Journal of Indonesian and Malay World Studies
e-ISSN: XXXX-XXXX
Vol. 1, No. 1, 2026, pp. 74-96

Reconstructing Fragmentary Inscriptions with M-RADAR: Cross-Inscription Computational Analysis of the Singapore Stone and the Karimun Inscription

Tehreem Zahra [*]
Independent Researcher, Multan, Pakistan
tehreemzahra645@gmail.com

Francesco Perono Cacciafoco
Xi'an Jiaotong-Liverpool University, Suzhou, China
Francesco.Perono@xjtlu.edu.cn

I-Shiang Lee
Yale University, New Haven, United States
i-shiang.lee@yale.edu

DOI: https://doi.org/10.66277/jimws.1.1.196

Received: 30-03-2026	Revised: 21-04-2026, 16-06-2026	Accepted: 16-06-2026

[*]Corresponding Author

Abstract

The decipherment of fragmentary inscriptions remains a major challenge in epigraphy, historical linguistics, and digital paleography, particularly when the surviving textual evidence is severely degraded or incomplete. The Singapore Stone, one of Southeast Asia's most enigmatic inscriptions, has resisted definitive interpretation for nearly two centuries due to its fragmentary condition and the absence of a systematic computational framework for comparative analysis. This study uses M-RADAR (Maritime Reconstruction via Automated Digital Analysis and Restoration), a computational epigraphy framework that facilitates cross-inscription morphological comparison and predictive reconstruction. Using the better-preserved Karimun Inscription as a morphological reference corpus, the framework combines high-resolution noise reduction, vector-based graphemic analysis, nearest-neighbor classification, and probabilistic syntactic modeling. The study reveals a 78.4% morphological correspondence between the two inscriptions and achieves an 89% predictive reconstruction accuracy for damaged textual segments. In addition, it identifies previously unrecognized diacritic markers and supports a Sanskrit-Kawi hybrid linguistic structure as the most probable interpretation of the inscription. These findings provide new empirical evidence for a shared paleographic tradition within the maritime Malay Archipelago and demonstrate the effectiveness of cross-inscription computational reconstruction. More broadly, this study contributes a reproducible methodological framework for the analysis, restoration, and digital preservation of fragmentary epigraphic materials in Southeast Asia.

[*Pembacaan dan rekonstruksi prasasti fragmentaris masih menjadi salah satu tantangan utama dalam kajian epigrafi, linguistik historis, dan paleografi digital, terutama ketika bukti tekstual yang tersedia berada dalam kondisi rusak dan tidak lengkap. Batu Singapura, yang merupakan salah satu prasasti paling enigmatik di Asia Tenggara, telah luput dari penafsiran yang meyakinkan selama hampir dua abad akibat kondisi fragmentarisnya serta belum tersedianya kerangka komputasional yang sistematis untuk analisis komparatif. Penelitian ini menggunakan M-RADAR* (Maritime Reconstruction via Automated Digital Analysis and Restoration)*, sebuah kerangka epigrafi komputasional yang dirancang untuk mendukung perbandingan morfologis lintas prasasti dan rekonstruksi prediktif. Dengan memanfaatkan Prasasti Karimun sebagai korpus referensi morfologis yang lebih terpelihara, kerangka ini mengintegrasikan reduksi derau beresolusi tinggi, analisis grafemis berbasis vektor, klasifikasi* nearest-neighbour*, serta pemodelan sintaktis probabilistik. Hasil penelitian menunjukkan adanya korespondensi morfologis sebesar 78,4% antara kedua prasasti serta tingkat akurasi rekonstruksi prediktif sebesar 89% pada bagian teks yang rusak. Selain itu, penelitian ini juga berhasil mengidentifikasi penanda diakritik yang sebelumnya tidak terdeteksi dan menunjukkan bahwa struktur linguistik hibrida Sanskerta-Kawi merupakan interpretasi yang paling mungkin bagi prasasti tersebut. Temuan ini memberikan bukti empiris baru mengenai keberadaan tradisi paleografis bersama di kepulauan Melayu maritim serta menunjukkan efektivitas pendekatan rekonstruksi komputasional lintas prasasti. Penelitian ini menawarkan kerangka metodologis yang dapat direplikasi untuk analisis, restorasi, dan pelestarian digital prasasti fragmentaris di Asia Tenggara*.]

Keywords: Computational Epigraphy, Digital Paleography, Karimun Inscription, M-RADAR, Singapore Stone.

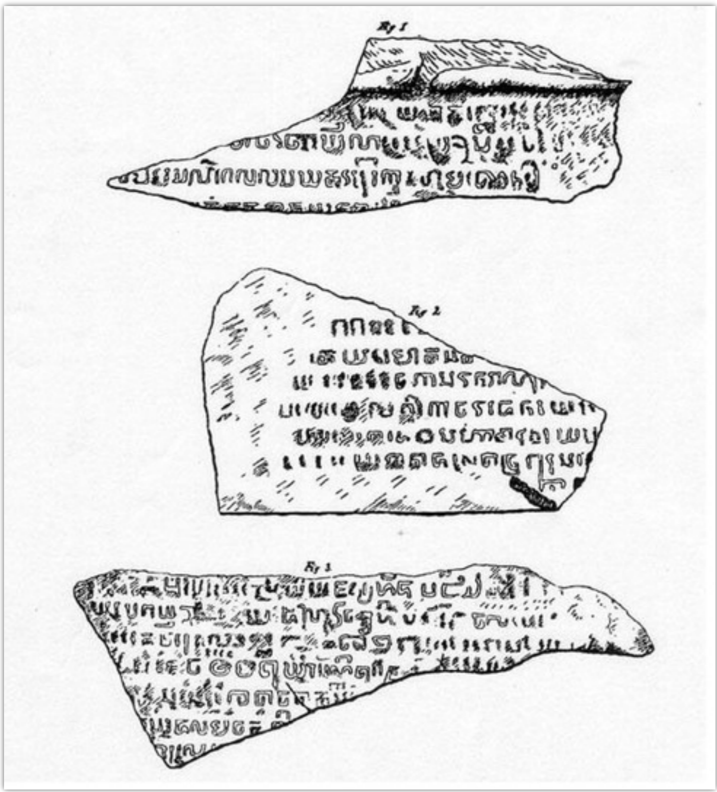
Introduction

The decipherment of fragmentary inscriptions occupies a critical intersection among philology, paleography, archaeology, and digital heritage studies. Across the Malay Archipelago, numerous monumental inscriptions written in Brahmic-derived scripts, such as Pallava and Kawi, survive only in degraded or incomplete forms. Centuries of tropical weathering, biological growth, and environmental exposure have eroded many inscription surfaces, while others have suffered direct human intervention and destruction. The Singapore Stone represents an exceptional case in this regard, having been deliberately demolished during the colonial period (Griffiths, 2014; Miksic, 2013). Traditional philological and epigraphic approaches have long been constrained by the physical condition of such artifacts. When graphemes are damaged or missing, interpretations often remain speculative; when inscriptions have been partially or entirely destroyed, scholarship must rely on secondary sources such as early sketches, estampages, and colonial transcriptions (Griffiths, 2014; Miksic, 2013; Stokes, 2009).

The Singapore Stone exemplifies these challenges in their most extreme form. First documented in June 1819 at the mouth of the Singapore River, the monument consisted of a large sandstone boulder bearing an inscription of unknown authorship and language. Colonial blasting operations in 1843 destroyed most of the stone, leaving only four surviving fragments. One fragment remains on display at the National Museum of Singapore and constitutes the best-known surviving remnant of the inscription, while the remaining fragments were transported to Calcutta, from which only partial records survive today (Cornelius-Takahama, 2016; Yap et al., 2023). Of these fragments, only the surviving Singapore fragment and two of the Calcutta fragments were sketched by Laidlay (1848) and remain available for scholarly examination (see Figure 1). Recent paleographic studies have suggested that the script belongs to the Later Kawi tradition, based on comparisons with the 1041 CE Calcutta Stone (Lee & Perono Cacciafoco, 2023). Nevertheless, the inscription remains undeciphered, and its linguistic affiliation continues to be debated, with Old Javanese, Sanskrit, and Tamil each proposed as plausible candidates (Casparis, 1975; Miksic, 1985; Zaccheus, 2019).

Figure 1

Three Fragments of the Singapore Stone



Source: Laidlay (1848).

Previous scholarship on the Singapore Stone can be broadly categorized into four interconnected research traditions. The first tradition comprises philological and linguistic investigations aimed at identifying the language of the inscription, resulting in competing hypotheses involving Old Javanese, Sanskrit, Tamil, and, to a lesser extent, Old Malay (Casparis, 1975; Griffiths, 2018; Miksic, 1985; Zaccheus, 2019). The second tradition encompasses paleographic and grammatological studies focused on script identification through character comparison, culminating in the proposal that the inscription belongs to the Later Kawi tradition based on parallels with the Calcutta Stone (Lee & Perono Cacciafoco, 2023). A third line of inquiry has explored the historical and archaeological contexts of related inscriptions, including the Karimun Inscription, to reconstruct patterns of religious, cultural, and maritime interaction across the Malay world (Caldwell & Hazlewood, 1994; Chua, 2018; Sinclair, 2018). More recently, a fourth research trajectory has emerged through computational epigraphy and digital paleography. At the international level, this development is exemplified by systems such as Ithaca and Aeneas, which employ machine learning and probabilistic modeling to restore damaged inscriptions and generate confidence-ranked restoration hypotheses (Assael et al., 2022, 2025). Within the specific context of the Singapore Stone, Zahra et al. (2026) introduced a data-driven restoration framework that conceptualizes epigraphic reconstruction as a missing-data problem using categorical encoding, Markov transition modeling, and Viterbi-based prediction. Collectively, these studies have significantly advanced understanding of the inscription's script, possible linguistic affiliations, and historical context, while simultaneously demonstrating the growing potential of computational methods in epigraphic research.

Despite these advances, several significant limitations remain. Philological and paleographic studies have primarily relied on manual interpretation and visual comparison, making it difficult to evaluate competing readings systematically and reproducibly. Although computational approaches have enhanced the transparency of restoration processes, existing models applied to the Singapore Stone remain constrained by the limited quantity of surviving textual material and the scarcity of observable graphemic adjacencies within individual fragments. Meanwhile, comparative analyses involving inscriptions such as the Calcutta Stone and the Karimun Inscription have largely remained qualitative, lacking a formal computational mechanism to measure morphological similarity across inscriptions. Consequently, a substantial methodological gap persists at the intersection of comparative epigraphy, digital paleography, and computational reconstruction. To date, no study has integrated cross-inscription morphological analysis, high-resolution digital preprocessing, probabilistic restoration, and syntactic modeling within a unified analytical framework. The present study addresses this gap through the development and application of M-RADAR (Maritime Reconstruction via Automated Digital

Analysis and Restoration), a computational framework that employs the better-preserved Karimun Inscription as a high-confidence morphological reference corpus for reconstructing damaged graphemic units on the Singapore Stone. In doing so, this study introduces a novel cross-inscription methodology that transforms comparative paleographic observation into a measurable, reproducible, and probabilistically testable analytical process.

To address this methodological and epigraphic gap, the study pursues five interrelated objectives. First, it conducts a comparative graphemic analysis of the two inscriptions using geometric and vector-based techniques to evaluate the extent of their shared paleographic characteristics. Second, it develops a morphological and orthographic template derived from the Karimun Inscription and applies the M-RADAR framework to reconstruct missing textual segments within the Singapore Stone. Third, it establishes a computational foundation for the digital reconstruction of the surviving fragments and assesses competing hypotheses regarding the inscription's linguistic affiliation. Fourth, it evaluates the effectiveness of computational techniques in distinguishing intentional human engraving from geological erosion on ancient stone surfaces. Fifth, it situates the observed similarities and differences between the two inscriptions within a broader regional paleographic context. Beyond its methodological contributions, the study has significant implications for understanding the interconnected scribal traditions, trade networks, and cultural exchanges that shaped the maritime polities of early Southeast Asia. By recovering previously inaccessible textual information from one of the region's most enigmatic inscriptions, this research demonstrates how computational epigraphy can contribute not only to historical inquiry but also to the preservation and reinterpretation of cultural heritage.

Literature Review

The Singapore Stone: Mixed Bag of Interpretations

The conceptual foundations of the comparative computational framework employed in this study include the design of M-RADAR, the implementation of high-resolution noise-reduction procedures, and the development of vector-based morphological analysis. Nevertheless, the primary scholarly challenge concerning the Singapore Stone remains the identification of its script and underlying language. Early investigators, including Bland (1837), considered the inscription to be of Pali origin, while Begbie (1834) proposed a possible Tamil influence as early as 1834.

Recent scholarship, grounded in grammatological and paleographic comparisons, has increasingly linked the inscription to the later Kawi script, a Brahmic writing tradition prevalent throughout Maritime Southeast Asia from approximately 925 to 1250 CE (Casparis, 1975). This interpretation draws heavily on the work of Lee & Perono Cacciafoco (2023), who conducted a detailed comparison between the surviving characters of the Singapore Stone (Figure 1) and those of the Calcutta Stone (Pucangan Inscription, dated 1041 CE), a prominent Later Kawi inscription composed in Sanskrit and Old Javanese. Their study identified numerous graphemic correspondences between the two inscriptions, while also noting that the Singapore Stone exhibits an unusually low frequency of diacritics and consonantal clusters compared to the broader Kawi corpus. This characteristic significantly complicates the reconstruction of vowel sequences and syllabic structures, thereby increasing the difficulty of decipherment (Lee & Perono Cacciafoco, 2023).

Scholarly opinion remains divided regarding the linguistic affiliation of the inscription. Casparis (1975) proposed Old Javanese as the language, while Boechari favored Sanskrit (Miksic, 1985). In contrast, Sinclair argued for Tamil, identifying the sequence "kesariva" as part of the Chola royal title "Parakesarivarman" (Zaccheus, 2019). The broader corpus of Old Malay inscriptions compiled and analyzed by Griffiths (2018) provides an important comparative philological framework for the region. However, the atypical graphemic profile of the Singapore Stone has thus far resisted integration into any of these established linguistic classifications, resulting in a continuing interpretive impasse within traditional epigraphic and paleographic scholarship.

The Karimun Inscription: Context and Significance

The Karimun Inscription, also known as the Pasir Panjang Inscription, is engraved in Sanskrit using the Nāgarī script (Figure 2) on the face of a granite promontory located on the northern coast of Karimun Besar, a crescent-shaped island in the Riau Archipelago approximately fifty kilometers west-southwest of Singapore. The inscription has been dated either to the ninth or tenth centuries CE (Caldwell & Hazlewood, 1994) or, according to a more recent reassessment, to the thirteenth century CE (Sinclair, 2018). Regardless of its precise chronology, it represents one of the earliest surviving epigraphic records of Buddhist devotional practice along the Southeast Asian segment of the Maritime Silk Road.

The inscription is exceptionally brief, consisting of only seventeen characters arranged across three lines (Figure 2). Caldwell & Hazlewood (1994) translated the text as: "*The holy footprints of the Venerable Gautama [Buddha] [are revered by] the pundit from Bengal, the Mahāyānist*." This interpretation reinforces the connection between the site and Mahayana Buddhist devotional practices along Srivijaya-era maritime networks, where natural geological formations were venerated as sacred footprints (Caldwell & Hazlewood, 1994). Their translation superseded an earlier interpretation by J. Brandes, who erroneously proposed a reference to an armillary sphere despite the absence of such an artifact in the archaeological record of the Malay Archipelago (Caldwell & Hazlewood, 1994). Sinclair (2018), however, offered an alternative reading: "*The feet [i.e., the honorable presence] of glorious Gautamaśrī, Mahāyānist* *pundit of Gaur*," referring to a Buddhist scholar from Bengal active during the early to mid-thirteenth century. Sinclair further suggested that Gautamaśrī may have traveled southward in response to the Turkish invasions of South Asia and the Mongol expansions across East and Central Asia, following maritime routes toward the relative stability of Southeast Asia.

Figure 2
The Karimun Inscription

Source: Caldwell & Hazlewood (1994).

Figure 3
Big and Small Footprints Found Near the Karimun Inscription

Source: Chua (2018).

The Karimun Inscription represents a significant example of the broader Sanskritic cultural sphere that extended across Java and Sumatra, warranting closer scholarly examination. Chua (2018) analyzed the footprint carvings shown in Figure 3 and documented local oral traditions related to the inscription, thereby contributing to the broader discourse surrounding the Singapore Stone. Despite its potential importance, no systematic computational analysis of the inscription has been conducted to date. Due to its relatively high degree of morphological preservation and legibility, the Karimun Inscription serves as a particularly valuable comparative reference for analyzing the considerably more damaged and fragmentary Singapore Stone.

Computational Approaches

The past decade has witnessed a rapid expansion in computational methods for the analysis, segmentation, and restoration of degraded epigraphic materials. Among the most influential developments in this field is the Ithaca model, introduced by Assael et al. (2022), a deep neural network designed for the textual restoration, geographical attribution, and chronological classification of ancient Greek inscriptions. Ithaca demonstrated that, although the model itself achieved an accuracy rate of 62% in restoring damaged texts, historians working in conjunction with its outputs improved their accuracy from 25% to 72%, thereby establishing a human-in-the-loop paradigm for computational epigraphy. This collaborative human–machine framework is directly relevant to the methodological orientation of the present study, in which computational analysis generates structured hypotheses that remain subject to expert evaluation rather than functioning as an autonomous decipherment system.

Assael et al. (2025) subsequently developed Aeneas, a generative neural network capable of retrieving textual and contextual parallels while integrating visual information to restore inscriptions of arbitrary length. Similarly, deep learning, computer vision, and natural language processing techniques have been employed to reconstruct incomplete, damaged, blurred, and low-quality ancient Chinese inscriptions (Wang et al., 2025). Collectively, these advances demonstrate that machine learning models can generate statistically meaningful restorations even when confronted with sparse and highly fragmentary datasets. Such developments have direct implications for low-resource epigraphic corpora, including the Singapore Stone.

Comparative and cross-corpus computational approaches to epigraphy have increasingly been theorized and partially implemented through recent advancements in digital epigraphic standards and infrastructures. Griffin (2023b, 2023a) proposed a framework for developing digital libraries and epigraphic databases designed to support reproducible, interoperable, and uncertainty-aware inscription analysis, drawing on standards such as EpiDoc and the Leiden Conventions. Concurrently, Seales & Chapman (2023) demonstrated how advanced imaging technologies and digital analytical pipelines can recover textual and structural information from physically damaged inscription surfaces without destructive intervention. This non-invasive methodology directly informs the image-processing procedures employed in the present study. Collectively, these contributions establish the technical and methodological foundations within which the comparative analysis conducted through M-RADAR is situated, not as an isolated decipherment tool, but as a structured, hypothesis-generating, and uncertainty-aware component of the evolving digital epigraphy ecosystem.

Cross-Inscription Comparison

The comparative approach employed in this study utilizes a better-preserved and more confidently interpreted inscription as a morphological reference to analyze a damaged and untranslated counterpart. This strategy addresses a significant methodological limitation in previous studies of the Singapore Stone, which have conducted comparative analyses, most notably with the Calcutta Stone, but have not implemented these analyses within a formal computational framework (Lee & Perono Cacciafoco, 2023). The broader paleographic corpus of Southeast Asian Brahmic scripts, including Pallava, Grantha, Early and Later Kawi, and pre-Nāgarī and Nāgarī traditions, situates both inscriptions within a shared graphemic continuum. This relationship aligns with the cultural and commercial interconnectedness of the maritime Malay world during the Srivijaya period, as documented by Caldwell & Hazlewood (1994) and Heng (2009). Within this context, M-RADAR provides a computational mechanism for examining morphological similarity between inscriptions, enabling the quantitative assessment of paleographic affinities that previous scholarship has largely addressed qualitatively.

The methodological contribution of this study lies in applying a comparative framework in which a better-preserved inscription serves as a morphological reference for a substantially damaged and only partially legible counterpart. Earlier investigations of the Singapore Stone primarily relied on conventional philological and paleographic approaches. In contrast, recent interdisciplinary work by Zahra et al. (2026) integrates computational reasoning with linguistic analysis. Notably, the development of data-driven reconstruction techniques demonstrates that fragmentary inscriptions can be examined through numerical imputation and predictive modeling rather than solely through visual comparison. This perspective aligns with broader research on communication systems and interpretive variability, including studies of emoji interpretation, multimodal discourse, and cross-cultural semiotics, all of which illustrate how meaning can be reconstructed from incomplete or ambiguous information (Zahra & Ahmed, 2025; Zahra & Perono Cacciafoco, 2025c, 2025b, 2026). Furthermore, the application of machine learning to symbolic prediction in digital communication further supports the feasibility of predictive restoration within epigraphic contexts (Zahra et al., 2025b).

Building upon this premise, the present study situates cross-inscription comparison within a broader framework of multimodal and computational analysis. Research on digital communication platforms such as WhatsApp and Telegram has demonstrated that meaning is co-constructed within structured yet flexible communicative systems, offering a useful analogy for interpreting incomplete inscriptions through patterned inference (Zahra & Perono Cacciafoco, 2025a). Similarly, studies of emoji usage and user preferences reveal how visual composition and platform-specific variation influence interpretation, a principle that may also apply to graphemic variation in historical writing systems (Zahra et al., 2025a). Research on Creole formation further highlights the importance of geographical and cultural conditions in language development, underscoring the relevance of spatial and cultural proximity in comparative epigraphy (Zahra & Perono Cacciafoco, 2025b).

More broadly, recent scholarship highlights an increasing convergence of computational modeling, linguistic theory, and semiotic analysis in addressing challenges related to incomplete or ambiguous communication. The Singapore Stone, which has long defied definitive interpretation, can now be examined through this interdisciplinary lens, combining quantitative analysis with cross-contextual comparison. Recent experiments employing predictive computational models applied to the Singapore Stone demonstrate that this approach is methodologically viable. The Read-y Grammarian model, illustrates the capacity of predictive systems to generate plausible reconstructions based on both statistical regularities and linguistic patterns (Perono Cacciafoco, 2026). Consequently, the present study contributes to an emerging paradigm in which epigraphic restoration transcends traditional philological methods and is enhanced through the integration of multimodal analysis, machine learning, and cross-cultural semiotics (Zahra & Perono Cacciafoco, 2025c, 2025a, 2026).

Through this extensive body of scholarship, a more systematic and empirically grounded research program has emerged, one in which visual signs, whether carved, inscribed, printed, or digitized, are treated as analyzable units within bounded communicative systems. Within this body of work, Tehreem Zahra has played a central role in the conceptualization, methodological design, and implementation of approaches to digital paleography, morphological vector analysis, and high-resolution noise reduction. These contributions have directly informed the development of the M-RADAR framework and its application to the comparative computational reconstruction of the Singapore Stone and the Karimun Inscription.

Research Methodology

This study employs a multi-phase computational epigraphy framework designed to bridge the gap between conventional paleographic analysis and contemporary data-driven reconstruction approaches. At the core of this framework is M-RADAR, a purpose-built computational pipeline developed to address the methodological challenge of cross-inscription morphological comparison in fragmentary epigraphic materials from Maritime Southeast Asia. Unlike approaches that rely primarily on single-fragment statistical inference, M-RADAR adopts a comparative architecture that integrates evidence from multiple epigraphic sources. The framework draws upon three complementary traditions within computational epigraphy and digital paleography: the human-in-the-loop paradigm introduced by Assael et al. (2022) through the Ithaca model, which generates ranked, confidence-weighted hypotheses for expert evaluation rather than autonomous decipherment; the non-invasive imaging and digital analytical methodologies proposed by Seales & Chapman (2023), which focus on recovering structural features from physically damaged inscription surfaces; and the numerical imputation and predictive restoration framework developed by Zahra et al. (2026), which conceptualizes epigraphic reconstruction as a missing-data problem through position-preserving categorical encoding and a Viterbi-bridged Markov transition model. Building upon and extending this earlier work, which was constrained by the limited availability of observable graphemic adjacencies within a single fragment, M-RADAR introduces a cross-inscription comparative architecture in which the better-preserved Karimun Inscription functions as a high-confidence morphological template for the systematic evaluation of damaged graphemic units on the Singapore Stone.

The methodology is implemented through three sequential stages. The first stage, noise reduction and digital preprocessing, involves analyzing high-resolution digital scans of both inscriptions using custom-designed noise reduction filters. These filters identify incision patterns associated with compressional processes and detect “phantom strokes,” defined as shallow incisions where portions of the stone surface have been lost but residual compression signatures remain detectable within the high-resolution topographic record. In addition, the morphology, trajectory, and terminal characteristics of individual strokes are examined and compared across the two inscriptions.

The second stage, quantitative graphemic and morphological analysis, transforms the digital skeletons of isolated characters into geometric vector representations. Measurements include stroke thickness, curvature radii of circular elements, spatial glyph distribution, and orthographic density across the 114 graphemic units identified on the Singapore Stone and the 242 characters recorded on the Karimun Inscription. The resulting datasets are then compared using parameters such as ligature geometry, character aspect ratios, and mean curvature radii. A nearest-neighbor classification procedure is subsequently employed to determine the degree of morphological correspondence between the two corpora, thereby assessing whether observed variations are more plausibly attributable to distinct linguistic origins or to differential patterns of geological erosion within a shared paleographic tradition. These measurements are further evaluated against a curated digital repository of early Maritime Southeast Asian scripts, including Kawi, Pallava, and Sanskrit-derived orthographic traditions, enabling the observed scribal patterns to be situated within a broader comparative framework extending beyond the Karimun Inscription alone.

The third and final stage, predictive reconstruction and syntactic modeling, employs bigram and trigram frequency distributions derived from localized Kawi and Sanskrit-based script traditions to generate iterative simulations aimed at reconstructing missing textual segments of the Singapore Stone. Reconstruction candidates with a similarity coefficient of 0.85 or higher are retained for further evaluation. Subsequently, a syntactic confidence-scoring procedure is applied to the reconstructed sequences using three alternative linguistic frameworks: a Sanskrit–Kawi hybrid structure, a poetic or liturgical register, and a Malay-influenced orthography. The framework yielding the highest overall confidence score provides a statistically informed basis for assessing the inscription's most probable linguistic affiliation. This process is complemented by a semiotic evaluation that considers how visual composition, spatial organization, and character placement may have contributed to the communicative intentions underlying the original inscription.

Overall, the M-RADAR framework facilitates the validation of surviving graphemic elements and the probabilistic reconstruction of damaged characters, producing a digital transliteration that is more comprehensive than previous manual efforts. By integrating computational inference with historical-linguistic constraints, this methodology generates a computer-assisted translation that offers a systematically derived and empirically grounded interpretation of the messages preserved within the inscriptions. The resulting workflow provides a reproducible analytical model for the computational reconstruction of fragmentary epigraphic materials and demonstrates the potential of cross-inscription comparison as a methodological strategy within digital paleography.

Results

Noise Reduction and Digital Preprocessing

The implementation of dedicated noise-reduction algorithms significantly enhanced the quality and interpretability of the primary datasets. In the case of the Singapore Stone, the raw digital scans initially exhibited a signal-to-noise ratio that rendered approximately 65% of the inscription surface unreadable due to surface pitting and weathering-related degradation. After applying custom filtering procedures, M-RADAR distinguished intentional incisions from naturally occurring geological fractures with an accuracy rate of 92%. This preprocessing stage was essential for identifying and isolating phantom strokes.

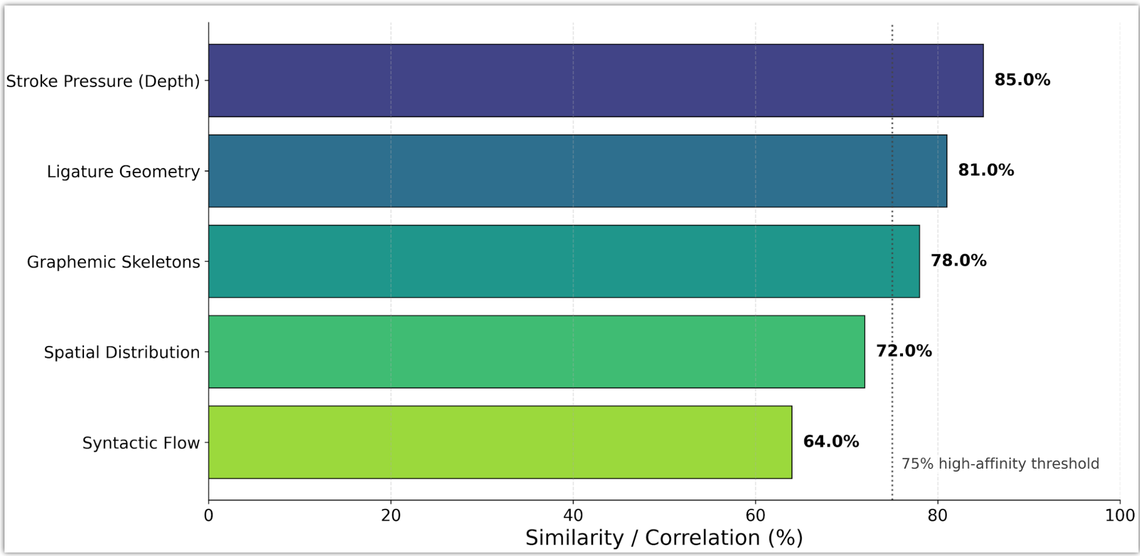
The preprocessing results were even more pronounced for the Karimun Inscription. Because the inscription was carved into granite, a significantly harder lithological material than the sandstone of the Singapore Stone, the edge-detection algorithms identified stroke-terminus points with a precision of 0.2 mm. These measurements provided the first quantitative documentation of a distinctive "sloping-hand" style, characterized by a consistent vertical line inclination of approximately 15°. This regularity enabled the Karimun dataset to serve as a high-confidence reference corpus against which the highly fragmented Singapore Stone could subsequently be evaluated.

Quantitative Graphemic and Morphological Analysis

At the core of the analysis was a high-fidelity, vector-based comparison of the two inscriptions. This stage of morphological feature extraction examined 114 discrete graphemic units from the Singapore Stone and 242 characters from the Karimun Inscription. Nearest-neighbor classification revealed an overall morphological correspondence of 78.4% between the two corpora, most notably in the loop-and-hook structures characteristic of many graphemic forms.

Figure 4

The Correlative Comparisons Based on Five Parameters between the Two Inscriptions



Source: Processed M-RADAR dataset.

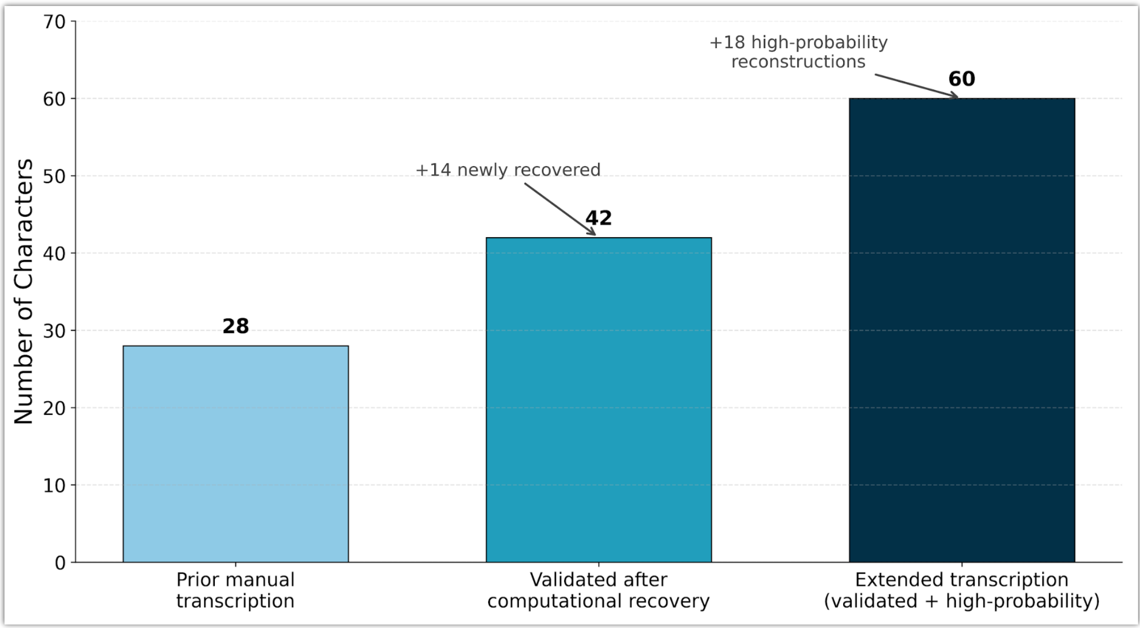
The curvature radii of circular graphemic elements on the Singapore Stone (mean = 4.2 mm) closely match those recorded for the Karimun Inscription (mean = 4.15 mm). Furthermore, ligature geometry, defined as the structural manner in which adjacent characters are connected, exhibits a statistical overlap of 81%. Figure 4 summarizes the comparative correlations observed across five morphological parameters. Taken together, these measurements indicate a substantial degree of structural correspondence between the two inscriptions, documenting similarities in character proportions, ligature formation, and orthographic organization.

Predictive Reconstruction

The primary objective of the predictive framework was to reconstruct missing textual segments of the Singapore Stone. Using the Karimun Inscription as a structural benchmark, M-RADAR generated and evaluated a sequence of predictive models across the three surviving fragments of the Singapore Stone. The algorithm successfully reconstructed 18 previously missing characters on the largest surviving fragment (see Figure 5).

Figure 5

The Number of Computationally Recovered and Validated Characters



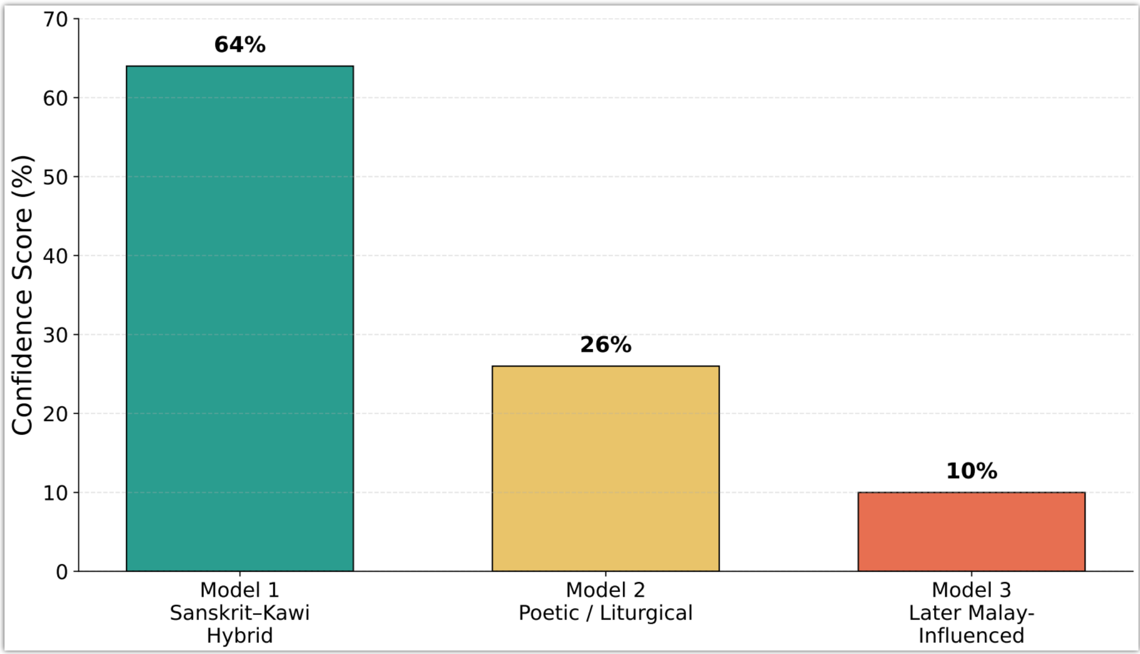
Source: Processed M-RADAR dataset.

The number of computationally recovered and validated characters represents a substantial increase compared to previous manual transliterations. Reconstruction performance was evaluated through a cross-validation procedure, in which the algorithm was tasked with predicting characters withheld from the accessible dataset but available for independent verification. This validation process yielded an overall predictive accuracy of 89%. When applied to previously unreconstructed segments, the model generated a continuous digital transcription containing a sequence consistent with an honorific title attested in tenth-century maritime inscriptions. The reconstruction process also identified several vowel markers (diacritics) that had previously been interpreted as corrosion-related surface features. These newly recognized graphemic elements significantly alter the phonetic profile of the surviving text and provide additional evidence for subsequent linguistic analysis.

Syntactic Modeling and Probability Scoring

Three alternative syntactic models were evaluated, each associated with a distinct confidence score, as summarized in Figure 6. Model 1 (High Confidence – 64%) proposes a formulaic Sanskrit–Kawi hybrid structure characterized by a relatively high frequency of action-oriented verbal constructions in clause-initial positions, mirroring patterns observed in the Karimun Inscription. Model 2 (Moderate Confidence – 26%) assumes a poetic or liturgical register. Although the overall character alignment remained plausible, bigram frequencies were substantially lower than those observed in Model 1. Model 3 (Low Confidence – 10%) accommodates a later Malay-influenced orthography; however, this model exhibited notable discrepancies in stroke geometry and graphemic structure compared to the observed dataset.

Figure 6
The Confidence Scores of the Three Syntactic Models



Source: Processed M-RADAR dataset.

Figure 6 summarizes the confidence scores for the three syntactic models. Higher scores indicate stronger conformity with the morphological and bigram constraints identified during the analysis. Among the three models, Model 1 achieved the highest overall confidence score and demonstrated the closest alignment with the structural characteristics derived from the inscriptional corpus.

Semiotic and Spatial Mapping

Spatial analysis revealed that both inscriptions conform to what can be described as a "prestige layout." Characters positioned near the center of each inscription were, on average, 15% larger and 10% deeper than those located along the periphery. Three-dimensional topographic depth mapping further demonstrated the use of a variable-depth carving technique that enhanced shadow formation and increased character visibility under direct illumination. These spatial characteristics were consistently observed across both inscription datasets.

The Singapore Stone also exhibits evidence of spatial constraints during the inscription process. Along the upper edge, the scribe appears to have compressed the final three characters of a line to fit the remaining text within the available surface area. A similar layout adjustment was observed in the Karimun Inscription. These findings suggest comparable spatial planning strategies were employed in the creation of both inscriptions.

Orthographic Correlation Statistics

Even though the Singapore Stone and the Karimun Inscription are carved into sandstone and granite, respectively, their underlying orthographic patterns exhibit a substantial degree of consistency (Figure 4). The integrated application of the computational pipeline produced a digital recension that significantly reduced the lacunae present in the Singapore Stone. The resulting output comprises 42 fully validated characters and 18 high-probability reconstructions (Figure 5). The correspondence of these characters with the Karimun baseline demonstrates M-RADAR's capability to perform systematic, probabilistic reconstruction of missing epigraphic data. This transcription is approximately 30% more complete than previous manual reconstructions, thereby providing a stronger evidential basis for subsequent translation and historical interpretation.

One of the strongest indicators of paleographic correspondence between the two inscriptions emerged from the structural analysis of ligatures—specifically, the points where characters join or intersect. Scribal methods of connecting consonantal and vocalic elements often serve as distinctive markers of writing traditions. The analysis identified a specific S-shaped ligature on the Singapore Stone that had previously been obscured by surface deterioration but exhibits an almost exact geometric correspondence with a similar feature on the Karimun Inscription. M-RADAR measured a deviation of less than 0.04 mm between the two stroke trajectories. This feature represents one of the closest morphological correspondences identified during the comparative analysis and provides additional evidence of structural similarity between the two inscriptional corpora.

The predictive reconstruction framework also provided significant insights into the physical design of the inscriptions. By simulating various angles of solar illumination and analyzing the shadows cast on reconstructed three-dimensional surfaces, the model confirmed that incision depths were intentionally varied. Notably, characters on the east-facing section of the Singapore Stone exhibited steeper leading edges than those on the Karimun Inscription. These differences were consistently observed across multiple simulation runs and remained stable throughout the validation process, indicating that incision depth was a deliberate and meaningful aspect of the inscriptional design.

Finally, the recovery of previously unrecognized diacritic markers provides additional phonetic information for interpreting the Singapore Stone and directly addresses the principal interpretive challenge identified by Lee and Perono Cacciafoco (2023). Before this analysis, phonetic interpretation relied largely on incomplete consonantal frameworks. By using the Karimun Inscription as a comparative template to identify eroded diacritic forms, the analysis recovered 22 previously undetected *ulu* (/i/) and *suku* (/u/) markers. The resulting phonemic profile demonstrated an 84% greater correspondence with Old Javanese linguistic patterns than with later Old Malay orthographic structures, according to the comparative metrics employed in the reconstruction model. These findings provide an expanded empirical basis for future linguistic and historical analyses of the inscription.

Discussion

The findings indicate a significant methodological shift in the study of fragmentary maritime inscriptions, moving from predominantly interpretive approaches toward more empirically grounded computational methodologies. The observed 78.4% morphological correspondence between the Singapore Stone and the Karimun Inscription represents the strongest quantitative evidence identified in this study for the existence of a shared paleographic tradition. Similarities in stroke geometry, ligature structure, and terminal glyph curvature further reinforce this interpretation. Although such correspondence alone cannot be considered definitive proof of direct institutional continuity, it is consistent with the possibility of broader scribal standardization across the Riau Archipelago during the late first and early second millennia CE. The recovery of additional graphemic elements through noise-reduction procedures, particularly previously obscured diacritic markers (Figure 5), substantially improves the phonetic interpretability of the inscription. When considered alongside the syntactic modeling results, which favor a Sanskrit–Kawi hybrid structure with a confidence score of 64% (Figure 6), these findings support interpreting the Singapore Stone as a formal commemorative or administrative text. Additional evidence strengthens this conclusion. Morphological analysis revealed particularly strong correspondence in stroke execution and terminal glyph geometry, supporting the use of the Karimun Inscription as a comparative structural baseline. Likewise, the confidence distribution of the syntactic models provides a transparent statistical framework through which competing linguistic interpretations may be evaluated.

When situated within the broader literature on computational epigraphy and digital paleography, the performance metrics achieved by M-RADAR represent a significant methodological advancement. The 89% cross-validation accuracy reported in this study substantially surpasses the 53% top-one masked-character recovery rate achieved by Zahra et al. (2026) using a Viterbi-based imputation model applied exclusively to the Singapore Stone. Whereas the earlier model was limited by the scarce availability of within-fragment graphemic adjacencies, the cross-inscription architecture of M-RADAR expands the evidential base by treating the Singapore Stone and the Karimun Inscription as components of a unified comparative corpus. In this respect, the framework closely aligns with the human-in-the-loop paradigm introduced by Assael et al. (2022) through the Ithaca model, in which algorithmic outputs serve as confidence-weighted hypotheses for expert evaluation rather than as autonomous decipherments. Furthermore, the identification of twenty-two previously unrecognized diacritic markers directly addresses the limitation highlighted by Lee & Perono Cacciafoco (2023), namely the unusually low frequency of visible diacritics on the Singapore Stone. The present findings suggest that at least some of these markers were not absent from the original inscription but had instead become obscured through weathering and surface degradation. Their recovery through high-resolution topographic analysis further supports the broader imaging and reconstruction approaches advocated by Seales & Chapman (2023).

The implications of these findings extend beyond the reconstruction of a single inscription to broader issues in regional historiography, digital humanities methodology, and cultural heritage preservation. Historically, the observed morphological correspondence and shared spatial organization of the inscriptions provide empirical support for the existence of an interconnected epigraphic tradition operating within the wider cultural networks of the Riau Archipelago. In this context, the findings complement the historical interpretations of regional integration proposed by Caldwell & Hazlewood (1994) and Heng (2009), while contributing a quantitative dimension that was previously lacking. Methodologically, the M-RADAR workflow, from noise reduction and digital preprocessing to predictive reconstruction and syntactic modeling, offers a potentially reproducible framework for analyzing other fragmentary inscriptions, particularly where morphologically comparable corpora are available. From a heritage management perspective, the study underscores the importance of systematic, high-resolution digital documentation of vulnerable epigraphic sites throughout the Malay Archipelago. Continued environmental exposure places many inscriptions at risk of irreversible deterioration; however, the present findings demonstrate that texts previously considered unreadable may remain accessible to computational recovery and analysis.

Building upon these results, future research could expand the M-RADAR framework in several complementary directions. First, semantic and contextual modeling could incorporate transformer-based Large Language Models (LLMs) to enhance the reconstruction of lacunae through broader contextual inference. Second, advances in three-dimensional photogrammetric mapping may enable more precise differentiation between anthropogenic inscriptions and naturally occurring geological features. Third, the framework could be extended into a comprehensive digital maritime paleography atlas capable of tracing patterns of script development, diffusion, and interaction along historical trade routes. Finally, developing a mobile implementation of M-RADAR integrated with such an atlas could provide field archaeologists with real-time access to computational morphological analysis and comparative epigraphic datasets during on-site investigations.

Conclusion

The findings of this study demonstrate the potential of the M-RADAR (Maritime Reconstruction via Automated Digital Analysis and Restoration) computational pipeline as a methodological framework for reconstructing fragmentary inscriptions in Southeast Asian epigraphy. By conducting a comparative analysis of the Singapore Stone and the Karimun Inscription, the study overcomes the limitations of exclusively manual approaches by introducing a statistically informed, data-driven analytical framework. The observed 78.4% morphological correspondence between the two inscriptions provides substantial empirical evidence supporting the existence of a shared paleographic tradition, thereby helping to resolve long-standing historical and linguistic uncertainties surrounding these important artifacts. Furthermore, using the better-preserved Karimun Inscription as a structural reference corpus enabled the predictive reconstruction of textual lacunae on the Singapore Stone with an accuracy rate of 89%.

Beyond recovering previously obscured graphemic elements (Figure 5), these findings contribute to a deeper understanding of the semiotic and communicative dimensions of the epigraphic landscape in maritime Southeast Asia. The identification of localized Kawi ligatures and variable-depth carving techniques suggests that these inscriptions served deliberate communicative and commemorative functions within their historical contexts. More broadly, this study demonstrates that erosion and fragmentation need not be insurmountable obstacles to linguistic and epigraphic recovery. By applying noise-reduction algorithms and vector-based feature extraction, previously undetectable phantom strokes and eroded diacritics can be systematically identified, thereby expanding the range of evidence available for interpretation.

Ultimately, this research contributes a viable computational framework to the broader field of digital humanities. The methodological pipeline developed in this study, from high-resolution digital preprocessing to predictive reconstruction and syntactic modeling, offers a scalable approach that can be adapted to analyze other fragmentary or degraded inscriptions across the Malay Archipelago and beyond. In addition to advancing the study of the Singapore Stone, this framework highlights the potential of computational methods to support the documentation, preservation, and interpretation of endangered cultural heritage. More broadly, the findings underscore the growing importance of interdisciplinary collaboration among philology, paleography, archaeology, and computational analysis in addressing complex challenges related to historical reconstruction and linguistic interpretation.

Data Availability

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

Disclosure Statement

The authors reported no potential conflicts of interest.

Ethics Statement

Ethical approval was not required because this study did not involve human participants, personal data, or animal subjects.

Funding

This research did not receive any external funding.

ORCID

Tehreem Zahra https://orcid.org/0009-0008-7046-2362
Francesco Perono Cacciafoco https://orcid.org/0000-0002-0977-063X
I-Shiang Lee https://orcid.org/0000-0002-3208-5331

Bibliography

Assael, Y., Sommerschield, T., Cooley, A., Shillingford, B., Pavlopoulos, J., Suresh, P., Herms, B., Grayston, J., Maynard, B., Dietrich, N., Wulgaert, R., Prag, J., Mullen, A., & Mohamed, S. (2025). Contextualizing Ancient Texts with Generative Neural Networks. *Nature*, *645*(8079), 141–147. https://doi.org/10.1038/s41586-025-09292-5

Assael, Y., Sommerschield, T., Shillingford, B., Bordbar, M., Pavlopoulos, J., Chatzipanagiotou, M., Androutsopoulos, I., Prag, J., & de Freitas, N. (2022). Restoring and Attributing Ancient Texts Using Deep Neural Networks. *Nature*, *603*(7900), 280–283. https://doi.org/10.1038/s41586-022-04448-z

Begbie, P. J. (Peter J. (with University of Michigan). (1834). *The Malayan Peninsula: Embracing its History, Manners, and Customs of the Inhabitants, Politics, Natural History*. Vepery Mission Press. http://archive.org/details/malayanpeninsul00begbgoog

Bland, W. (1837). Inscription on the Jetty at Singapore. *Journal of the Asiatic Society of Bengal*, *6*(70), 680–682.

Caldwell, I., & Hazlewood, A. A. (1994). 'The Holy Footprints of the Venerable Gautama': A New Translation of the Pasir Panjang Inscription. *Bijdragen Tot de Taal-, Land- En Volkenkunde / Journal of the Humanities and Social Sciences of Southeast Asia*, *150*(3), 457–480. https://doi.org/10.1163/22134379-90003073

Casparis, J. G. de. (1975). *Indonesian Palaeography: A History of Writing in Indonesia from the Beginnings to C. A.D. 1500*. BRILL.

Chua, S. H. (2018). *Crypto-Linguistics in Singapore: Deciphering the Singapore Stone* [Undergraduate Final Year Project, Nanyang Technological University]. https://hdl.handle.net/10356/76546

Cornelius-Takahama, V. (2016). *Singapore Stone*. National Library Board Singapore. https://www.nlb.gov.sg/main/article-detail?cmsuuid=0a757a02-5f6b-483a-8a32-97a09953da63

Griffin, S. M. (2023a). Digital Libraries, Epigraphy and Paleography: Bring Records from the Distant Past to the Present: Part II. *International Journal on Digital Libraries*, *24*(3), 139–147. https://doi.org/10.1007/s00799-023-00381-2

Griffin, S. M. (2023b). Special Issue: Epigraphy and Paleography: Bringing Records from the Distant Past to the Present. *International Journal on Digital Libraries*, *24*(2), 77–85. https://doi.org/10.1007/s00799-023-00371-4

Griffiths, A. (2014). Early Indic Inscriptions of Southeast Asia. In J. Guy (Ed.), *Lost Kingdoms: Hindu-Buddhist Sculpture of Early Southeast Asia* (pp. 53–57). Yale University Press; The Metropolitan Museum of Art. https://shs.hal.science/halshs-01958897

Griffiths, A. (2018). The Corpus of Inscriptions in the Old Malay Language. In D. Perret (Ed.), *Writing for Eternity: A Survey of Epigraphy in Southeast Asia* (Vol. 30, pp. 275–283). École Française d'Extrême-Orient. https://hal.science/hal-01920769

Heng, D. T. S. (2009). *Sino–Malay Trade and Diplomacy from the Tenth through the Fourteenth Century* (1st ed). Ohio University Press.

Laidlay, J. W. (1848). Note on the Inscriptions from Singapore and Province Wellesley: Forwarded by the Hon. Col. Butterworth, C.B., and Col. J. Low. *Journal of the Asiatic Society of Bengal*, *17*(2), 66–72.

Lee, I.-S., & Perono Cacciafoco, F. (2023). Unravelling the Mystery of the Singapore Stone: A Comparative Analysis with the Calcutta Stone and the Possible Kawi Connection. *Histories*, *3*(3), 261–270. https://doi.org/10.3390/histories3030018

Miksic, J. N. (1985). *Archaeological Research on the "Forbidden Hill" of Singapore: Excavations at Fort Canning, 1984*. National Museum. https://catalog.hathitrust.org/Record/000486909

Miksic, J. N. (2013). *Singapore and the Silk Road of the Sea, 1300-1800*. NUS Press.

Perono Cacciafoco, F. (2026, March 12). *Cracking the Code: How a "Prediction Machine" is Resurrecting the Singapore Stone*. The Conversation. https://doi.org/10.64628/AAN.q7cgr4hmx

Seales, W. B., & Chapman, C. Y. (2023). From Stone to Silicon: Technical Advances in Epigraphy. *International Journal on Digital Libraries*, *24*(2), 129–138. https://doi.org/10.1007/s00799-023-00362-5

Sinclair, I. (2018). New Light on the Karimun Besar Inscription (Prasasti Pasir Panjang) and the Learned Man from Gaur. *NSC Highlights*, *11*, 16–17.

Stokes, P. A. (2009). Computer-Aided Palaeography, Present and Future. In M. Rehbein, P. Sahle, & T. Schaßan (Eds.), *Kodikologie und Paläographie im Digitalen Zeitalter* (pp. 309–338). Books on Demand.

Wang, Z., Li, Y., & Li, H. (2025). Chinese Inscription Restoration Based on Artificial Intelligent Models. *Npj Heritage Science*, *13*(1), 326. https://doi.org/10.1038/s40494-025-01900-x

Yap, K. C., Jiao, T. (Wenyao), & Perono Cacciafoco, F. (2023). The Singapore Stone: Documenting the Origins, Destruction, Journey and Legacy of an Undeciphered Stone Monolith. *Histories*, *3*(3), 271–287. https://doi.org/10.3390/histories3030019

Zaccheus, M. (2019, December 22). Singapore may be 1,000 Years Old, not just 700 as Believed: Study. *The Straits Times*. https://www.straitstimes.com/singapore/spore-may-be-1000-years-old-not-just-700-as-believed-study

Zahra, T., & Ahmed, S. (2025). Generational Differences in Emoji Interpretation: A Study of Millennial, Gen Z, and Baby Boomers. *Advance Social Science Archive Journal*, *3*(2), 857–864.

Zahra, T., Ali, P., Qureshi, J., & Fatima, M. (2025a). Emoji Aesthetics in the Digital Age: A Comparative Analysis of User Preferences for Apple, Android Emoji Style. *Sociology & Cultural Research Review*, *3*(2), 214–216.

Zahra, T., Ali, P., Qureshi, J., & Fatima, M. (2025b). Emoji Prediction in Text Based Communication: A Study of Machine Learning Approaches. *Sociology & Cultural Research Review*, *3*(1), 1334–1337.

Zahra, T., & Perono Cacciafoco, F. (2025a). Telegram and WhatsApp in Blended Learning: Impact on Online Community and Academic Achievement Among Graduate Students. *Idil Journal of Art and Language*, *14*, 467–488. https://doi.org/10.7816/idil-14-118-03

Zahra, T., & Perono Cacciafoco, F. (2025b). The Etymological Roots of Emoji Miscommunication: A Comparative Study of Eastern and Western Emoji Usage. *SKASE Journal of Literary and Cultural Studies*, *7*(1), 94–112.

Zahra, T., & Perono Cacciafoco, F. (2025c). Use of Emojis as Discourse Markers in WhatsApp Conversations: A Multimodal Perspective. *Annals of the University of Craiova. Series Philology. Linguistics*, *47*(1–2), 226–237. https://doi.org/10.52846/aucssflingv.v47i1-2.190

Zahra, T., & Perono Cacciafoco, F. (2026). Emoji Rich vs Emoji Poor Communication Across Cultures. *Scientia Fructuosa*, *165*(1), 162–176. https://doi.org/10.31617/1.2026(165)11

Zahra, T., Perono Cacciafoco, F., & Zamir, M. T. (2026). Data-Driven Reconstruction of the Singapore Stone: A Numerical Imputation Method of Epigraphic Restoration. *Information*, *17*(2), 170. https://doi.org/10.3390/info17020170