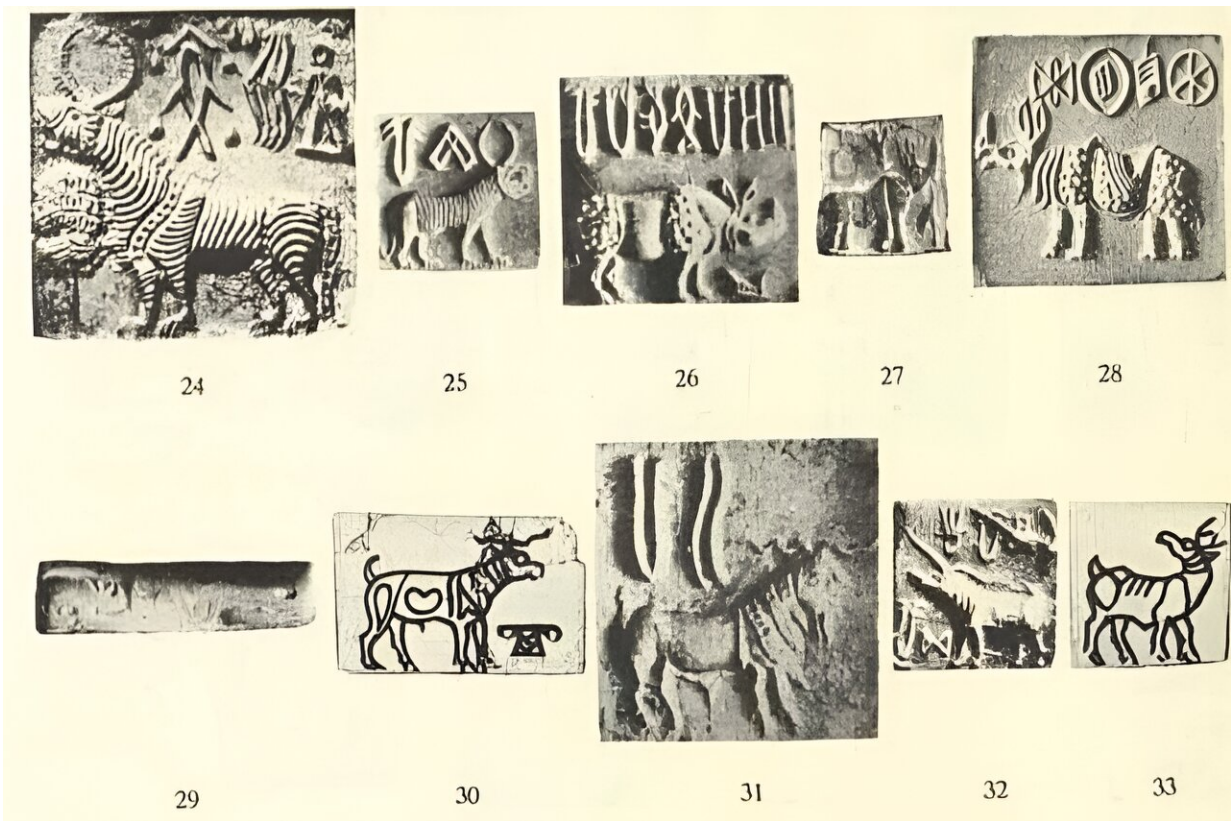# Researcher uses machine learning to help digitize ancient texts from Indus civilization

March 22 2024, by Adam Lowenstein

A series of Indus seals from Iravatham Mahadevan, a scholar and author who has studied Indus script for decades. Credit: Florida Institute of Technology

The civilization of Indus River Valley is considered one of the three earliest civilizations in world history, along with Mesopotamia and

Egypt. Bigger geographically than those two as it unfolded starting in 3300 BCE across what is now Pakistan and India, the Indus civilization boasted uniform weights and measures, skilled artisans, a multifaceted system of trade and commerce, and upwards of 500 symbols and signs for communicating.

But one question has vexed scholars for decades and hindered attempts to learn more about this civilization: Were those characters a language or more akin to pictograms? Even as some experts begin to translate the right-to-left script found in Indus inscriptions, there is little agreement.

"That's a controversy which is not yet settled," said Debasis Mitra, a professor of computer science who is now connected to this quest thanks to a novel grant he was awarded from the National Endowment for the Humanities: "Ancient Script Digitization and Archival (ASDA) of Indus Valley Artifacts using Deep Learning."

Graduate student assistant Deva Atturu, who will defend his master's thesis in April, is assisting Mitra with conducting the grant-funded research. Just this month he and Mitra virtually attended the South Asian Archaeology Conference 2024 from the University of Chicago, where Atturu presented on their work.

The writings they are studying may be a series of symbols like the equivalent of dollar signs and business transaction images, or those symbols may be graphemes, the individual letters or groups of letters that represent speech sounds.

"Both sides have very strong arguments," Mitra said.

He is not looking to solve the argument but to empower those who will by developing a machine learning algorithm for identifying and digitizing the Indus civilization's ancient script. There is a paucity of

digitized data that Mitra is hoping to address.

The process uses an automated script recognition (ASR) system to extract coded sequences of graphemes from a dataset of more than 1,000 photographs of Indus seals. Using two-staged artificial neural networks, the ASR has achieved 88% success in detecting graphemes.

Still, the process has been challenging. Often machine learning is empowered by inputting huge amounts of data to basically train the system. In this case, however, there is not much data to enter. And what data there is can sometimes be "noisy" or distorted.

"I work on medical imaging and some of the challenges are similar," Mitra said.

Mitra applies different machine learning elements to the project to try to generate new data or see if another approach may work better. And he also finds himself at conferences not usually on the schedule for computer scientists, like last year's Annual Conference of South Asia hosted by the University of Wisconsin in Milwaukee, where he presented on this machine learning project.

Attending these keeps him in contact with archaeologists who can feed him more data. "I go to these conferences and try to talk to them," he said.

He also enlists the help of students at the Indian Statistical Institute in his native India. Together they are making progress. They can digitize some motifs and graphemes and, depending on the amount of data, even create a script. Doing that and getting it into a database is the goal of the initial grant funding.

The next phase? Create a system that allows archaeologists in the field to

snap a smartphone photo of a text or symbols and have it routed into the database for digitization.

That these efforts are designed to help illuminate and better understand one of the great civilizations in the history of his country is added motivation for Mitra.

"It's part of my history, so there is extra motivation for that. And obviously I see Indian students are very interested because of the same reason," he said. "But one of the first breakthroughs was by a couple of American students who had strong interest in India, and some of them said they visited India afterwards."

Provided by Florida Institute of Technology

Citation: Researcher uses machine learning to help digitize ancient texts from Indus civilization (2024, March 22) retrieved 27 April 2024 from https://phys.org/news/2024-03-machine-digitize-ancient-texts-indus.html