Determination of language families using deep learning

P. B. Lerner[1]

Abstract

We use a c-GAN (convolutional generative adversarial) neural network to analyze transliterated text fragments of extant, dead comprehensible, and one dead non-deciphered (Cypro-Minoan) language to establish linguistic affinities. The proposed approach is based on forming a two-dimensional digital "fingerprint" of the fragment. The analysis of the fingerprint employs the methods developed for the analysis of visual images. At the post-processing stage, we define several metrical distances between original and simulated fingerprints. This paper is agnostic with respect to translation and/or deciphering. However, there is hope that the proposed approach can be useful for decipherment with more sophisticated neural network techniques.

1. Introduction

The availability of the deep learning instruments made possible the types of analysis unachievable with other methods. Deep learning currently is used in LLMs (Large Language Models), for image identification, creation of deepfakes and analyses of astrophysical and financial information (Krizhevsky, 2012), (Sutskever, 2014), (Vaswani, 2017), (Wang, 2015), (George, 2018), (Li, 2010).

When the instruments of deep learning became widely available, it was decided that the decipherment of all dead languages was only a matter of time. However, the progress is slow and application specific (see the excellent review (Sommershield, 2023)). Deciphered dead languages are still the ones which have bilingual or trilingual texts available with minor exceptions.[2]

This paper implements the c-GAN-based methodology, earlier published by this author in the context of the analysis of causality relationships of the financial time series for the establishment of the language families. The significance of this method is that comprehension of the text fragments plays no role in the determination of their phonetic and grammar affinity and can be equally applied to the deciphered and undeciphered languages, if only the texts can be placed in a uniform framework. The author uses texts in their Latin transliteration, but it simply

[1] Unaffiliated, Ithaca, NY. Contact: pblerner18@gmail.com, peter.lerner@faculty.umgc.edu.

[2] S. A. Starostin, private communication to the author, c. 1987. The only exception known to the author was Ventris' decipherment of the Linear B, which was, after all, an archaic version of a well-known language (Chadwick, 1990).

reflects limitations of the author's resources. In principle, any fragment can be analyzed as long as the signs admit a uniform mapping of any kind between them.

Adversarial generative networks (GANs) are founded on the following principle. The algorithm consists of two parts: generator and discriminator (critic). The generative part of the network produces arrays from a random noise, which must satisfy some criteria of affinity, usually based on entropic considerations, to the training dataset. Critic must distinguish between a real dataset and its random noise replication through adjusting a large number of parameters (several hundred thousand in my home PC-based simulation and up to trillions for LLMs of Google, etc.). The c-GAN and other neural networks (NNs) allow a significant plethora of computational methods developed with great success for the analysis of visual information.

The author, in his previous analysis of the causality of the financial time series proposed the following protocol (Lerner, 2023). One chooses the datasets to compare as the training and test datasets, respectively. An entropic comparison measure determines whether a fake, built from a random noise by a generator of the neural network, fits better one or the other dataset. In the case of the financial time series, a dataset, which requires more information to fake it from a random noise, is considered more informative. In the case when both datasets are synchronous, a more informative dataset is declared causally determinative to a less informative dataset.

This protocol in the case of text fragments can be intuitively explained as follows. Imagine a speaker of e.g. Finnish language being asked to produce a series of nonsensical texts claimed to be English. An English-language speaker must determine whether it is nonsense or a broken English. Adversely, an English-language speaker is asked to produce series of nonsensical Finnish fragments, and they are analyzed by a Finn.

If the statistical asymmetry between recognition capabilities is large, we consider the languages as linguistically unrelated and vice versa. Obviously, the more sophisticated are the fakes, the better is an ostensible determination of linguistic affinity.

Unlike my financial time series where c-GAN algorithm produced quite well-fitted fakes having a relatively high correlation (~20%, Fig. 1) with the original data, the correlation of pre-processed linguistic samples is small. Below we provide tests to improve the qualities of the fakes but so far, statistical significance of comparisons of original language samples is small. Despite that, there is a robust reproduction of the correlation statistics in independent tests. Furthermore, the fakes created by the network algorithm demonstrate a reasonably high correlation among each other.[3]

The paper is structured as follows. In Section 2 we explain the construction of the samples. In Section 3, we provide the measures of the distance between datasets. We discuss post-processing metrics in Section 4. The measure of asymmetry between datasets built from the fragments of different languages is used for our analysis in Section 5. The test of robustness is described in Section 6. Section 7 concludes the presentation.

[3] See Section 6.

2. **Choice of the samples**

We use samples of eight languages. The samples were chosen deliberately small as to 1) be amenable for analysis on a standard PC and 2) conform to a very small existing corps of Cypro-Minoan texts, the only undeciphered language among the eight. Our dataset contains two extant Indo-European (IE) languages (English and Spanish), two extant non-IE (NIE) languages geographically unrelated to the Mediterranean-Fertile Crescent area (Tagalog and Finnish; for simplicity of transliteration), one dead IE (Luwian), and two classified non-IE Middle Eastern languages from the geographic area of Cypro-Minoan (Babylonian—Semitic, and Hurrian—language isolate) and one unclassified and undeciphered dead language—Cypro-Minoan. Finnish language was added at the end of simulations to verify the robustness of the classification scheme to adding of a new language.

Luwian was selected because it was the geographically closest IE area, which was not archaic Greek. Hurrian was selected for the following reasons. The historic period associated with the Cypro-Minoan inscriptions (1500-1150 BCE) almost exactly coincided with Kassite domination of Mesopotamia. Kassite language is not decisively attributed to any language family, but some researchers related it to Hurro-Urartian (Schneider, 2003), (Fournet, 2011).

In our study, we do not touch the problem, whether Cypro-Minoan inscriptions were written in one, or several languages (see, e.g. (Palaima, 1989)). Our chosen sample of Cypro-Minoan is too limited to address this possibility.

The choice of Tagalog and Finnish was made because of the original Latin-based writing system simplifying conversion into number lists. The English sample was selected from a finance paper (reviewed by the author) written by the Spanish-speaking authors to complicate recognition. The Spanish sample was taken from Wikipedia article 'Peres Galdos'. The Tagalog text was chosen from a Tagalog Wikipedia article on the Philippines. Further data on the sources is provided in Table 1.

Table 1

Language	**Primary source**	**Content/article**	**Size (ns/sp)***
English	Draft manuscript	-	8,356/9,936
Spanish	Wikipedia	Peres Haldos	8,019/9,563
Tagalog	Wikipedia	Philippines	8,441/9,937
Finnish	Wikipedia	Kalevala	11,714/13,187
Luwian	(Cambel, 2022)	A prayer to relieve toothache	9,065/10,387
Babylonian	(Clay, 1910)	Random selection	5,563/6,228
Hurrian	(Wegner, 2020), (Campbell, 2016)	Random selection	5,255/6,219
Cypro-Minoan**	University of Barcelona Dissertation	(Valerio, 2016)	2,070/2,371

*ns/sp – size without (ns) and with (sp) spaces in digits.

**All-numbers transiteration

At the preprocessing stage, we fit our transliterated texts into 4×64×64 arrays where text was split into four excerpts to be comparable in size with (usually shorter) excerpts from the dead languages. For preliminary investigation, a standard UTF-8 conversion routine was used.

A standard deep learning procedure consists of choosing training, tests and predictive stages (Aggrawal, 2023). Sometimes, the data undergo pre- and/or post-processing for analysis. This is, in part, because the multidimensional tensors, which are frequently the numeric output of the neural nets, are difficult to convey in a usual two-dimensional format of a text or an image. Our pre-processing stage included digitizing of the input texts and fitting individual excerpts into a standard 64×64=4196 signs list, for which some of the entries could be empty. This list has a visual form, which the author called a "fingerprint" in a financial context (Lerner, 2023). A fingerprint is a unique image, which we prepared to analyze by the c-GAN network. We choose one of the 4-tuples of a given language fingerprint as a training and another 4-tuple as a test sample.

c-GAN network is run to produce fake fingerprints from a random noise from a training file and analyzes it with a discriminator/critic. The outputs of the neural network were assigned certain affinity criteria (quasi-distances, see the next section), which we used to ascribe hypothetical relationship between languages or a lack thereof.

3. **Distances on the state space**

Because of the "black box" nature of the neural networks, their output is hard to rationalize. First, the human mind has evolved to analyze two- or three-dimensional images in three-dimensional space. Most humans cannot directly comprehend tensor inputs, intermediate results, and outputs typical for neural networks. Second, the results of neural network analyses are necessarily stochastic and depend on the large number of estimated intrinsic parameters, which are frequently inaccessible, but in any case, too numerous for the humans to rationalize. Third, deep learning results can depend on how the training and testing samples are organized, even if they represent identical datasets. All of this can indicate the failure of a deep learning procedure, but it can also show additional information we fail to recognize (Brownlee, 2021). Because neural networks are the "black boxes", instead of the interpretation of a hundred thousand—in my case, and trillions of—parameters in the case of Google and Microsoft deep learning networks, one must design numerical experiments and analyze the output from a deep learning algorithm in its entirety.

To systematize the results, we propose two measures of divergence of images as follows: After the c-GAN generated fake images ("fingerprints") of the session, we considered these images as (1) matrices and (2) nonnormalized probability distributions.

The first approach is to treat arrays as matrices (tensors). We computed the pseudo-metric cosine between the image arrays X and Y according to the following formula:

$$C_{XY} = \frac{\|X+Y\|^2 - \|X-Y\|^2}{4\|X\|\cdot\|Y\|} \tag{1}$$

In the above formula, the norm $\|\cdot\|$ is a Frobenius matrix norm representing each image array. In the first stage, we computed the distance as the average of each twentieth or sixteenth of the last

400 images in the sequence. Because, sometimes, the fake image is an empty list having a zero norm, we modified this formula according to the following prescription:

$$C_{train,fake} = \frac{\|train + fake\|^2 - \|train - fake\|^2}{4\|train\| \cdot \|test\|} \tag{2}$$

$$C_{test,fake} = \frac{\|test + fake\|^2 - \|test - fake\|^2}{4\|train\| \cdot \|test\|}$$

Equation (2) provides answers close to the correct geometric formula (2), but it does not fail in the case of an empty fake image. The pseudo-metric measure, calculated according to Equation (2), provides a fair picture of the affinity of the fake visual images to the originals, but it is still unstable with respect to different stochastic realizations of the simulated images.

From the expressions of Equation (2), we can create an entropic measure of affinity between samples and the fakes by the formula:

$$\rho = ln\left(\frac{C_{train,fake}}{C_{test,fake}}\right) \approx \frac{C_{train,fake}}{C_{test,fake}} - 1 \tag{3}$$

The second approximate formula is true for most of our correlations because the correlation ratios are close to 1.

Unlike the financial series fingerprints, for which net-generated fakes are correlated at the level of 20% with the originals and visually they are virtually indistinguishable from the originals, the correlation of linguistic images is very small (typically, 0.2-0.3%) and they are easily distinguishable from the originals on eyesight (compare Figures 1, 2 and 3). This circumstance can totally invalidate the substantive conclusions of our analysis, but the essence of the method can be outlined irrespective of the low correlation. Additional considerations and robustness checks will be discussed and implemented below.

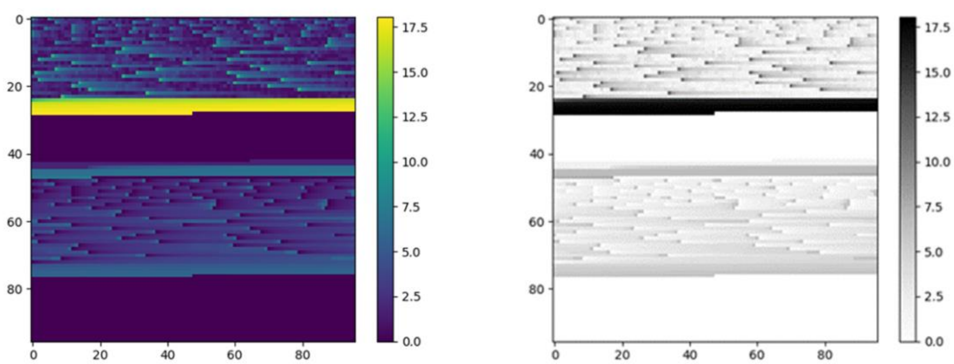
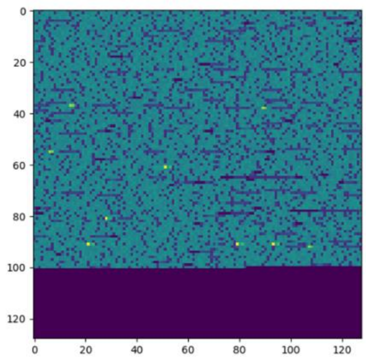
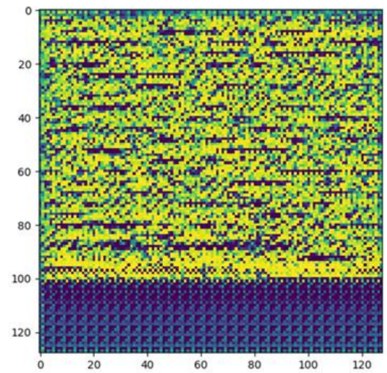


Fig. 1. The visual quality of c-GAN synthesis of the fingerprints for the financial time series (Lerner, 2023). The correlation of arrays is approximately 20%. Original data is on the left, a fake fingerprint synthesized from a random noise is on the right.





A) B)

Fig. 2. A) Fingerprint of an original English 128x128 data file. It splits into four 64×64 training files. A dark band indicates the end of the sample. B) Fingerprint of a Spanish text. Visual difference with the English fingerprint is evident.

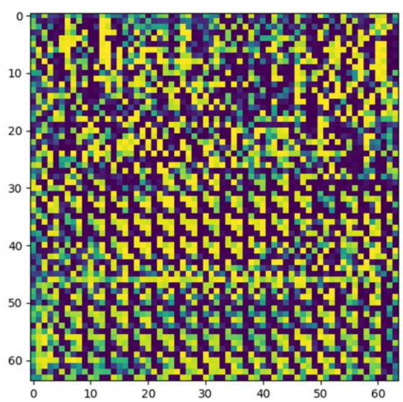


Fig. 3. The c-GAN output for a Babylonian fake fingerprint. A typical correlation with an original array is 0.2-0.3% (compare to Fig. 9).

4. **Post-processing metrics**

To analyze affinity between languages, we need distance metrics between fake images, i.e. the arrays by which the c-GAN simulates relations between symbols of an original text. For any pair of languages A and B, we analyze their relationship by the following protocol:

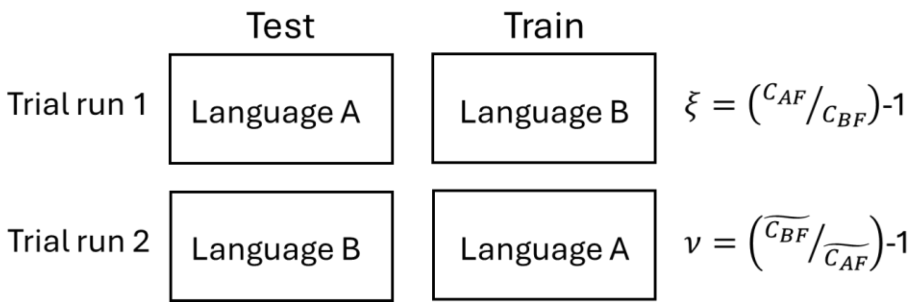


Fig. 5. The protocol of comparison of the language pair A and B. The distances between languages are expressed through Equation (4).

Obviously, this protocol is applied to every pair of languages in our sample. Classification is done by applying two sets of distances. One set is the Euclidean-style or l^2 distance.

$$d_1 = \sqrt{\frac{\xi^2 + \nu^2}{2}}$$

$$d_2 = |\xi - \nu| \tag{4a}$$

Another is the Manhattan-style (Black, 2019) or l^1 metrics:

$$\tilde{d}_1 = |\xi| + |\nu|$$

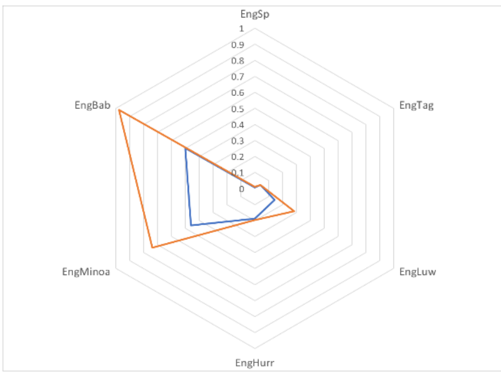
$$\tilde{d}_2 = ||\xi| - |\nu|| \tag{4b}$$

Precise functional form for d_1 and d_2 is not essential. We could, for instance, replace d_1 and d_2 by the Shannon-Kullback-Leibler type metric analogous to Equation (3) with little change for a qualitative picture. From the definition, it is clear that for the identical training and test files in the protocol, both distances would be exactly zero. The network creates fake fingerprints from a random noise. Heretofore, the fact that the fakes "inherit" this property from the original samples is not entirely trivial.

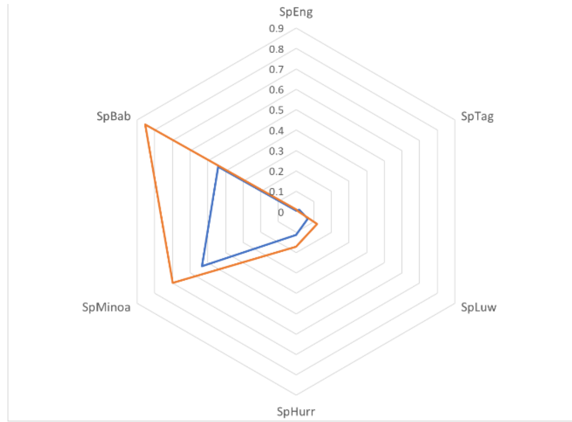
The difference between d_1 and d_2 is that while the distance d_1 measures combined distance inside the Trials 1 and 2, the distance d_2 measures the asymmetry between the Trials 1 and 2. In the next section we shall see that identification of language affinity by the two measures is very similar. We attribute it to a high symmetry of the protocol in Fig. 5.

5. **Identification of languages**

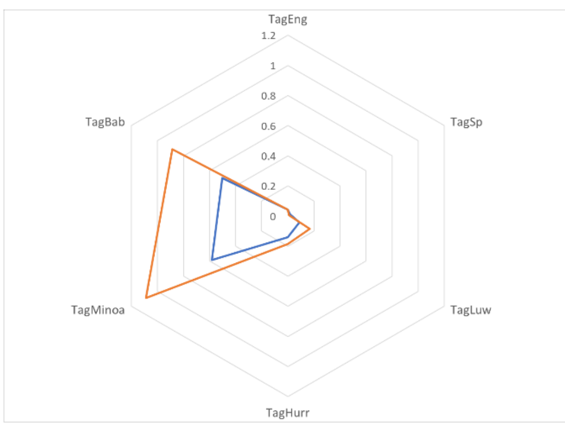
Euclidean distances between language pairs are plotted in Fig. 6. We observe that, while d_1 and d_2 are disparate in numerical values, the classification based on average asymmetry inside and between the Trials 1 and 2 is largely consistent. The results of our classification are provided in Table 2.



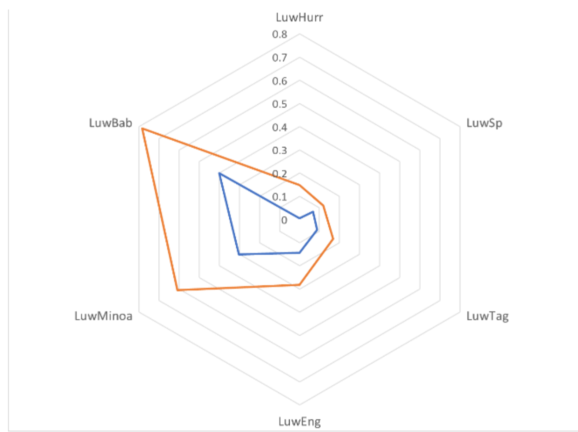
A)



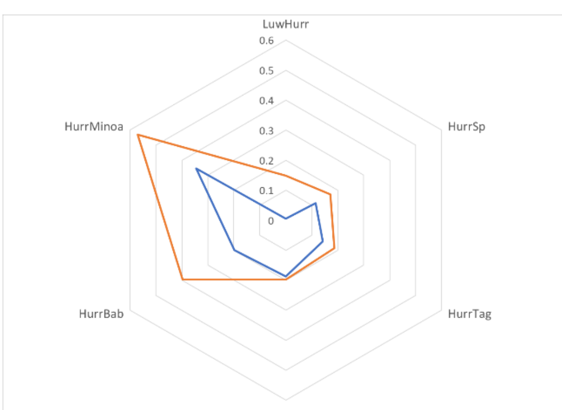
B)

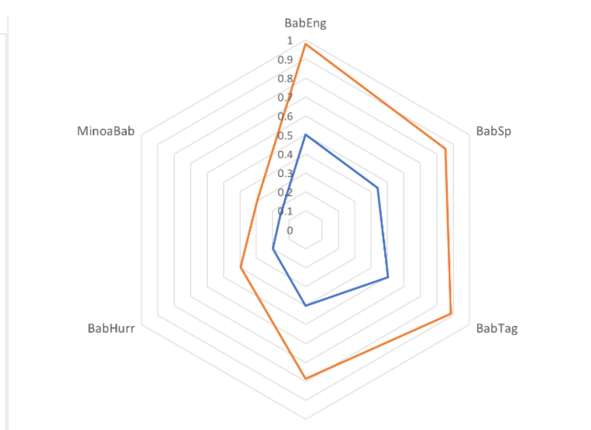


C)

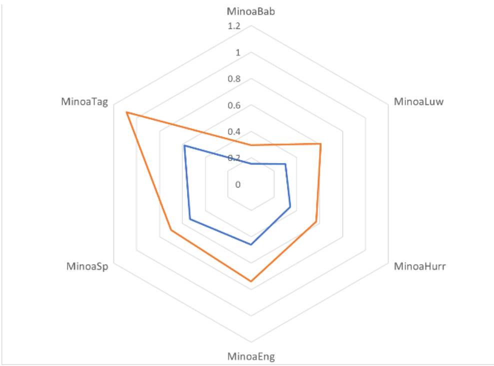


D)





E) F)



G)

Fig. 6 Radar plots of the distances d_1 and d_2 between the languages. A) English; B) Spanish; C) Tagalog; D) Luwian; E) Hurrian; F) Babylonian; G) Cypro-Minoan. Description of the samples is given in Table 2.

Table 2. Classification of the languages by the Euclidean distances d_1 and d_2.

Language	***Affinity by d_1***	***Affinity by d_2***
1. *English*	2, 3, 4, 5, 7, 6	2, 3, 5, 4, 7, 6
2. *Spanish*	1, 3, 4, 5, 7, 6	3, 1, 4, 5, 7, 6
3. *Tagalog*	2, 1, 4, 5, 6, 7	2, 1, 4, 5, 6, 7
4. *Luwian*	5, 2, 3, 1, 7, 6	2, 5, 3, 1, 7, 6
5. *Hurrian*	4, 2, 3, 1, 6, 7	4, 2, 3, 1, 6, 7
6. *Babylonian*	7, 5, 4, 2, 1, 3	7, 5, 4, 2, 1, 3
7. *Cypro-Minoan*	6, 4, 5, 1, 2, 3	6, 5, 4, 1, 2, 3

Note: Affinity to the given language, i=1,…7, is given by the string of the letters j in the order of increasing distance. The discrepancies in classification by two distances are marked in red.

Table 3. Classification of the languages by the Manhattan distances $\tilde{d}_1$ and $\tilde{d}_2$.

Language	***Affinity by d_1 from closer to more distant***			***Affinity by d_2 (ibid.)***		
1. *English*	2, 3	4, 5	7, 6	2, 3	5, 4	6, 7
2. *Spanish*	1, 3	4, 5	7, 6	3, 1	4, 5	6, 7
3. *Tagalog*	2, 1	4, 5	6, 7	2, 1	4, 5	6, 7
4. *Luwian*	2, 3	1, 5	7, 6	5, 3	1, 2	6, 7
5. *Hurrian*	2, 3	4, 1	6, 7	4, 6	2, 3	1, 7
6. *Babylonian*	7, 5	4	2, 3, 1	5, 7	4	3, 2, 1
7. *Cypro-Minoan*	6, 5	4, 1	2, 3	6, 5	3, 1	4, 2

Note: This table is not as consistent between $\tilde{d}_1$and $\tilde{d}_2$, which should not be considered as a defect. It could be that these metrics provide a better distinction between the tests inside and between the samples (see Figure 2). Classification of closest clusters is more consistent than for the Euclidean distance for most samples.

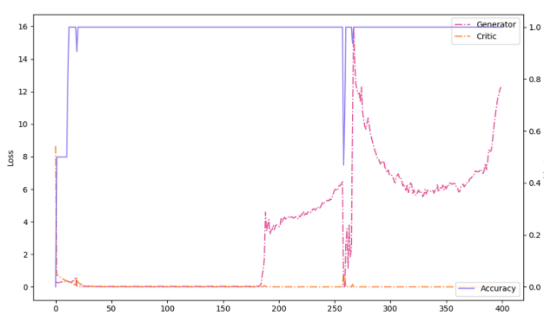
The Euclidean metric results are inconsistent with the data of comparative linguistics in only two cases.[4] First, the distance d_2 identifies Tagalog as closer to Spanish than to English. The results of Luwian sample somehow identify Luwian as closer to NIE Tagalog than to IE English but the difference in distances is small. We consider these results as remarkable given that 1) the NN does not know anything about linguistics, 2) the author is not a professional linguist, and 3) the correlations between original fingerprints and their machined fakes are very small.

Unexpectedly, the closest language to Cypro-Minoan is Babylonian, i.e. a Semitic language by three metrics out of four. The Manhattan metric $\tilde{d}_2$. identifies Hurrian and Babylonian as the closest to Cypro-Minoan with very little difference in respective distances. Note that statistical analyses are not usually designed as positive proofs. All they do is to reject some hypotheses, in our case, a closer affinity of our sample of Cypro-Minoan to modern IE and NIE languages, IE Luwian, and language-isolate Hurrian than to the Semitic Babylonian.

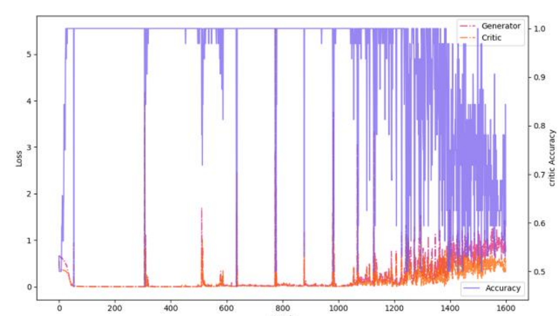
This could be an artifact of, for example, our random choice of Babylonian instead of other Semitic language. Or this could be a real hint that our sample of Cypro-Minoan texts was written in a Semitic language. Even, if this is true, it cannot be an indication what language was used by Cypriots at the time. However, the author expresses hope that the development of this method using a larger language database and a better NN algorithm can be useful for decipherment.

6. **Tests of robustness**

Comparison of the input and output images in Figs. 2, 3 and the typical loss curves in Fig. 7 demonstrates that the NN suffers from a severe mode collapse. We tried to increase correlation between a fingerprint of a real language and its computer-generated fake through several commonly applied methods.

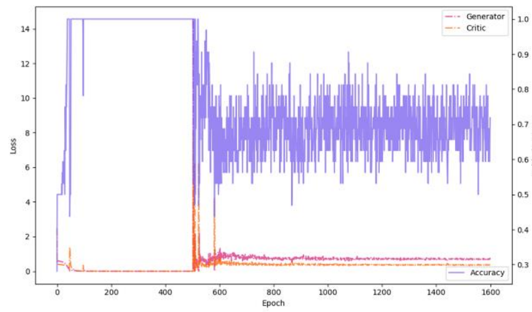


A)



B)

[4] The distances between the dead languages are generally large (see Fig. 6) so that, for instance, the closeness between Luwian and Hurrian does not necessarily indicate the same language family.



C)

Fig. 7. A) Typical output of the c-GAN for the first 400 epochs. Generator loss increases, while the critic's accuracy stays near zero. Critic does not learn cumulatively; B) Accuracy of a learning critic. The accuracy of critic grows with the number of epochs and identification of fake falls, which usually indicates improvement of the image quality. Nevertheless, the correlation between fake and original images remains low; C) The plot demonstrates that the loss can be stabilized but learning procedure demonstrates an optimum at the number of epochs between 400 and 600 and does not improve thereafter.

1) Smoothing/filtering of the fingerprint. If the image contains too many spurious details, it makes sense to produce a smoother image. We tried 1) Fourier transforming the fingerprint; 2) Gaussian filtering along the horizontal axis of the array – following a usual direction of the text before digitization; 3) Gaussian filtering along the vertical axis and 4) Gaussian filtering along both axes. Fourier transform amplifies long-distance correlations at the expense of short-distance ones, Gaussian filtering along horizontal axis amplifies correlation between syllables – assuming most of them consist of two letters, and filtering along vertical axis amplifies correlation between subsequent lines of the text. None of these methods produced an improvement according to our chosen post-processing metric.
2) Making learning of the critic cumulative. We changed the settings of the critic from unlearnable to learnable. The statistics of fake recognition changed significantly, but not the correlation of the output.
3) We applied different distance settings for the generator and the critic, for instance, binary cross entropy measure for one, and mean square error (MSE) for another. The results showed no improvement, according to our post-processing measure of Section 4 and we reverted the settings to original ones ('binary cross entropy' for both).
4) Increasing the number of simulation epochs from 1600 to 4800. No tangible improvement was noticed.

As additional tests of robustness we added modern NIE Finnish to our list (Figure 8). It demonstrated almost equal distance from all the tested languages, except the Cypro-Minoan, for which the distances were much higher than for the rest. This can be conditionally attributed to the fact that our sample of Cypro-Minoan was much smaller than all other languages. The only classification outlier with respect to the Tables 2 and 3 is an extremely small distance $\tilde{d}_2$ between Finnish and Hurrian. A measured distance with most languages was similar to the distances between non-extinct and extinct languages irrespective of their language family.

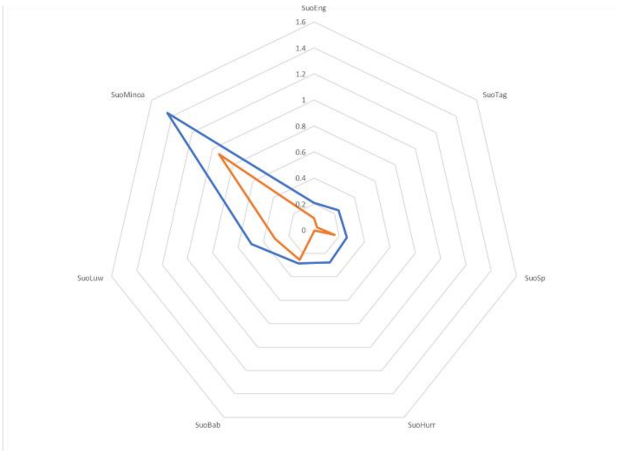
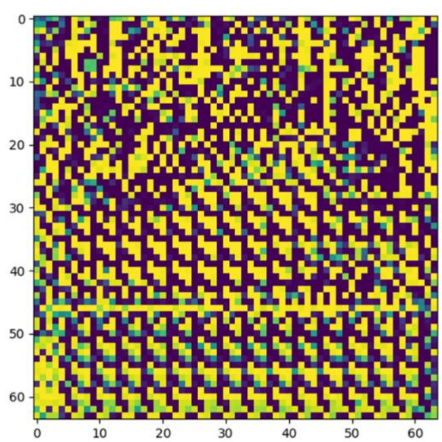


Fig. 8. Radar plot of a sample, which includes Finnish. The only classification outlier with respect to the Table 2 is an extremely small distance $\tilde{d}_2$ between Finnish and Hurrian.

Finally, we used the following procedure. We compared distances not between original samples but between fakes of the languages in question. Correlation between fakes generated from the languages in question has demonstrated a drastic improvement (Fig. 9 and Table 4). The number and length of fake samples can be increased at will, unlike the original samples of dead languages. In essence, this test is similar to the application of bootstrap procedure in traditional statistics (Efron, 1993), (Davison, 1997). However, we cannot conclude whether it signifies real possibilities of the method or is an artifact of our implementation of c-GAN network.



.

Fig. 9 A secondary fake Babylonian fingerprint obtained from a fake training sample analogous to the primary fake on Fig. 3. Visual differences are small and typical correlations are in tens of percent.

Table 4. Jackknife testing of the Minoan-Babylonian pair of languages. Training files were taken from the first-stage simulation. Correlations between fingerprints (original and secondary fakes) are listed in the table below. The correlation between Cypro-Minoan and Babylonian fakes are

high and almost symmetric with respect to exchanging between “Cypro-Minoan” and “Babylonian” samples. The only asymmetry concerned critic’s recognition of fakes.

Train / *Test direction*	*Minoan*	*Babylonian*	*Notes*
Min-Bab	0.44	0.18	Poor discrimination at the end of the epochs*
Bab-Min	0.17	0.41	100% discrimination by the critic

*Poor discrimination by the critic means high-quality fakes and vice versa. “Good” fakes must exhibit ~50% discrimination.

7. Conclusion

In this paper, the author proposes a deep learning-based approach to language classification. This approach uses the comparison of digital fingerprints (see Section 2) of language samples. We define a set of quasi-distances on the state space to quantify the affinity between languages in our sample. This method does not require that the classified text be comprehensible or even transcribed, as long as signs/characters in it are uniformly digitized.

The author very preliminarily identifies samples of Cypro-Minoan as being written in a Semitic language through the comparison of the relative distances between languages of different families. Hope is expressed that a more advanced applications of his approach can bring Cypro-Minoan inscriptions closer to decipherment (Ferrara, 2019).

References

Aggrawal, C. C. (2023). *Neural Networks and Deep Learning: A Textbook.* Heidelberg: Springer .

Black, P. E. (2019). *Manhattan distance.* (P. E. Black, Ed.) Retrieved from Dictionary of Algorithms and Data Structures.

Brownlee, J. (2021). *Machine Learning Mastery With Python: Understand Your Data, Create Accurate Models and Work Projects.* Independently published .

Cambel, J. a. (2022, 04 21). *https://archive.org/details/hieroglyphic-luwian-inscriptions*. Retrieved from Internet Archive.

Campbell, D. (2016). Hurrian šug=am ("to unite") and its context. *Journal of Near Eastern Studies, 75*(2), 339-347.

Chadwick, J. (1990). *The Decipherment of Linear B.* Cambridge UP.

Clay, A. T. (1910). *Babylonian tablets & C. in the British Museum*.

Davison, A. C. (1997). *Bootstrap Methods and Their Application (Cambridge Series in Statistical and Probabilistic Mathematics Series, Number 1).* Cambridge: Cambridge UP.

Efron, B. a. (1993). *An Introduction to Bootstrap, Monographs on Statistics and Applied Probability.* CRC Press.

Ferrara, S. (2019). *The Greatest Invention.* Farrar, Straus and Giroux.

Fournet, A. (2011). The Kassite language in a comparative perspective with Hurrian and Urartian. *Macro-Comparative Journal, 2(1)*, 1-19.

George, D. E. (2018). Deep neural networks to enable real-time multimessenger astrophysics. *Phys. Rev. D*, 044039.

Krizhevsky, A. I. (2012). Image net classification with deep convolutional networks. *Advances in Neural Processing Systems, 25*.

Lerner, P. B. (2023). A New Entropic Measure for the Causality of the Financial Time Series. (MDPI, Ed.) *Journ. Risk and Financial Management, 16(7)*, 338. doi:https://doi.org/10.3390/jrfm16070338

Li, Y. W. (2010). Applications of artificial neural networks in financial economics: a survey. In I. Explore (Ed.), *International Symposium on Computational Intelligence and Design.* IEEE. doi:10.1109

Palaima, T. G. (1989). Cypro-Minoan scripts: problems of historical context. *Problems of Decipherment, 49*, 121-187.

Schneider, T. (2003). Kassitisch and Hurro-Urartaisch. Ein Discussionbeitrag zu moglichen lexikalischen Isoglossen. *Altorientalische Forschungen, 30*, 372-391.

Sommershield, T. Y. (2023). Machine learning for ancient languages: a survey. (MIT, Ed.) *Computational Linguistics, 49*(3), 703-747.

Sutskever, I. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems, 27*.

Valerio, M. F. (2016). *Investigating the Signs and Sounds of Cypro-Minoan (Doctoral Dissertation)*. Barcelona: Universitat de Barcelona.

Vaswani, A. N. (2017). Attention is all you need. *Advances in Neural Information, 30*.

Wang, S. Y.-L. (2015, December). Feedforward kernel neural networks, generalized least learining machine, and its deep learning with application to image classification. *Applied Soft Computing , 37*, 125-141.

Wegner, I. a. (2020). *Introduction to the Hurrian language.* Florence.

Xusen, Y. N. (2019). Decipherment of historical manuscript images. *International Conference on Document Analysis and Recognition.* IEEE. doi:10.1109