing various devices, each with its models trained on private data as stated previously, to work together as a bigger, better-performing model. This feature can offer cooperation between various research teams, which can now not only keep their data private but also assist each other by adding a model trained on their dataset, possibly improving performance and preventing biases from forming as more varied data is added overall.

Our goal is to achieve a satisfying performance in motif detection and classification, which is why we look into ensemble ML, specifically the PATE framework's potential to be used with various models to improve performance in identifying simple and complex motifs when dealing with rock art images from various countries. For that, we must also look into how well models perform when it comes to rock art classification. As PATE allows us to fit a student model using sensitive data, there is a need to apply noise while training, which may heavily impact performance, which is why a baseline would be useful using the teacher model performance. This would also provide insight for improvements and allow comparison of the performance of the trained student model.

3 Theoretical Background

According to Robert Bednarick [3], rock art is defined as human-made markings on natural rock surfaces with the use of additives (pictograms by applying material) or through a reductive process (also defined as petroglyphs, created through the removal of rock surface). The former regards rock paintings, pigment drawings, stencils, and beeswax figures, while the latter term may cover engravings, percussion petroglyphs, and finger flutings. These markings occur in nearly all countries, although in an uneven distribution, which can be attributed to cultural convention differences and a result of preservation bias (taphonomic attribute). An example of rock art can be seen in Figure 1, which showcases stencils discovered at Apruanga site in the year 2018, depicting a hand, a thumb and respectively leaves. Another example can be seen in Figure 2, which instead depicts different techniques, including carving, used to display objects and animals.

4 Methodology

The papers selected in this report have been mainly selected using research platforms such as Arxiv and ResearchGate, or through the searching tool Google Scholar.

The keywords used for searching for these papers are *ensemble ml*, *federated learning*, *rock art classification*, *rock art taxonomy*, *heterogeneous*, *PATE*, *rock art ml*, *privacy ml*, *individualized privacy ml*, *digital heritage ml*, *ml security*. The focus was on gathering papers on various ensemble ML models to be compared when used in the taxonomy of rock art, but this required more information on said domain. Thus, we also gathered some keywords to look into more information about the domain of digital heritage and more specifically the classification of rock art. We prioritized the context when it came to filtering the found ML approaches, as we wanted to look into what has already been tried for the task of rock art classification. If that was not the case, like for the ensemble ML approaches, we simply focused on the most recent and high-impact work to be used for this topic.

The datasets we decided to use for these experiments consist of images from various public sites in the countries US, Spain, Italy, Bulgaria, and Armenia. These count up to a total of 257 images.

Table 1. Statistics of the private rock art used data

Country	*Count*	*Train Split*	*Test Split*	*Person*	*Goat*	*Stag*	*Circle*	*Spiral*	*Zigzag*	*Cross*
Armenia	106	74	32	91	132	18	53	9	27	10
Italy	47	33	14	83	11	50	39	3	3	4
US	91	64	27	108	44	4	104	58	66	31
Total	244	171	73	282	187	72	196	70	96	45

5 State of the art

5.1 Rock Art

Most of the papers found featuring the use of machine learning methods that looked into the topic of rock art focused on image classification, similar to our problem. At the same time, the ones focused on ensemble ML used datasets that can be considered highly different in characteristics from the rock art dataset we used in our experiment. Thus, the presented work only offers some possibly useful approaches when it comes to ensemble ML, but does not provide enough insight for our case. We expect worse results than the ones described in this subsection.

We begin with a research paper delving into the identification and classification using Faster R-CNN of 3D rock art data processed into images [6]. Although this ensemble ML approach achieved a mAP of 32.47%, its results on the simple common labels such as boats were promising, achieving over 60% in precision for said class. This can be considered a minimal score we aim to achieve with the models we use for classification.

The rest of the objects suffered from under-representation and thus were harder to conclude results from. An important noticed issue to be aware of in our experiment was the inability to capture all objects in the data with their predictions, which may stem from the need to adjust the IoU threshold. This was left for future research as it was outside of the paper's scope and required further evaluation by setting the value lower than the one used in the experiment, which was 0.7, looking into the possibility of a relation between lower threshold values and higher recall values.

Visual inspection of the results also showed clear difficulty when it came to the model predicting larger objects with one single bounding box. As they averaged overlapping bounding boxes of the same class, the blame may stem from the bounding boxes not overlapping to a sufficient degree and causing misalignment with the motif during the box's plotting.

Also, the worst overall performing object class seemed to be "animal", mainly because of the considerable variation, possibly combined with the limited number of training examples. This could also be the case for other classes such as "human", that suffered from less performance compared to object classes representing much simpler figures, like "boat", or "circle", since in rock art research, humans or other anthropomorphic figures tend to come in combinations of bodily features and associated objects, causing them to be complex and varied. We should be aware of this possible issue, as our labels feature human and animal figures.

We follow with another recent study on Australian rock art which use the models VGG, Inception, and ResNet in their approach [7]. These models achieve very promising accuracy and F1 scores of around 80-90%. Although this research does not focus on the classification of motifs, more explicitly just rock art detection, it proves that the challenge lies solely in identifying the different represented shapes.

These papers showcase overall promising results achieved in classifying rock art for a country, whether only for specific types of motifs or just simply for detecting possible motifs in an image, but a lack of focus on the bigger apparent issues when it comes to dealing with rock art. These issues are cultural diversity and lack of data privacy, as stated previously.

5.2 Ensemble ML

Next, we looked into what approaches deal with the problem of data heterogeneity, as we suffer the same challenge from the cultural diversity of rock art. This challenge stems from the fact that motif depiction may vary depending on region, which can cause data overfitting and a bias towards specific depictions of certain cultures. At the same time, the lack of data privacy is a big issue for researchers, as the competitive environment of rock art research means no trust in sharing datasets that may be used in the development of assisting tools for classification.

In our search, there was a lack of ensemble approaches that fit our topic. As a result, we looked into the topic of medicine, which uses datasets that share similar challenges to the ones we discuss about rock art datasets. One such trait is data heterogeneity, as some diseases can be represented by varied shapes that can be either simple or complex. Another trait is the need to protect the privacy of the dataset, as datasets from this topic may contain patient info that needs to be kept confidential. This is similar to the requirement to keep rock art datasets private, to allow the use of private datasets in training and testing.

Another found paper attempts to deal with skin lesion classification, which has similar traits to the data we are working with, as they are heterogeneous, varying in size, color, shape, and complexity [11], through an ensemble learning approach which combines three deep convolutional neural network (DCNN) architectures known as Inception V3, Inception ResNet V2 and DenseNet 201. This is done to improve overall performance compared to a single DL model approach, producing promising classification performance, resulting in 97.23% accuracy, 90.12% sensitivity, 97.73% specificity, 82.01% precision, and 85.01% F1-Score.

Assiri et al. [2] proposed an algorithm for breast tumor classification using a voting mechanism, first selecting the 3 best-performing models out of 8 evaluated classifiers, based on the F3 score due to the importance of false negatives for this subject. The following winners, simple logistic regression learning, support vector machine learning with stochastic gradient descent optimization, and multilayer perceptron network, are then used for ensemble classification using various voting mechanisms. This results in the majority-based voting mechanism achieving an accuracy of 99.42% on the publicly available Wisconsin Breast Cancer Dataset (WBCD).

Kang et al. [9] focused on a different approach for the classification of brain tumors, which once again could share similar traits to our data due to their unpredictable and complex nature, by adopting the concept of transfer learning. CNNs are used to extract deep features from MRIs, which are then evaluated by

ML classifiers. The top three best performing are selected, concatenated then fed once again into ML classifiers as an ensemble of deep features to predict the final outcome. After testing on three different datasets, this method proves promising for overcoming limitations of a single CNN model, the best performance being achieved by support vector machine (SVM) with radial basis function (RBF) kernel, especially on larger datasets, with the highest accuracy reached on each dataset being between 93-98%.

This related work provides some solutions that could work if implemented for the domain of rock art, but even so, there is still one previously mentioned issue that we have yet to go through, which is the inability to use sensitive data. The need for privacy guarantees also can happen in the domain of medicine, as patient data privacy limits the available data to be used for training and testing models.[10] Also, due to mainly focusing on only a specific type of affliction, these works did not have to account for high data heterogeneity, similar to the rock art research papers only focusing on one country.

Compared to these related works, we plan on using the PATE framework. This offers the possibility of training teacher models that can be on different computers, each with its private dataset, to work together through a voting mechanism and with applied noise to train a student model. This results in the protection of the data used by the teachers to train the student.

This approach's design is also adaptable and offers scalability, allowing any model to be added to the network as long as they are properly configured, meaning there is the possibility of compensating for the weaknesses of certain models through this ensemble. PATE offers a voting system that evaluates each teacher model's response and trains the student model using selected queries for which an overwhelming consensus was reached, passing over a threshold T.

With this method, we aim for promising results in improving rock art classification accuracy by offering the possibility of using or adding sensitive data to the mix to combat the lack of public data. The framework presents the possibility of using sensitive data and features of scalability and adaptability, meaning a path to cooperation between rock art researchers and ML can be paved.

If the resulting student model performs well, the PATE framework also offers us the possibility of further research into the side of data privacy, with the possible implementation of individualized privacy [4]. This approach may further improve the student model's performance, as the teacher models would be able to pick from a range of noise levels, depending on the required strictness for the privacy budget of the used dataset. There are two existent mechanisms for this approach, the addition of either achieving the aforementioned task of providing individualized privacy.

Being allowed to use various levels of noise, instead of the highest required by the dataset, would in theory mean a better performance. One problem is that they require an entirely different way to evaluate from the standard PATE, which is already challenging enough due to the lack of available implementations to use for the framework's appraisal.

The reason as to why an entirely different method of evaluation is necessary is caused by it has to be done for particular data points or groups of data points separately instead of the entire dataset. Even so, they may be worth pursuing for this possible performance improvement, as their individualized variant of the framework showed better results when tested against the original implementation of PATE [4].

For our experiment, using the standard framework is satisfactory, as we simply aim to showcase the promise of using the ensemble approach to achieve solid results when it comes to rock art classification. Pursuing this approach of individualized privacy is better left for future works, or if proven to be required for the domain of rock art, due to necessary high differences in the levels of privacy between datasets.

6 Experiment

6.1 Research

Before our experiment began, we studied an example providing a PATE implementation in PyTorch[8] based on a research paper studying it as a generally applicable approach to providing strong privacy guarantees for training data [10]. The implementation showcases the use of this framework for the problem of image classification on the MNIST dataset, which would mean that we have to swap the model they use with the ones we plan to study and also implement a custom class to be able to work on our datasets.

Thus, we concluded that we are limited to using PyTorch, as the best approach would be to use the same packages used by Syft, the library providing a black box implementation through which we can analyze the performance of PATE in our experiments. This would ensure that as long as the models are functional and the experiment results are promising, testing these models using PATE would consist of simply replacing the models and datasets in one of its various PyTorch implementations.

The specific models that were planned to be used in this experiment were Resnet18, Resnext64, and DenseNet201, as these were the closest available choices provided by our used technologies to the similar work using ensemble ML in the task of classification.

6.2 Setup

The code used in our experiment is made available on GitHub.[1] As a start, we aimed to provide a performance comparison of each of the selected models. We used Anaconda to set up an environment running Python 3.7, which was required to be able to install and import the packages that would be used in this research. Most of the setup consisted of installing an older version of the Syft library, downloadable thanks to an example of code with PATE as its focus

[1] https://github.com/ovybe/paterockartsota/

subject [1], due to the latest version having been remade and not containing a PATE implementation at the time this experiment happened. For the dataset, we managed to

Overall, our experiment used the technologies: PyTorch 1.1.0, TorchVision 0.3.0, cudatoolkit 9.0, protobuf 3.20.1, matplotlib, and Pillow (< 7). The used implementations are run on a setup consisting of a desktop featuring an RTX 3060 ti equipped with 4,864 CUDA cores and 152 Tensor cores, an Intel® Core™ i9-11900F Processor, and 32GB of RAM DDR4 with a speed of 3200 MHz.

We separately trained the selected models, one instance for each country dataset on a newer version of PyTorch, which offered superior training times, then saved them as older variants so we could load them into the older version of PyTorch we're using for the implementation of the PATE experiment. This meant we had 9 saved model instances that could eventually be used to train the student model.

For the PATE implementation, we first loaded the US dataset, which we decided to reuse for the experiment, and then applied Laplace Distribution on it, to keep data confidentiality. We then used the teacher models to go through the dataset and vote on the predicted class. The possible predicted classes featured in the dataset were Person, Goat, Stag, Circle, Spiral, Zigzag, and Cross. This would provide the dataset used by the student model, validated with the help of the teacher models. All that is left for the experiment is to train the student model using the aforementioned dataset and provide its evaluation results.

For data pre-processing, we resized the images to 180 by 180, mainly to provide faster training and no size differences between them, and also used normalize with a sequence of means: 0.485, 0.456, 0.406 and standard deviations: 0.229, 0.224, 0.225.

We used a mean of 0, a scale of 0.1, and ϵ of 0.2 to apply the Laplace noise distribution. We used Stochastic Gradient Descent for optimization with a learning rate of $0.001 * 2.82 \approx 0.00282$ and momentum of 0.9. For the loss function, we used cross-entropy. The batch size used for training was 4 and the student model was trained for 100 epochs.

The full experiment consists of training a student model with the same model as the teacher ensemble, first formed homogeneously for each country from the selected models, then with an ensemble of the best-performing individual model for each country. In our case, we used three teachers, one for each dataset we had assembled.

The used teacher models all have an average accuracy of around $0.5-0.6$ with a loss between 1 and 2. These were trained using PyTorch version 2.5.1, with similar data pre-processing applied to the procured country datasets. We trained an instance of each model on three built datasets, respectively US, Armenia, and one combined from the data of the remaining countries.

6.3 Achieved Results

The homogeneous ensembles with this setup achieved around $0.1 - 0.3$ accuracy and 2 loss. We believe this is mainly caused by the lack of data for this experiment, as the datasets we procured are small and unbalanced, as seen in Table 1, and PATE is known to perform better on larger datasets.

For the Densenet201 homogeneous ensemble, based on Figure 3, training loss appears to fluctuate between 1.96 and 2.02, averaging around 1.99. Validation loss ranges from 2.00 to 2.06, with an estimated average of 2.03. Training accuracy seems steady at around 0.14. Validation accuracy is highly noisy but fluctuates between 0.08 and 0.16, averaging around 0.12.

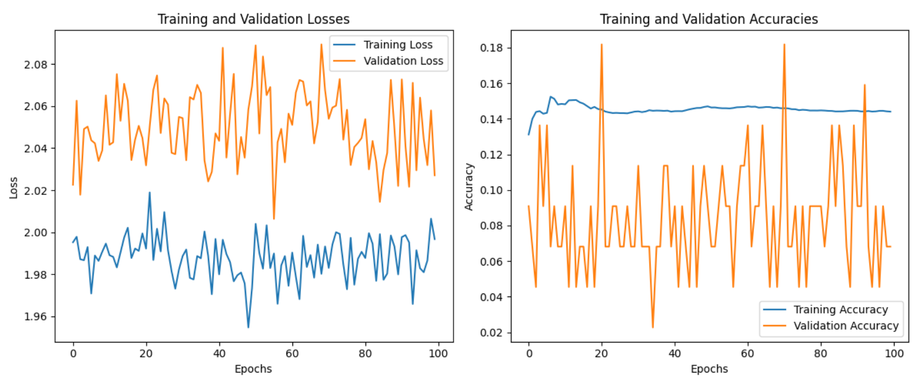

Fig. 3. A plot showcasing the training and validation losses and accuracy for the student model trained by the Densenet201 Ensemble on the Laplace distributed US dataset.

In Figure 4 we can see that the training loss for the Resnet18 homogeneous ensemble fluctuates between 2.17 and 2.23, averaging around 2.20. Validation loss shows significant variation between 2.15 and 2.30, with an average of about 2.22. Training accuracy stabilizes at around 0.10, while validation accuracy varies but averages close to 0.08.

For the Resnext64 homogeneous ensemble, as seen in Figure 5, training loss is around 2.0, while validation loss varies significantly between 2.0 and 3.4, averaging around 2.7. Training accuracy stabilizes near 0.10, while validation accuracy shows high variability (between 0.10 and 0.30) but averages around 0.15.

We believe artificially increasing the dataset through transformation could significantly improve results. This could be used alongside a GAN to procure a similar output to the datasets we used to provide even more data and to achieve a balance between the classes. This could result in at least a more promising performance while protecting the privacy of the existent dataset.

We could also use this approach for the teacher models, to further improve their performance, as they play a key role in training the student model on the noisy dataset. This could result in better training and further improvements to student model performance.

7 Conclusion

In conclusion, we showcased our current research progress and gained insights on the current state of the art when it comes to the task of rock art classification. We've also showcased the existent limitations of our approach aiming to make use of the PATE framework for its privacy guarantee, which allows the use of sensitive data in training.

We believe the main reason for the low achieved results to be caused by the small size of a dataset featuring classes that can vary from simple to complex. This, alongside the classes differing depending on country of origin, results in the student model under-performing in comparison to the individual teacher models, which are each trained and validated in only one country.

Fig. 4. A plot showcasing the training and validation losses and accuracy for the student model trained by the Resnet18 Ensemble on the Laplace distributed US dataset.

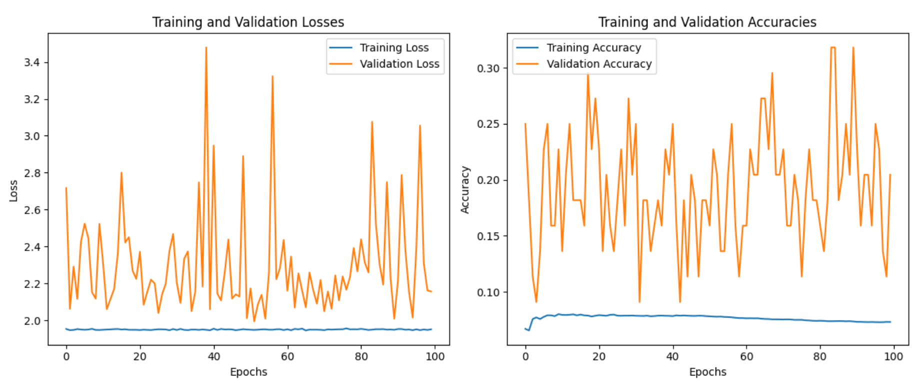

Fig. 5. A plot showcasing the training and validation losses and accuracy for the student model trained by the Resnext64 Ensemble on the Laplace distributed US dataset.

A good possible solution to these problems would be artificially increasing the dataset size through further transformations and GAN. These require more complex research and experimenting. We use both rock art paintings and carvings, which may be differently affected by more varied and complex transformations.

The same issue could arise with GANs when it comes to classes, as we would have to generate data based on the existent sets. Adding the generated data to the training would likely improve the models' performance. Unfortunately, as it is synthetic, the resulting images could be unrealistic and cause bias, as we lack a large enough dataset to assure the robustness of the artificial output.

Even so, we believe using a voting system mechanism formed by an ensemble of models, whether used to train a student model or for data prediction is a promising path to continue researching for the task of rock art classification. This is due to allowing the mitigation of some of the most challenging aspects of rock art datasets. One such aspect is its heterogeneous data based on the culture of the region of origin.

Some ensemble ML frameworks also offer a possible solution for the apparent lack of images causing class unbalance and their resulting small size. As this data can be usually considered by researchers as sensitive until their research is published, it could allow them to train teacher models privately and simply offer the resulting trained instance to be used in a framework such as PATE, which could allow the creation of a larger, better-performing model through cooperation, if enough data is procured. This would mean a more general model, that could be used to identify data from various countries and regions.

References

1. Aristizábal, A.: Making PATE Bidirectionally Private (2019), https://github.com/aristizabal95/Making-PATE-Bidirectionally-Private
2. Assiri, A.S., Nazir, S., Velastin, S.A.: Breast tumor classification using an ensemble machine learning method. Journal of Imaging **6**(6) (2020). https://doi.org/10.3390/jimaging6060039, https://www.mdpi.com/2313-433X/6/6/39
3. Bednarik, R.G.: Rock art science. Brepols (2001)
4. Boenisch, F., Mühl, C., Rinberg, R., Ihrig, J., Dziedzic, A.: Individualized pate: Differentially private machine learning with individual privacy guarantees (2022)
5. Domingo Sanz, I., May, S., Smith, C.: Communicating through rock art : an ethnoarchaeological perspective (02 2016). https://doi.org/10.13140/RG.2.1.4539.6244
6. Horn, C., Ivarsson, O., Lindhé, C., Potter, R., Green, A., Ling, J.: Artificial intelligence, 3d documentation, and rock art—approaching and reflecting on the automation of identification and classification of rock art images. Journal of Archaeological Method and Theory **29** (03 2022). https://doi.org/10.1007/s10816-021-09518-6
7. Jalandoni, A., Zhang, Y., Zaidi, N.A.: On the use of machine learning methods in rock art research with application to automatic painted rock art identification. Journal of Archaeological Science **144**, 105629 (2022). https://doi.org/https://doi.org/10.1016/j.jas.2022.105629, https://www.sciencedirect.com/science/article/pii/S0305440322000875

8. Kamath, H.: PATE Example (2019), https://github.com/kamathhrishi/PATE
9. Kang, J., Ullah, Z., Gwak, J.: Mri-based brain tumor classification using ensemble of deep features and machine learning classifiers. Sensors **21**(6) (2021). https://doi.org/10.3390/s21062222, https://www.mdpi.com/1424-8220/21/6/2222
10. Papernot, N., Abadi, M., Úlfar Erlingsson, Goodfellow, I., Talwar, K.: Semi-supervised knowledge transfer for deep learning from private training data (2017)
11. Pratiwi, R., Nurmaini, S., rini, d., Naufal Rachmatullah, M., Darmawahyuni, A.: Deep ensemble learning for skin lesions classification with convolutional neural network. IAES International Journal of Artificial Intelligence (IJ-AI) **10**, 563–570 (09 2021). https://doi.org/10.11591/ijai.v10.i3.pp563-570
12. Tsang, R., Brady, L.M., Katuk, S., Taçon, P.S., Ricaut, F.X., Leavesley, M.G.: Agency, affect and archaeologists: Transforming place with rock art in auwim, upper karawariarafundi region, east sepik, papua new guinea. Rock Art Research: The Journal of the Australian Rock Art Research Association (AURA) **38**(2), 183–194 (2021)