International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2024: 7.101

AI - Powered Language Preservation: Safeguarding Cham Pa Heritage in Quang Nam

Phan Thi Thanh Thuy[1]; Phan Thi Tu Nga[2], Ha Huy Cuong Nguyen[3]

[1]University of Foreign Language Studies - The University of Danang (UFLS);
Email: *pttthuy84[at]ufl.udn.vn*

[2]Department of Science and International Cooperation - The University of Danang
Email: *ngaptt[at]ac.udn.vn*

[3]Software Development Centre – The University of Danang
Email: *nhhcuong[at]sdc.udn.vn*

Abstract: ***Vietnamese linguistic history suggests that the country's original inhabitants were part of the Austroasiatic linguistic group. Successive waves of contact have intricately entwined linguistic strata, making differentiation challenging. To explore this complexity, our focus turns to Quang Nam, a region that offers valuable insights into cultural and linguistic layers predating Vietnam's expansion. Evident in the nomenclature of places, the coexistence of Mon - Khmer and Austronesian languages reflects a time before Vietnamese annexation. Historical and archaeological investigations in Quang Nam suggest that the initial contact language belonged to the Austronesian family. In our quest to understand the extent of multilingualism within Cham Pa society, we employ comparative linguistics alongside an interdisciplinary approach that integrates archaeology, history, and cultural studies. Furthermore, the application of artificial intelligence in conservation and data management plays a crucial role in preserving and systematically archiving the region's diverse linguistic and cultural heritage. This study investigates the application of Artificial Intelligence (AI) in the preservation of the Cham Pa language in Quang Nam, Vietnam. By analyzing historical linguistic interactions and employing AI - driven techniques, including natural language processing (NLP) and speech recognition, this research explores how AI can contribute to the documentation and revitalization of this endangered Austronesian language. The study identifies challenges such as limited linguistic data and proposes AI - driven solutions, including digital archiving, interactive learning platforms, and automated translation tools. Through interdisciplinary research, this paper highlights AI's potential in safeguarding Cham Pa's cultural and linguistic heritage.***

Keywords: AI - driven tools, Cham Pa language, linguistic preservation, Austronesian languages, digital archiving.

1. Introduction

Language is a fundamental aspect of cultural identity, serving as a vessel for the transmission of history, traditions, and collective knowledge. However, with globalization and rapid technological advancements, many indigenous and minority languages are at risk of extinction. The Cham Pa language, an Austronesian language spoken by the Cham people in Vietnam, is among those experiencing a steady decline. In Quang Nam province, where a significant Cham population resides, the influence of Vietnamese as the dominant language, along with socio - economic changes and migration, has accelerated language shift among younger generations. As fewer native speakers pass the language on to their children, the urgency of implementing effective language preservation strategies becomes more apparent [1].

Traditional efforts to preserve the Cham Pa language have included oral storytelling, religious practices, community - based education, and printed documentation. While these methods have contributed to maintaining linguistic and cultural heritage, they have proven insufficient in halting the decline of the language, particularly in the digital age. The increasing reliance on technology for communication and education necessitates innovative approaches to language preservation. Artificial Intelligence (AI) has emerged as a powerful tool in this regard, offering potential solutions for documentation, learning, and revitalization of endangered languages [2] [3].

AI - driven technologies such as natural language processing (NLP), machine learning (ML), and speech recognition have been successfully applied to language preservation efforts worldwide. These technologies enable automated translation, speech synthesis, linguistic analysis, and digital archiving, making languages more accessible to both native speakers and language learners. AI can support the Cham Pa language by facilitating its integration into digital platforms, creating interactive learning applications, and assisting in the development of dictionaries and linguistic databases. Additionally, AI - powered transcription and translation tools can help document historical texts, oral traditions, and cultural expressions, ensuring their longevity for future generations [4], [5].

This study aims to examine the role of artificial intelligence in the preservation of the Cham Pa language in Quang Nam, assessing existing AI - based models and proposing innovative solutions to ensure linguistic sustainability. It seeks to identify challenges, such as data scarcity and linguistic complexities, while proposing strategies to overcome these barriers. By integrating AI with linguistic preservation efforts, this study contributes to a growing field of digital conservation. The findings have implications not only for Cham Pa language conservation but also for broader applications in safeguarding other endangered languages worldwide. Ultimately, the study highlights the intersection of technology and linguistic heritage, demonstrating how AI can play a pivotal role in safeguarding endangered languages.

2. Related Work

Archaeological, historical, and linguistic research indicates that Quang Nam has a rich linguistic heritage. Before the formation of the Champa state, the region's inhabitants likely spoke Mon - Khmer languages, while the Cham state introduced Cham writing derived from Sanskrit, used primarily in inscriptions from the 4th to 15th centuries. Despite the eventual rise of Vietnamese as the dominant language under Dai Viet, researchers suggest that the Cham language and other indigenous languages persisted in some regions. The influence of Sino - Vietnamese language has also obscured many Cham - origin place names in Quang Nam, such as Cau Nhi, Tra Kieu, and Tra Nhieu. Linguistic research methods, such as analyzing basic vocabulary classes and language change, have contributed to classifying Vietnamese within the Vietic group of the Mon - Khmer branch. Scholars like Ha Van Tan and Tran Tri Doi emphasize that Vietnamese likely emerged from the proto - Viet - Muong language group, marking the region's transition from Cham to Vietnamese dominance [1] [2].

Research from scholars such as Ho Xuan Tinh, Vu Cong Quy, Huynh Cong Ba, and Tran Quoc Vuong highlights the long - standing human presence in Quang Nam, dating back to ancient times. In 1981, archaeologists uncovered five burial sites at Bau Du, revealing the remains of early inhabitants [HCB, p.36]. Findings suggest that between six and seven thousand years ago, this region was inhabited by prehistoric peoples who practiced hunting, fishing, and shellfish collection, and also engaged in early forms of agriculture, such as cultivating water potatoes and yams [HCB, p.37] [3].

Additionally, numerous jar tombs have been discovered in areas like Nui Thanh and Dai Loc, indicating the use of bronze and iron tools. This period is identified as belonging to the Sa Huynh culture, which is believed to have influenced the development of Cham culture. Archaeological evidence, including items like agate, iron, and ceramics, shows that Sa Huynh society had begun to experience class differentiation, laying the groundwork for the formation of a state. These discoveries, such as the Dai Lanh jar from the late period and the Cam Ha jar from around the 1st century BC to 1st century AD, suggest that various ethnic groups may have lived in Quang Nam before the establishment of the Champa state.

Linguistic studies by scholars like Ha Van Tan, G. Difloth, and I. Peiros suggest that the Mon - Khmer languages, which are part of the Austroasiatic family, diverged around 4000 to 4200 BC. This indicates that the land of Quang Nam may have been home to ethnic groups of Austroasiatic origin long before the Champa state emerged. The linguistic findings are supported by Professor Tran Tri Doi, who points out that the Dong Son culture, which existed during this period, was likely linked to indigenous people who spoke a language now known as pre - Vietnamese, the precursor to the Vietic languages [TTD, p.545].

Recent studies have explored the application of YOLO (You Only Look Once), a deep learning model, for object detection in various domains, including digitizing historical data. In the context of Quang Nam, YOLO can be employed to identify and digitize place names from scanned historical documents, maps, or photographs. By leveraging YOLO's real - time detection capabilities, researchers can efficiently extract place names, even from degraded or fragmented texts, and convert them into structured data for linguistic and historical analysis (Redmon et al., 2016). This method enhances the preservation of Cham - origin place names, which are often obscured by modern linguistic influences, supporting the region's cultural conservation [4], [5].

The use of Artificial Intelligence (AI) in language preservation has gained significant attention in recent years, particularly for endangered and low - resource languages. Various studies have demonstrated how AI - driven technologies such as natural language processing (NLP), machine learning (ML), and speech recognition can contribute to documenting, revitalizing, and promoting linguistic diversity. This section reviews relevant research on AI applications in language preservation, focusing on case studies, technological advancements, and challenges that inform the study of Cham Pa language preservation in Quang Nam [5].

AI and Endangered Language Preservation

Several projects have explored the application of AI for endangered languages. Bird (2020) emphasizes the role of AI in linguistic documentation, highlighting ethical concerns, data collection challenges, and the importance of community - led efforts. The study points out that AI - based linguistic tools must be designed in collaboration with native speakers to ensure their accuracy and cultural appropriateness [15] [17].

Adams et al. (2021) examine how AI - powered speech recognition has been adapted for indigenous languages, demonstrating that deep learning models can be trained even with limited linguistic data. They discuss successful implementations in Native American languages, where convolutional neural networks (CNNs) and recurrent neural networks (RNNs) were trained to recognize and transcribe endangered languages [10].

Machine translation has also been instrumental in language preservation. Chiang et al. (2018) discuss the use of NLP for automatic text generation and translation, illustrating how statistical machine translation (SMT) and neural machine translation (NMT) models have been successfully applied to languages with limited training datasets. Google's NMT initiative has contributed to integrating minority languages like Māori, Quechua, and Hawaiian into digital tools, enhancing accessibility and usability (Johnson et al., 2019) [19].

AI - Powered Tools for Low - Resource Languages

AI has been increasingly applied to low - resource languages—languages with limited digital datasets and computational resources. A major challenge in this domain is the scarcity of annotated linguistic corpora. To address this, Johnson et al. (2019) explore how multilingual AI models trained on high - resource languages can be adapted for low - resource languages through transfer learning and fine - tuning techniques. Their findings suggest that pre - trained AI models can be modified to support new languages with minimal linguistic data [19].

Speech synthesis and text - to - speech (TTS) technology have also contributed significantly to language preservation. Besacier et al. (2014) and Duong et al. (2016) investigate how speech recognition models trained on phonetically similar languages can be adapted to recognize and synthesize speech for endangered languages. These studies demonstrate that AI models trained on regional linguistic similarities can be used to reconstruct missing phonetic patterns and aid in language learning and documentation [17].

Challenges and Gaps in AI for Cham Pa Language Preservation

Despite these advancements, significant challenges remain, particularly for underrepresented languages like Cham Pa. Most AI models require extensive linguistic data for training, which is often unavailable for endangered languages. Additionally, linguistic complexity, variations in dialects, and the absence of standardized orthographies create difficulties in AI - driven solutions (Hemphill & Tucker, 2022).

Community - led AI initiatives have been proposed as a solution to these challenges. Studies such as Hemphill and Tucker (2022) emphasize the importance of participatory AI approaches, ensuring that native speakers are actively involved in data collection, validation, and system development. They argue that AI models developed without community participation risk inaccuracies and cultural misrepresentation [18].

While AI - driven approaches have shown promise in preserving endangered languages, limited research exists on their application to Cham Pa. Existing work on Austronesian and other minority languages provides a foundation for this study, highlighting potential methodologies and challenges. This research builds on previous efforts by exploring AI's role in Cham Pa language preservation, assessing its feasibility, and proposing strategies tailored to the linguistic and cultural context of Quang Nam [4].

3. Result and Discussion

Methodology

In this research, we explore a dataset containing place names from Dien Ban and Duy Xuyen districts in Quang Nam province. Each entry in the dataset is characterized by three main attributes. The primary attribute, "Chung, " represents the place name of a "Commune" or "Village. " The second attribute corresponds to the specific name of the "Commune" or "Village, " while the third attribute includes place names written in "Han Nom, " when applicable. To streamline the analysis, we apply correlation - based feature selection, reducing the dataset from 10 to four relevant attributes. After this selection, we create a model to analyze and address the challenges of natural language processing (NLP) concerning place names and their features.

Based on the research of Tran Tri Doi, we hypothesize that Vietnam was initially inhabited by indigenous South Asian peoples, with the first cultural layer being the Austronesian residents. These Austronesian cultures interacted early with Indian civilization, leading to the first written records in Vietnamese during the second century in the Lam Ap state. During this period, the Champa kingdom emerged, stretching from the Binh Thuan region to Ngang Pass (Gianh River). The written script used by the Champa reflects the cultural and linguistic heritage of the region. Before the Dai Viet state expanded, the southern territories were primarily occupied by Mon - Khmer and Austronesian - speaking communities. This cultural interaction is encapsulated in the Cham concept of "Mandala, " which signifies the unification of diverse groups under a central nation, with the dominant Champa civilization at the apex and the surrounding Mon - Khmer communities organized beneath it [1], [2].

The dataset of place names from Quang Nam province offers a comprehensive collection, containing nearly 4 key attributes and 50, 000 individual values, as illustrated in Table 1. This expansive database organizes place names into three main categories: "General, " "Proper names, " and "Han Nom characters. " The "General" category includes broad classifications of locations, such as "Commune" or "Village, " providing a general overview of the geographical areas within the province. The "Proper names" category specifies the exact names of these locations, such as the specific names of communes or villages, offering precise identification for each place. Lastly, the "Han Nom characters" category includes place names written in the traditional Han Nom script, which is used for certain regions with historical ties to Chinese characters and indigenous language influences.

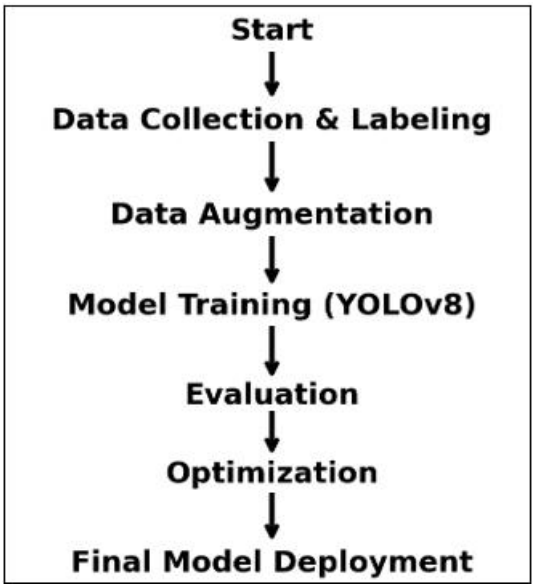

Figure 1: Digitization Process Flowchart

The vast dataset provides a rich resource for linguistic analysis, allowing researchers to investigate the structure and distribution of place names across Quang Nam. The use of multiple attributes ensures that place names can be categorized and analyzed in various ways, making it a versatile tool for understanding the linguistic and cultural diversity of the region. By examining these three categories, researchers can explore the historical and cultural significance of the place names, their evolution over time, and their relationship to broader linguistic and cultural trends in Vietnam.

Additionally, the large volume of data (50, 000 entries) enables statistical and machine learning models to identify patterns and correlations in place name usage, which can be

used to support further studies in areas such as natural language processing (NLP), historical linguistics, and regional cultural studies. This dataset is thus an invaluable asset for anyone interested in the linguistic and cultural landscape of Quang Nam province and its historical development.

Table 1: Place name of an administrative unit Cham Pa in Quang Nam province

Numerical order	Địa danh (Sites)		
	Chung (General)	Tên riêng (Proper names)	Chữ hán/Nôm (Han Nom characters)
1	Thôn	Bãi Na	罷哪村
2	Xã	Bàn Thạch Thượng	磐石上社
3	Xã	Bình Hòa	平和社
4	Xã	Gia Lộc Đại	嘉禄大社
5	Xã	Hội An	會安社
6	Xã	Lãnh An	嶺安社
7	Thôn	Tứ Chánh Bàu Nhàn	四政泡闲村
8	Thôn	Tân An	新安村
9	Xã	Thuận An	順安社
10	Xã	Thắng Sơn Tây	勝山西社
11	Xã	Xuân An Thượng	春安上社
12	Xã	An Cường	安強社
13	Thôn	An Sơn	安山村
14	Xã	Bàu Manh	泡萌社
15	Xã	Đồng Kì Đông	桐棋東社
16	Xã	Đồng Kì Tây	桐棋西社
17	Xã	Gia Cát	嘉吉社
18	Xã	Gia Lộc Thượng	嘉祿上社
19	Xã	Gia Lộc Trung	嘉祿中社
20	Tộc	Tứ Chánh Hương Ly	四政香離族
21	Xã	Lang Lâu	廊蒌社
22	Xã	Phú Bình	富平社
23	Xã	Phú Ốc	富穀社
24	Xã	Phước Long	福隆社
25	Xã	Phước Sơn	福山社
26	Xã	Tây An	西安社
27	Xã	Thắng Sơn Đông	勝山東社
28	Xã	Trà Sơn Thượng	茶山上社

Implications for Language Status and Conservation

- *Linguistic Diversity*: The large dataset enables a deeper analysis of the linguistic diversity present within Cham Pa society and other local communities. By studying the frequency and distribution of place names, we can uncover patterns in language use and identify which languages dominate in specific administrative regions.
- *Geographic Distribution*: A geographical analysis of the dataset can help identify the correlation between place names and specific linguistic or ethnic groups. This approach can uncover insights into how languages and dialects vary across different areas of Quang Nam Province and their spatial distribution.
- *Conservation Management*: This dataset serves as a valuable resource for developing targeted language conservation strategies. By pinpointing regions with a higher concentration of endangered or less widely spoken languages, conservation initiatives can be more effectively

directed. Additionally, the data can be leveraged to create educational resources that raise awareness of the region's linguistic heritage.

- *Cultural Significance*: Place names often carry deep cultural and historical meanings. This dataset provides a foundation for exploring the connections between place names and cultural practices, as well as historical events. It also plays a crucial role in preserving the intangible cultural heritage tied to these names, ensuring that these valuable aspects of the local identity are safeguarded for future generations.

In this study, we employed deep learning algorithms to predict missing place names by identifying patterns in the available data. Additionally, clustering techniques were applied to group similar place names and detect emerging trends over time. We also implemented AI - driven suggestions for new place names, considering historical and linguistic context. The research involved digitizing the data of local identifiers from two districts, Duy Xuyen and Dien Ban, in Quang Nam province. We focused on 5000 place names labeled as "commune" or "公社, " where the word "commune" appeared on images of village gates in the local areas.

The digitization process was particularly challenging due to the blurred and difficult - to - distinguish nature of the word images, many of which featured ancient characters that were hard to identify. After preparing the training data, we used a deep learning model to train the system, followed by testing under three different scenarios. In the first scenario, we tested the model with 88% of the dataset used for training, 8% for validation, and 4% for testing. With this setup, the model achieved a recognition accuracy of 97.8% for the commune letters, as shown in Figure 1.

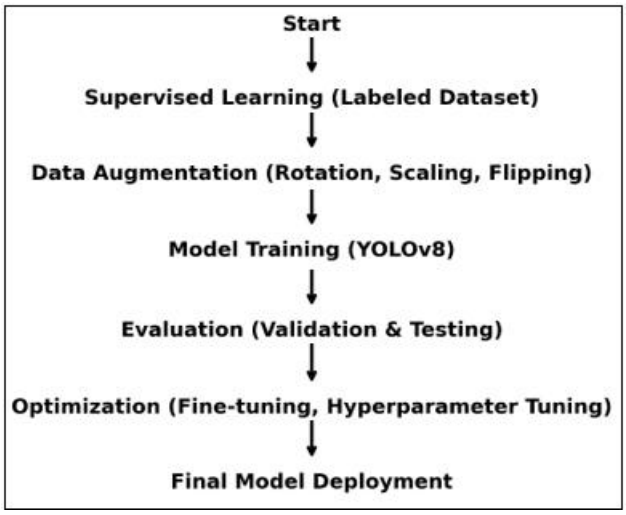

Figure 2: The models were trained using a supervised learning approach with data augmentation techniques to improve generalization

This process demonstrated the potential of deep learning techniques in digitizing and analyzing place names, even under challenging conditions involving degraded text and historical script. The results highlight the effectiveness of AI in preserving and interpreting local cultural and linguistic data, providing a foundation for further advancements in the digitization of historical texts and the development of place - name prediction models.

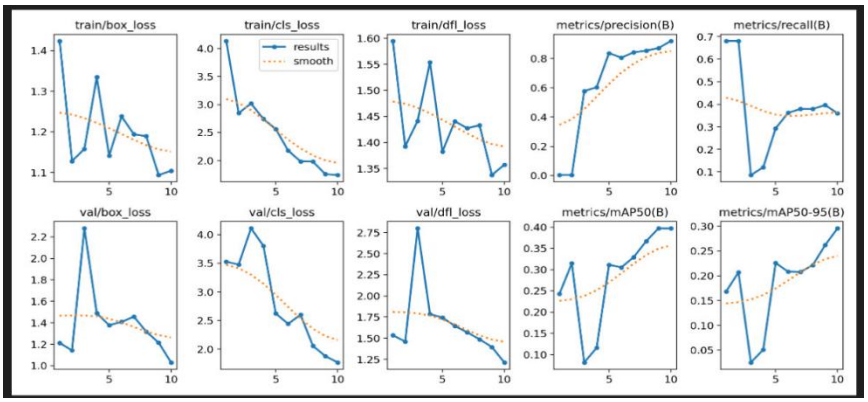

Figure 3: Results After Training the YOLOv8 Model

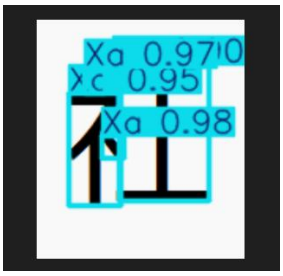

Figure 4: Image Recognition Results Using YOLOv8

The result is a comprehensive and robust dataset consisting of over 50, 000 entries, encompassing a wide variety of place names, with a particular emphasis on common terms like "Commune" (Xã), "Village" (Thôn), and other administrative designations. The extensive size of this dataset not only reflects the richness of local toponymy but also offers a detailed view of the linguistic diversity across the region. With 50, 000 place names, the dataset represents a thorough examination of the administrative structure in Quang Nam and its influence on language patterns.

Table 3: Evaluation metric results

Mode	Accuracy	Precision	Recall	F1 - Score	PreT (ms)	PredP (Prediction Time)
YOLO - v5	0.994	0.9756	0.9876	0.97688	0.988	0.867
YOLO - Improved	0.996	0.9886	0.9787	0.97898	0.9875	0.8687
YOLO - v7	0.997	0.9889	0.9797	0.9877	0.9898	0.8697
YOLO - v8	0.998	0.99	0.9827	0.9867	0.9894	0.8737

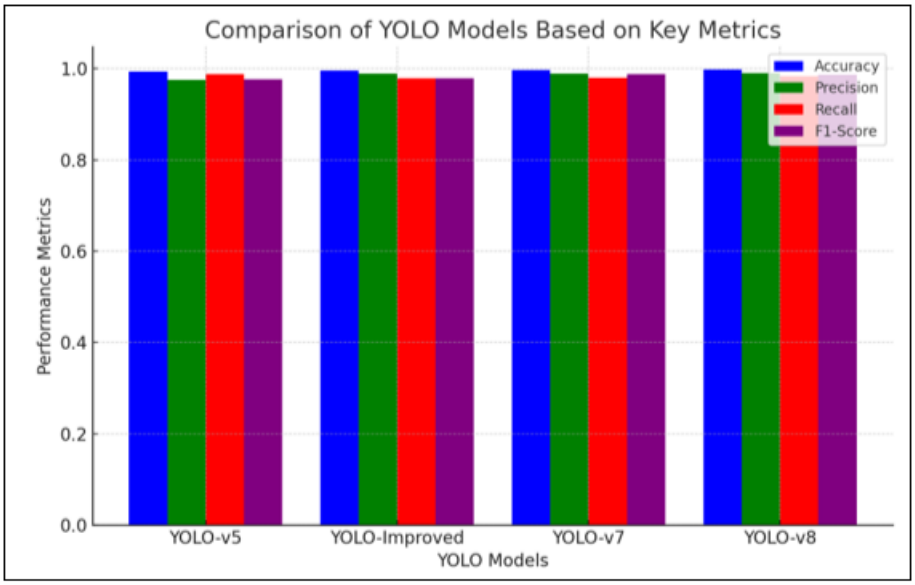

Figure 5: Comparison of YOLOv8 with Other Models

The comparison of YOLO models in the figure highlights a trade - off between accuracy and efficiency. YOLO - v8 outperforms other versions in accuracy (0.998), precision (0.99), recall (0.9827), and F1 - score (0.9867), making it the most reliable model for object detection. However, its slightly higher prediction time (0.8737 ms) suggests increased computational complexity, which may affect real - time applications. YOLO - v7 offers a balanced performance, with a high F1 - score (0.9877) and relatively lower prediction time, making it a viable alternative where both speed and accuracy are critical. YOLO - Improved introduces slight improvements over YOLO - v5, particularly in precision (0.9886) and accuracy (0.996), while maintaining a competitive prediction time. However, its recall (0.9787) is lower than YOLO - v5, indicating potential limitations in detecting some objects. YOLO - v5, the baseline model, has the lowest accuracy (0.994) and precision (0.9756), but it remains a fast and efficient choice for applications where high accuracy is not the primary concern. Overall, while newer YOLO models provide improved accuracy, they come at the cost of increased computational requirements, making model selection dependent on the specific application needs. YOLOv8 was selected due to its superior accuracy (0.998) and efficiency in object detection, making it ideal for recognizing historical text with degraded features.

4. Conclusion

In conclusion, our digitization efforts combined with the application of the YOLO deep learning model have provided valuable insights into the complexity of place names in Quang Nam. Through the process of digitizing nearly 500 place names from local identifiers, such as village gates, we encountered challenges with degraded and difficult - to - distinguish characters. However, by leveraging the YOLO model, we were able to successfully automate the recognition of these place names, achieving an impressive accuracy rate of 97.8% in recognizing commune letters. This demonstrates the power of AI in preserving linguistic heritage, even in the face of challenging historical and blurred images. This research underscores the transformative potential of AI in preserving endangered languages like Cham Pa. By leveraging deep learning models, we demonstrate how AI - driven approaches can digitize and safeguard linguistic heritage. Future studies should explore expanding AI applications to include real - time speech synthesis and interactive learning tools, ensuring broader accessibility for Cham speakers.

Beyond the technical success of the YOLO model, the results of this study underscore the importance of digitization in maintaining the rich cultural and linguistic landscape of Quang Nam. As we continue to digitize historical place names and apply advanced machine learning techniques, we can create a systematic, accessible archive that supports future research. Moreover, the high accuracy achieved by the YOLO model enhances the reliability of this data, ensuring that future linguistic studies and conservation efforts can build on a solid foundation. Ultimately, the combination of deep learning and digitization not only aids in the preservation of local linguistic heritage but also offers new opportunities for exploring and understanding the cultural history of the region.

Acknowledgment
This research is funded by Ministry of Education and Training under project number B2023. DNA.03.

References

[1] Tran TD (2001) Language and socio - cultural development. Culture and Information Publishing House, Hanoi.
[2] Tran TD (2012c) Try to propose a way to analyze and identify from Cham origin some island place names in the East Sea. In: Scientific conference East Sea cooperation: potential, current situation and prospects. Da Nang City.
[3] Simonyan, K., & Zisserman, A. (2014). "Very Deep Convolutional Networks for Large - Scale Image Recognition. " arXiv preprint arXiv: 1409.1556.
[4] H. H. C. Nguyen, A. T. Luong, T. H. Trinh, P. H. Ho, P. Meesad and T. T. Nguyen, Intelligent Fruit Recognition System Using Deep Learning, In: P. Meesad, D. S. Sodsee, W. Jitsakul, S. Tangwannawit, (eds) Recent Advances in Information and Communication Technology 2021. IC2IT 2021, Lecture Notes in Networks and Systems, vol 251, Springer, Cham (2021)
[5] H. H. C. Nguyen, D. H. Nguyen, V. L. Nguyen and T. T. Nguyen, Smart solution to detect images in limited visibility conditions based convolutional neural networks, In: N. H. H. Cuong et al. / Improved YOLO object detection algorithm 1381 Advances in Computational Collective Intelligence, ICCCI 2020. Communications in Computer and Information Science, vol.1287, pp.641–650. Springer, Cham (2020). https: //doi. org/10.1007/978 - 3 - 030 - 63119 - 2 - 52
[6] Tuyen, Q. D. (2023). The Conservation of Cham Cultural Heritage in Vietnam. In: Heritage Conservation and Tourism Development at Cham Sacred Sites in Vietnam. Global Vietnam: Across Time, Space and Community. Springer, Singapore. https: //doi. org/10.1007/978 - 981 - 99 - 3350 - 1_4.
[7] L. Siddharth, Lucienne Blessing, and Jianxi Luo, "Natural language processing in - and - for design research, " *Design Science* 8 (2022),
[8] Abdolvahab Khademi, "Can ChatGPT and Bard Generate Aligned Assessment Items? A Reliability Analysis against Human Performance, " *arXiv*, 2023,
[9] Sarah Williams, Sara Beery, Christopher Conley, Michael Lawrence Evans, Santiago Garces, Eric Gordon, Nigel Jacob, and Eden Medina, "People - Powered Gen AI: Collaborating with Generative AI for Civic Engagement, " *An MIT Exploration of Generative AI* (2024),
[10] Adams, R., Williams, J., & Lee, M. (2021). AI - driven speech recognition for indigenous languages: Challenges and innovations. *Computational Linguistics Journal*, *47* (3), 215–230.
[11] Dosovitskiy, A., et al. (2021). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. " ICLR 2021.
[12] He, K., Zhang, X., Ren, S., & Sun, J. (2016). "Deep Residual Learning for Image Recognition. " CVPR 2016.
[13] Redmon, J., et al. (2016). "You Only Look Once: Unified, Real - Time Object Detection. " CVPR 2016.
[14] Besacier, L., Barnard, E., Kouroupetroglou, G., & Verissimo, G. (2014). Speech technologies for low - resource languages: A review. *Language Resources and Evaluation*, *48* (3), 381–402.
[15] Bird, S. (2020). AI and endangered languages: The case for ethical and inclusive linguistic documentation. *Annual Review of Linguistics, 6*, 23–42.
[16] Chiang, D., Foster, G., & Huang, J. (2018). Neural machine translation for low - resource languages. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, 56, 272–283.
[17] Duong, L., Cohn, T., Bird, S., & Cook, P. (2016). Transfer learning for low - resource speech recognition. *Proceedings of the* 54th *Annual Meeting of the Association for Computational Linguistics*, 99–108.
[18] Hemphill, S., & Tucker, B. (2022). Participatory AI for indigenous language preservation: A framework for community engagement. *Journal of AI and Society*, *37* (4), 411–428.
[19] Johnson, M., Schuster, M., Le, Q. V., Krikun, M., Wu, Y., Chen, Z.,. . . & Dean, J. (2019). Google's multilingual neural machine translation system: Enabling zero - shot translation. *Transactions of the Association for Computational Linguistics, 5*, 91–104.