TẠP CHÍ KHOA HỌC
TRƯỜNG ĐẠI HỌC QUY NHƠN

Hiểu rõ hơn về khoáng sản: chương trình phân loại khoáng sản nâng cao tích hợp AI giải thích được và mô hình ngôn ngữ lớn

Nguyễn Trương Thành Hưng[1], Trương Thị Cẩm Mai[2,*]

[1]Phòng thí nghiệm Analytics Everywhere, Trường Đại học New Brunswick, Canada
[2]Khoa Khoa học tự nhiên, Trường Đại học Quy Nhơn, Việt Nam

Ngày nhận bài: 29/08/2023; Ngày sửa bài: 24/10/2023;
Ngày nhận đăng: 27/10/2023; Ngày xuất bản: 28/10/2023

TÓM TẮT

Khoáng sản, với thành phần hóa học phức tạp và cấu trúc tinh thể, đóng một vai trò then chốt trong nhiều quá trình hóa học, ứng dụng, và nghiên cứu. Truyền thống, việc phân loại chúng được thực hiện thông qua các kỹ thuật quan sát và hóa học. Tuy nhiên, với việc tăng số lượng mẫu, các phương pháp này thường mất nhiều thời gian. Những tiến bộ gần đây trong Trí tuệ nhân tạo (AI) và Học sâu (DL) hứa hẹn những cải tiến đột phá về tốc độ và độ chính xác của việc phân loại khoáng sản. Tuy nhiên, các mô hình DL, mặc dù chính xác, thường hoạt động như những "hộp đen", làm cho quyết định của chúng không tường minh. Để giải quyết điều này, nghiên cứu của chúng tôi giới thiệu một khung chương trình dựa trên AI cho việc phân loại khoáng sản, kết hợp các mô hình tiên tiến với AI Giải thích được (XAI) và mô hình AI sinh ngôn ngữ lớn (LLMs) như GPT-4. Chương trình này không chỉ phân loại một số lượng lớn các khoáng sản mà còn giải thích lý do phía sau mỗi lựa chọn phân loại. Thông qua sự kết hợp của mô hình Swin Transformer V2 cho việc nhận dạng khoáng sản, GradCAM cho tính minh bạch của mô hình, và GPT-4 để truy xuất thông tin khoáng sản chi tiết, chương trình cung cấp sự kết hợp cân đối giữa hiệu suất, khả năng giải thích và thông tin hướng tới người dùng. Chương trình có thể được truy cập công khai, nhấn mạnh tiềm năng của AI trong việc cách mạng hóa việc phân loại khoáng sản trong khi vẫn đáp ứng nhu cầu về sự rõ ràng, minh bạch và giáo dục người dùng. Đường dẫn truy cập công khai tại *https://huggingface.co/spaces/minatosnow/mineral_framework.*

Từ khóa: *Phân loại khoáng sản, AI giải thích được, mô hình AI sinh ngôn ngữ lớn.*

**Tác giả liên hệ chính.*
Email: truongcammai@qnu.edu.vn

QUY NHON UNIVERSITY
JOURNAL OF SCIENCE

Understanding minerals better: advancing mineral classification framework through explainable AI and large language model integration

Truong Thanh Hung Nguyen[1], Thi Cam Mai Truong[2,*]

[1]Analytics Everywhere Lab, University of New Brunswick, Canada
[2]Faculty of Natural Sciences, Quy Nhon University, Vietnam

Received: 29/08/2023; Revised: 24/10/2023;
Accepted: 27/10/2023; Published: 28/10/2023

ABSTRACT

Minerals, with their intricate chemical compositions and crystalline structures, play a pivotal role in diverse chemical processes, applications, and research. Traditionally, their classification was achieved through observational and chemical techniques. However, with increasing sample sizes, these methods often proved time-consuming. Recent advances in Artificial Intelligence (AI) and Deep Learning (DL) promise transformative improvements in the speed and accuracy of mineral classification. However, DL models, for all their precision, often operate as "black boxes", making their decision-making opaque. To address this, our study introduces an innovative AI-powered framework for mineral classification, integrating state-of-the-art models with Explainable AI (XAI) and generative AI large language models (LLMs) like GPT-4. This framework not only categorizes a wide-ranging number of minerals but also elucidates the reasoning behind each classification. Through a combination of Swin Transformer V2 models for mineral identification, GradCAM for model transparency, and GPT-4 for detailed mineral information retrieval, the framework offers a balanced blend of performance, interpretability, and user-centric information. Available for public access, this system underscores the potential of AI to revolutionize mineral classification while staying attuned to the demands of clarity, transparency, and user education. The framework can be publicly accessed via *https://huggingface.co/spaces/minatosnow/mineral_framework.*

Keywords: *Mineral classification, explainable AI, generative AI large language models.*

1. INTRODUCTION

Minerals are naturally occurring inorganic substances with a specific chemical composition and crystalline structure.[1] Mineral classification is the systematic categorization of minerals based on their physical and chemical properties.[2,3] This classification provides detailed insights into the chemical composition and structure of minerals. By categorizing minerals, chemists can predict their behavior, reactivity, and stability.[4] This understanding is fundamental for various chemical processes, including synthesis, analysis, and industrial applications. Mineral classification is not only an academic exercise but also a vital practice in the chemical field. It underpins various industrial processes, medical applications, environmental protection, and research endeavors. Its importance continues to grow with the increasing complexity and specialization of chemical products and

**Corresponding authors.*
Email: truongcammai@qnu.edu.vn

https://doi.org/10.52111/qnjs.2023.17510

processes, making it an indispensable aspect of modern chemistry.[2,3,5]

Traditionally, mineral classification has been carried out through a combination of physical observation and chemical analysis.[6] Regarding the physical properties, minerals are often classified based on their hardness, luster, color, streak, and specific gravity. The Mohs scale, for example, is used to classify minerals based on hardness.[7,8] Minerals can be grouped into classes based on their primary anionic species, such as silicates, carbonates, and sulfates.[9,10] Chemical tests, such as flame tests and wet chemical analysis, are used to identify the presence of specific elements or compounds.[6,11] X-ray diffraction and other microscopic techniques are also employed to analyze the crystalline structure of minerals, further categorizing them into specific groups.[12–14] Additionally, another approach is to use polarizing microscopes to study the optical properties of minerals, such as birefringence and pleochroism, which can be essential for classification.[15–17]

However, conventional methods might be labor-intensive and time-consuming, particularly when dealing with a large number of samples. With the advent of Artificial Intelligence (AI) and Deep Learning (DL), the field of mineral classification has witnessed a significant transformation.[18–20] The application of DL techniques to mineral classification on images has opened new avenues for accurate and automated classification. DL models can easily scale to handle vast datasets, providing rapid classification without compromising accuracy.

Nevertheless, DL models, particularly complex neural networks (NNs), are often referred to as "black boxes" due to their lack of transparency in how they arrive at a particular decision.[21–24] While these models can achieve high accuracy, understanding the specific reasoning behind their decisions can be elusive. This lack of transparency poses significant challenges, particularly in understanding the rationale behind specific classifications and in ensuring trust and compliance with regulatory standards. Consequently, there is a growing imperative for the integration of Explainable AI (XAI) methods, which aim to unravel the intricate workings of DL models, providing insights into their decision-making processes.[25,26] Besides that, recent works in generative AI large language models (LLMs) have shown promising results in generating human-like text that can be leveraged to provide more information and facts about the model's decisions.[27]

Hence, in this paper we propose an AI-assisted mineral classification framework leveraging several state-of-the-art models in a multi-class classification task integrated with XAI techniques and generative AI LLMs. This integration not only enhances the interpretability of mineral classification but also provides clear and plausible insights into the decision-making process for the end-users. Our proposed framework is tailored to meet the specific needs of the chemical field, ensuring that the classifications are both scientifically robust and readily interpretable. Through our framework, we aim to address the critical challenge of transparency in AI-driven mineral classification, offering a solution that balances performance with interpretability, and understandability, tailored to the unique requirements of the chemical domain. The framework can be publicly accessed via *https://huggingface.co/spaces/minatosnow/mineral_framework*.

2. RELATED WORK

2.1. Deep learning in mineral classification

DL has emerged as a powerful tool in the field of mineral classification, leveraging the ability to learn complex patterns and relationships directly from data, which has been greatly facilitated by the availability of large datasets, powerful computing resources, and the development of sophisticated algorithms.[18–20]

Convolutional Neural Networks (CNNs)[28] are deeply structured feedforward NNs and one

of the representative algorithms of DL, which can be applied to automatically extract optical features of minerals for mineral identification or accelerate the microphase classification. A hybrid approach combining mineral photo image features extracted by CNN EfficientNet-b4 and mineral hardness features to identify minerals.[7] U-Net model is utilized to effectively and automatically extract deep feature information of ore minerals, realizing intelligent recognition and classification under the microscope.[29] ResNet-18 and ResNet-50 models is proposed for DL-based intelligent mineral recognition, enhancing data with image flipping and scale transformation.[30–32]

However, challenges related to interpretability and data dependence remain, where the generated models are complex and difficult to interpret and good accuracy is only guaranteed when the amount of data is large enough, limiting the application in scenarios with limited data, calling for further research and innovation in the field.[20,33]

2.2. Swin transformer – hierarchical vision transformer using shifted windows

Given the aim of our research to classify images according to their corresponding mineral specimen, we undertake this endeavor within the paradigm of image classification-a canonical yet persistently demanding task within the domain of computer vision (CV). For this purpose, we have chosen to utilize a leading-edge model known as the Swin Transformer. The Swin Transformer is a hierarchical vision transformer characterized by its use of shifted windows to compute its representations.[34] This model has been meticulously crafted to navigate the inherent challenges of transposing transformers from linguistic contexts to visual ones. These challenges encompass the vast disparities in scale among visual entities and the inherent high resolution of pixels in images, which stand in stark contrast to the relative simplicity of words within a textual context. The deployment of a shifted windowing scheme serves a dual purpose: it enhances computational efficiency by restricting self-attention computations to discrete, non-overlapping local windows, and concurrently, it facilitates cross-window connections. The hierarchical nature of this architecture bestows upon it the versatility to operate across multiple scales, all while maintaining linear computational complexity in relation to image size. Such attributes render the Swin Transformer a suitable candidate for an array of vision tasks, spanning from image classification to object detection and semantic segmentation.[34] These qualities make Swin Transformer compatible with a broad range of vision tasks, including image classification, object detection, and semantic segmentation.[35]

Furthermore, Swin Transformer V2 represents a sophisticated evolution of the original Swin Transformer model, with an emphasis on augmenting both its capacity and resolution.[36] The associated paper addresses three predominant challenges encountered during the training and application of expansive vision models: training instability, discrepancies in resolution between the stages of pre-training and fine-tuning, and an acute dependence on labeled data. To rectify these issues, the authors propose three primary strategies: 1) The combination of a residual-post-norm approach with cosine attention to bolster training stability; 2) The introduction of a log-spaced continuous position bias method, facilitating the seamless transference of models pre-trained on low-resolution images to downstream tasks necessitating high-resolution inputs; and 3) The deployment of a self-supervised pre-training technique named SimMIM, which mitigates the requirement for vast repositories of labeled images. Leveraging these strategies, the researchers were successful in training a Swin Transformer V2 model comprising a staggering 3 billion parameters, marking its

https://doi.org/10.52111/qnjs.2023.17510

position as one of the most voluminous dense vision models presently available. Impressively, this model has established new benchmarks in performance across four cardinal vision tasks: ImageNet-V2 image classification, COCO object detection, ADE20K semantic segmentation, and Kinetics-400 video action classification.[36,37]

2.3. Explainable AI

XAI is a field of research that aims to make the decisions and predictions of AI systems more transparent and interpretable to humans. There are several approaches to achieving this goal, including gradient-based, perturbation-based, and Class Activation Mapping (CAM)-based methods.

Gradient-based methods, such as LRP,[38] use gradient signals to assign the burden of the decision on the input features. These techniques can be evaluated for their robustness and the role that adversarial robustness plays in having meaningful explanations.

Perturbation-based methods investigate properties of deep neural networks (DNNs) by perturbing the input of a model. For example, part of the input image can be occluded with a mask or a word in a sentence can be replaced with its synonym, and the changes in the output of the model can be observed. Some notable perturbation-based methods are LIME,[23] RISE, D-RISE,[39] D-CLOSE.

CAM-based methods, such as CAM,[40] GradCAM,[41] GradCAM++, SeCAM,[24,42] ScoreCAM,[43] are visual explanation techniques that use class activation maps to highlight the regions of an input image that are most relevant to the model's prediction.

In this work, we employ GradCAM[41] for model debugging and to make CNN-based models more transparent to end-users, primarily in visual tasks like image classification. By visualizing the important regions in an image as a high-resolution heatmaps, developers and end-users can better understand if a model is focusing on the correct patterns or perhaps getting misled by noise or other irrelevant features. GradCAM offers easily interpretable visualizations that align well with human intuition.[44]

2.4. Generative AI with large language models

In the arena of AI, generative AI LLMs have garnered significant attention. Such models, underpinned by extensive datasets, possess the aptitude to synthesize text that is strikingly analogous to human-authored content. One of the most distinguished models in this domain is the Generative Pre-trained Transformer (GPT), a brainchild of OpenAI.[45] GPT has seen several iterations, with the latest being GPT-4.[46] In parallel, Llama 2 has emerged as a notable LLM, a product of collaborative efforts between Meta and Microsoft. This model stands out due to its

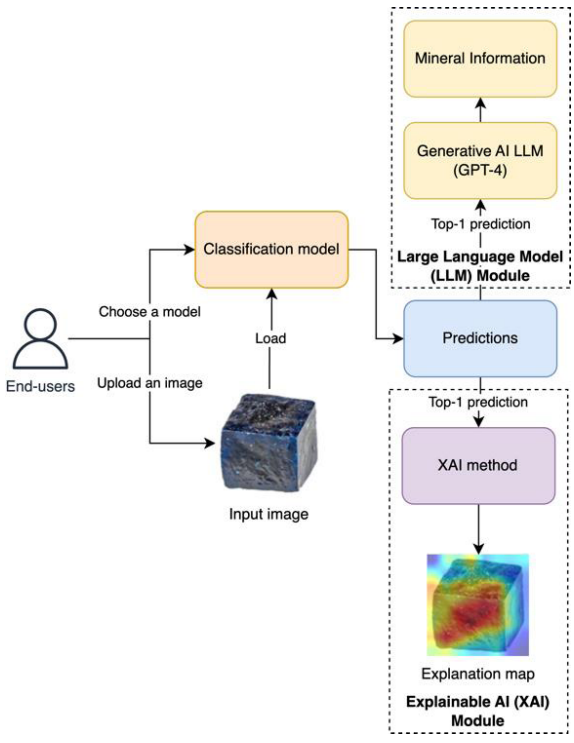


Figure 1. The flowchart representation of the proposed mineral classification framework. After the classification model receives the input image loaded by end-users, the top-1 prediction is fed into the XAI method to deliver the explanation map, and into the generative AI LLM to give information and facts about the classified mineral.

training on a contemporaneous and more eclectic dataset. Claude 2, heralded by Anthropic, is another LLM worth mentioning, boasting enhanced performance, safety, harmlessness and an aptitude for generating more extensive responses. Additionally, the BLOOM, an exemplar of open science and accessibility, was conceived by the BigScience team at Hugging Face. Specifically designed to elaborate on textual prompts, BLOOM capitalizes on industrial-grade computational capacities to produce coherent text across 46 languages and 13 programming languages, rivaling the fidelity of human-generated content.

These expansive LLMs exemplify the forefront of advancements in their uncanny capacity to emulate human text generation. Their implications are manifold, particularly within domains such as natural language processing (NLP) and machine learning (ML). Consequently, they remain at the epicenter of fervent academic inquiry and technological progression.[27,47]

3. PROPOSED FRAMEWORK

In this work, we introduce an innovative framework for mineral classification augmented by Swin Transformer V2 models. This framework seamlessly integrates XAI techniques with LLMs with the overarching aim of enhancing the interpretability and understandability of the generated models. A comprehensive illustration of the structural composition of our mineral classification framework is provided in Figure 1. Moreover, to offer a tangible glimpse into its real-world implementation, the user interface (UI) of our proposed framework is depicted in Figure 2.

The ensuing sections meticulously detail each phase of our methodology-ranging from data preparation and model training to the nuanced intricacies of integrating XAI and LLMs into our framework.

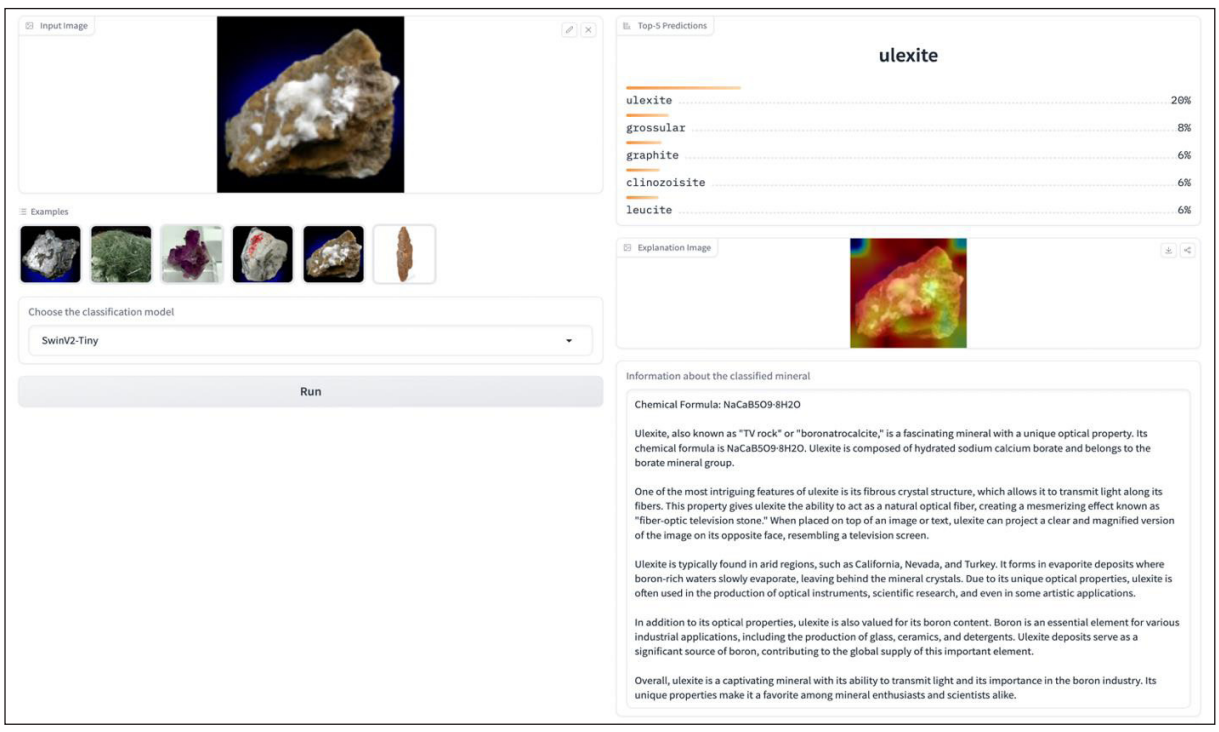


Figure 2. The mineral classification framework user interface (UI) deployed on the Huggingface platform with Gradio UI. The framework requires end-users to upload a mineral image and choose a classification model (the default model is set as SwinV2-Tiny) on the left panel. On the right panel, the top-5 predictions from models, explanation map of XAI methods on the model's prediction, and information retrieval about the top-1 classified mineral from GPT-4 are delivered.

https://doi.org/10.52111/qnjs.2023.17510

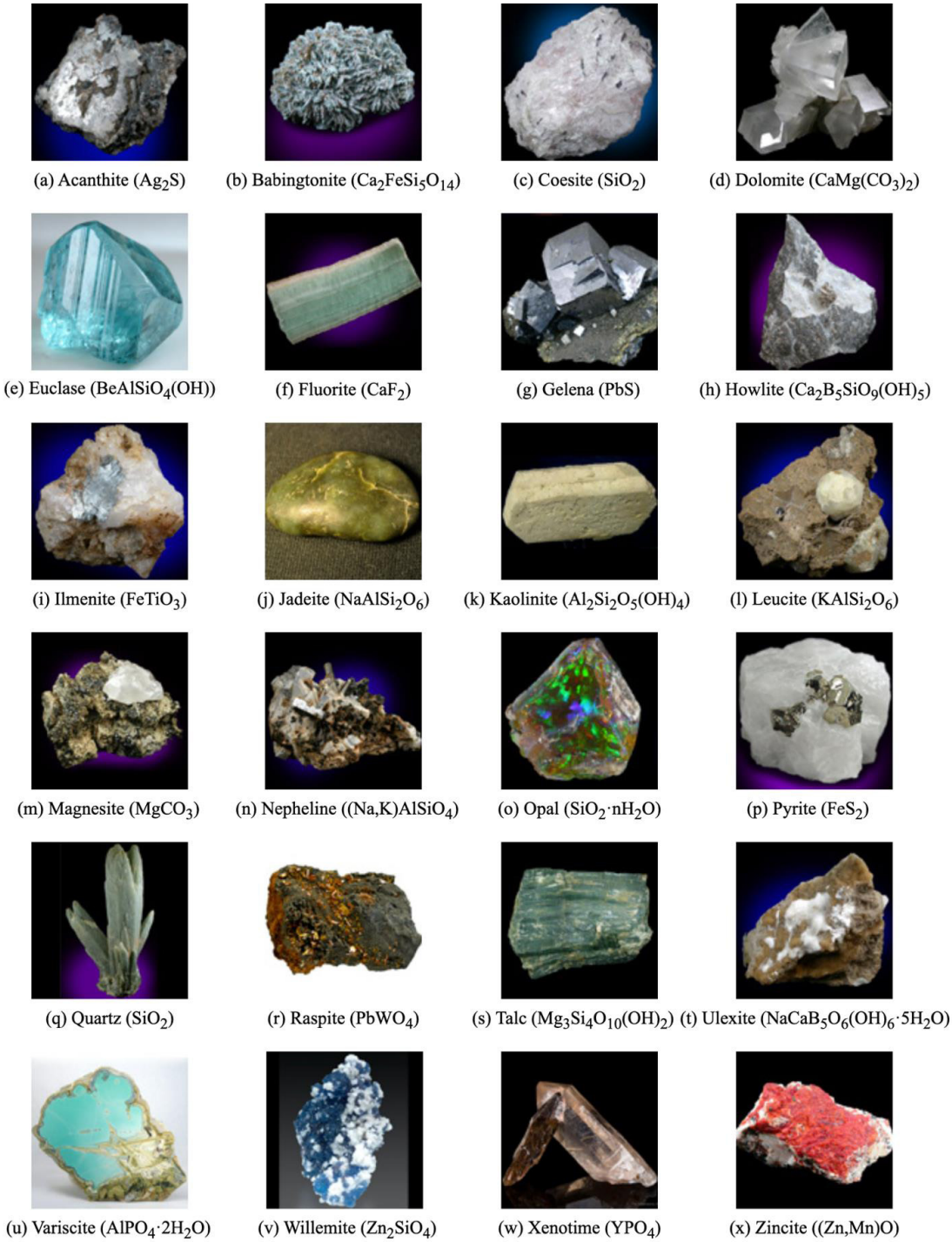


Figure 3. Samples of mineral specimens in the mineral dataset. Each mineral is shown in their name and formula.

3.1. Data preparation

Initiating with data acquisition, we embarked on a web-crawling exercise, amassing a rich dataset of mineral images, each meticulously annotated with their respective labels. The dataset contains around 4,000 images of 282 different minerals, each with labels. The dimensions of these images stand at 110 × 110 pixels. The labeling schema is comprehensive, encapsulating various attributes such as the mineral name, associated crystal system, chemical groupings, rock typologies, and fracture characteristics. For the purpose of model training and evaluation, the dataset was stratified into training and test sets, adhering to an 80% to 20% split ratio.

Given the inherent challenges posed by a limited number of images per mineral specimen (averaging about 14 images for each mineral type) and the relatively diminutive image dimensions, we employed a series of data augmentation strategies. Techniques such as Random Resized Crop and Random Horizontal Flip were judiciously applied to the training dataset to diversify and enhance its content.

3.2. Model training

Within the mineral classification framework, we incorporated three variants of the Swin Transformer V2 model, differentiated by their size: the Tiny-sized model (SwinV2-T), the Small-sized model (SwinV2-S), and the Base-sized model (SwinV2-B). Each of these models has undergone preliminary training on the ImageNet-1k dataset at a resolution of 256 × 256 pixels.[48] Recognizing the intricacies of a multiclass classification task, we elected the cross-entropy (CE) as our loss function, with the top-1 accuracy metric serving as the cornerstone of our evaluation process.

The training set, derived from our curated dataset, was harnessed to fine-tune these models. An advanced image preprocessing tool, the Vision Transformer (ViT), was deployed to ensure uniform normalization of images, thus harmonizing their resolution to align with the models' specifications. All associated hyperparameters pertinent to the fine-tuning process are systematically delineated in Table 1.

Table 1. The defined hyperparameters for finetuning the Swin Transformer V2 models.

Hyperparameter	Value
learning_rate	5e-5
warmup_ratio	0.1
gradient_accumulation_steps	4
batch_size	32

Subsequent to the fine-tuning phase, a rigorous evaluation was conducted to assess the performance of each model variant, employing the test set as the benchmark.

3.3. XAI integration

In this section, we leverage XAI to enhance the interpretability and transparency of Swin Transformer V2 models. We utilize GradCAM as the XAI method.[41] Given an input image, the forward pass computes activations at the chosen layer. The gradients of the class score concerning this layer's activations are then computed. These gradients are globally average-pooled to produce weights. Finally, a weighted combination of forward activation maps produces the GradCAM heatmap.

$$L_{GradCAM} = ReLU(\sum_{k} \alpha_k^c A^k)$$

where:

- $L_{GradCAM}$ is the explanation map for class c.
- α_k^c are the global-average-pooled gradients.
- A^k represents the forward activation maps for the chosen layer.
- $ReLU$ ensures that only positive influences on the class prediction are visualized.

3.4. Information retrieval with GPT-4

Given the multitude of mineral specimens that can be identified and categorized by our models, we recognized the imperative to supplement the raw classification with pertinent information. To this end, we employ the capabilities of GPT-4. This strategic integration is underpinned by the objective of furnishing end-users-who may lack prior familiarity with the specific mineral depicted in the image-with comprehensive and contextually relevant insights.

Upon obtaining the results from our primary classification model, we extract the top-most prediction, which is then utilized as an input for GPT-4. This methodology enables the provision of comprehensive and contextual data to the end-user. Notably, we have configured GPT-4 to emulate the expertise of a mineralogist, thereby ensuring that the generated information is not only informative but is also presented in a manner that is both engaging and cogent. It is worth emphasizing that vague or generic

https://doi.org/10.52111/qnjs.2023.17510

explanations are deliberately avoided, thereby enhancing the utility and reliability of the provided details.

To further bolster the authenticity and veracity of the information retrieved, we have imparted explicit instructions to GPT-4, directing it to rely solely on information from reputable sources. Among the preferred repositories are Wikipedia, an encyclopedia recognized for its vast and up-to-date content; The Mineral and Gemstone Kingdom, known for its exhaustive listings and detailed mineralogical insights; and the Mineral Resources Database, a repository hailed for its accuracy and comprehensive coverage. By anchoring our information retrieval process in such esteemed sources, we aspire to ensure that the knowledge disseminated to the users is both trustworthy and of the highest academic caliber.

4. RESULTS

In this section, we systematically present the empirical results and observations gleaned from the evaluations of the Swin Transformer V2 models. Initially, we will provide a quantitative assessment of the models based on the test set, followed by an exploration of the visual explanations in the form of saliency maps.

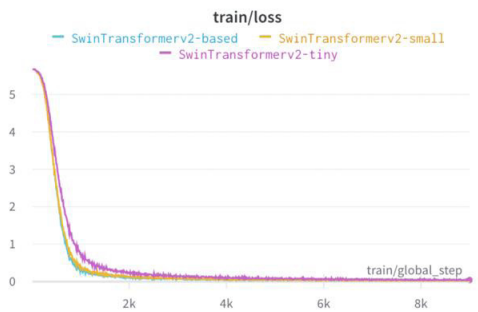
4.1. Quantitative assessment of model performances

We subject three distinct models - SwinV2-T, SwinV2-S, and SwinV2-B - to rigorous evaluation, both on training and test sets. As depicted in Figure 4, all three models demonstrate comparable CE loss on the training set. Notably, SwinV2-B emerges as the earliest to converge, trailed by SwinV2-S and SwinV2-T. Furthermore, SwinV2-B boasts the lowest CE loss among the trio.

However, a contrasting pattern emerges upon examining their performance on the test set, as shown in Figure 5. SwinV2-S achieves the lowest CE loss. Nevertheless, all three models showcase an analogous behavior; their CE losses manifest a steady uptick after the initial 1,000 training steps. This tendency suggests a pronounced overfitting to the training data and limited generalization to unseen datasets. This observation is further corroborated by accuracy metrics on the test set, with the most compact model, SwinV2-T, outperforming its counterparts.

In contemporary AI research, the efficiency of models, especially concerning GPU power consumption measured in Watts (W), has emerged as a crucial criterion. Lower power usage signifies a reduced carbon footprint, advancing the cause of sustainable and eco-friendly AI modeling. As one would anticipate, SwinV2-T, with its parsimonious parameterization, consumes the least power, trailed by SwinV2-S and then SwinV2-B, as evident from Figure 6.

Given the above empirical observations, factoring in both performance and efficiency, we advocate SwinV2-T as the primary model recommendation within our framework. However, we offer users the flexibility to leverage other models as per their requirements.



(a) The loss of three models on the training set

(b) The average loss of three models on the training set

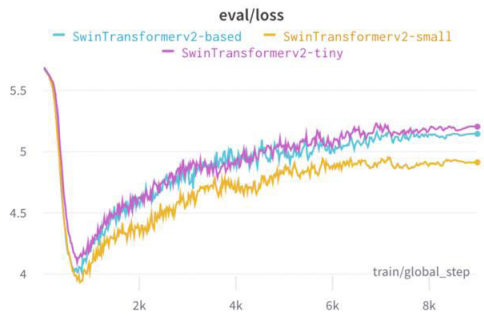
Figure 4. The (a) loss and (b) average loss of three classification models, namely SwinV2-T (pink), SwinV2-S (yellow), and SwinV2-B (blue), on the training set during the training phase.

4.2. In-depth qualitative analysis of classification explanations

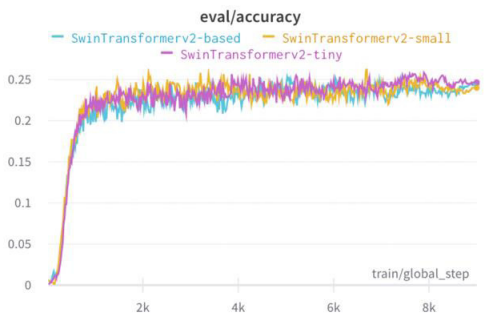
This section provides a meticulous qualitative dissection of the explanations underlying the classification decisions made by our selected model.

Figure 7 demarcates two distinct classification cases associated with the SwinV2-T model: an instance of accurate classification and a contrasting case of misclassification.

In scenarios where the classification proves accurate, the model's top-1 prediction perfectly resonates with the ground truth, illustrated by the case of the mineral Boleite. A closer examination reveals that the model, in its discernment, emphasizes specific features of the mineral. Specifically, it pays particular attention to the frontal facade of the mineral, which seems to be a key determinant in its accurate classification.



(a) The loss of three models on the test set



(b) The accuracy of three models on the test set

Figure 5. The (a) loss and (b) accuracy of three models, namely SwinV2-T (pink), SwinV2-S (yellow), and SwinV2-B (blue) on the test set during the training phase.

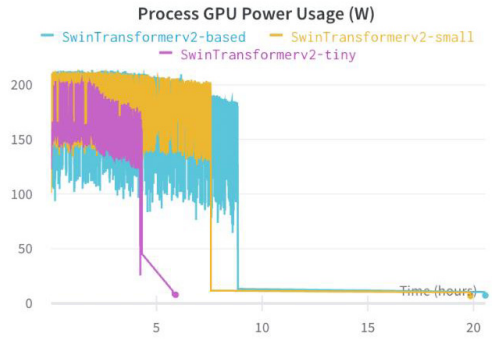


Figure 6. The process GPU power usage (in Watt) of three models, namely SwinV2-T (pink), SwinV2-S (yellow), and SwinV2-B (blue) during the training phase.

However, in instances where the classification goes awry, the model's decision-making intricacies become all the more fascinating. Taking the case where the model erroneously identifies an image as Monazite, with the second-highest probability assigned to Xenotime, the GradCAM explanation offers invaluable insights. The model seems to be unduly influenced by the image's background and specific sections of the Xenotime that are color-matched with training samples of Monazite. Such idiosyncrasies in model behavior, illuminated by XAI techniques, accentuate the need for interpretability. It ensures that even when the model falters, its decision-making rationale remains transparent, allowing users to comprehend and, potentially, correct or circumvent such pitfalls.

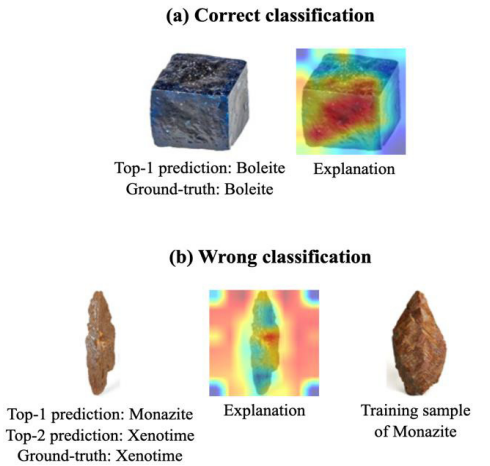


Figure 7. Examples of XAI-driven explanations for (a) an accurate classification and (b) an instance of misclassification as determined by SwinV2-T. Notably, the latter also integrates a training exemplar of the misclassified mineral, offering users an added layer of context.

https://doi.org/10.52111/qnjs.2023.17510

4.3. Comprehensive insights through information retrieval

Table 2. Information retrieval with GPT-4 on classified minerals. As being instructed to deliver only facts from credible sources, the GPT-4 can deliver human-like information to end-users.

Mineral: Monazite **GPT-4:** Monazite is a rare-earth phosphate mineral with the chemical formula (Ce,La,Nd,Th)PO_4. It is typically reddish-brown in color and has a resinous to vitreous luster. Monazite is an important source of rare-earth elements, which are used in various high-tech applications including electronics, magnets, and catalysts. It is also known for its radioactive properties, containing thorium and uranium. Due to its scarcity and economic value, monazite is often mined for its rare-earth content.
Mineral: Acanthite **GPT-4:** Acanthite is a silver sulfide mineral with the chemical formula Ag_2S. It is an important ore of silver and is commonly found in hydrothermal veins and low-temperature deposits. Acanthite is known for its metallic luster and dark gray to black color. It has a high density and is relatively soft. Acanthite is often associated with other silver minerals such as native silver and argentite. It is widely used in the production of silver metal and jewelry.

Building upon the top-1 classification of our mineral classification framework, GPT-4 acts as a sophisticated intermediary, extracting and presenting informative details about the identified minerals, such as Monazite and Acanthite, as shown in Table 2. Leveraging its vast training data, which encapsulates extensive knowledge on diverse mineral specimens, GPT-4 ensures that the information procured is not just accurate but is also curated to cater to users with varied levels of prior knowledge.

Furthermore, by incorporating safety protocols that ensure information retrieval solely from reputable sources, such as Wikipedia, The Mineral and Gemstone Kingdom, and the Mineral Resources Database, we guarantee the veracity and reliability of the procured data. Thus, users not only receive a rich tapestry of mineralogical information but also the assurance of its credibility. In essence, the synergy between our classification framework and GPT-4 creates an enriched user experience, fostering a more profound understanding and appreciation of the minerals.

5. CONCLUSION AND FUTURE WORK

Throughout this work, we have presented an AI-driven mineral classification framework characterized by its high interpretability and informative capabilities. This framework, bolstered by advanced XAI techniques and LLM, is strategically designed to cater to a wide audience, including those with limited or no prior expertise in mineralogy or AI. The incorporation of XAI proves invaluable, particularly in instances of incorrect model decisions, facilitating a more transparent and comprehensible insight into the model's reasoning. Such transparency is crucial in bolstering user trust and understanding, enabling them to more confidently engage with the system. Our future works revolve around broadening the scope of our dataset by integrating data from diverse and robust sources. This not only promises to enhance the model's precision but also its efficiency. Additionally, we aim to delve deeper into the human-centric aspect of our system. Specifically, we intend to orchestrate comprehensive human evaluations that will scrutinize both the plausibility and the faithfulness of explanations and information generated by XAI techniques and LLMs. Such evaluations will serve as a litmus test, assessing the real-world applicability and impact of our framework on its intended users.

REFERENCES

1. P. Patnaik. *Handbook of inorganic chemicals*, McGraw-Hill, New York, 2003.
2. S. T. Ishikawa, V. C. Gulick. An automated mineral classifier using Raman spectra, *Computers & Geosciences*, **2013**, *54*, 259-268.

3. A. S. Povarennykh. *Crystal chemical classification of minerals*, Springer, New York, 2014.

4. C. Owen, D. Pirie, G. Draper. *Earth lab: exploring the earth sciences,* Cengage Learning, Boston, USA, 2010.

5. C. Klein, B. Dutrow. *Manual of mineral science,* John Wiley & Sons, New Jersey, USA, 2007.

6. R. L. Shriner, C. K. Hermann, T. C. Morrill, D. Y. Curtin, R. C. Fuson. *The systematic identification of organic compounds,* John Wiley & Sons, New Jersey, USA, 2003.

7. X. Zeng, Y. Xiao, X. Ji, G. Wang. Mineral identification based on deep learning that combines image and Mohs hardness, *Minerals*, **2021**, *11*(5), 506.

8. M. E. Broz, R. F. Cook, D. L. Whitney. Microhardness, toughness, and modulus of Mohs scale minerals, *American Mineralogist*, **2006**, *91*, 135-142.

9. S. S. Lam, D. Fortin, B. Davis, T. Beveridge. Mineralization of bacterial surfaces, *Chemical Geology*, **1996**, *132*, 171-181.

10. I. V. Veksler, A. M. Dorfman, P. Dulski, V. S. Kamenetsky, L. V. Danyushesky, T. Jeffries, D. B. Dingwell. Partitioning of elements between silicate melt and immiscible fluoride, chloride, carbonate, phosphate and sulfate melts, with implications to the origin of natrocarbonatite, *Geochimica et Cosmochimica Acta*, **2012**, *79*(2), 20-40.

11. V. Apte. *Flammability testing of materials used in construction, transport, and mining*, Woodhead Publishing, Sawston, UK, 2021.

12. A. Ali, Y. W. Chiang, R. M. Santos. X-ray diffraction techniques for mineral characterization: a review for engineers of the fundamentals, applications, and research directions, *Minerals*, **2022**, *12*(2), 205.

13. G. W. Brindley. Identification of clay minerals by X-ray diffraction analysis, *Clays and Clay Minerals*, **1952**, *1*, 119-129.

14. J. Srodon, V. A. Drits, D. K. McCarty, J. C. Hsieh, D. D. Eberl. Quantitative X-ray diffraction analysis of clay-bearing rocks from random preparations, *Clays and Clay Minerals*, **2001**, *49*(6), 514-528.

15. S. Aligholi, R. Khajavi, M. Razmara. Automated mineral identification algorithm using optical properties of crystals, *Computers & Geosciences*, **2015**, *85*, 175–183.

16. C. D. Gribble. *Optical mineralogy: principles and practice,* Springer Science & Business Media, New York, 2012.

17. C. D. Gribble. *A practical introduction to optical mineralogy,* Springer Science & Business Media, New York, 2012.

18. A. G. Flores, S. Ilyas, G. W. Heyes, H. Kim. A critical review of artificial intelligence in mineral concentration, *Minerals Engineering*, **2022**, *189*, 107884.

19. H. P. Borges, M. S. D. Aguiar. *Mineral classification using machine learning and images of microscopic rock thin section,* Advances in Soft Computing: 18th Mexican International Conference on Artificial Intelligence, Xalapa, Mexico, 2019.

20. T. Long, Z. Zhou, G. Hancke, Y. Bai, Q. Gao. A review of artificial intelligence technologies in mineral identification: classification and visualization, *Journal of Sensor and Actuator Networks*, **2022**, *11*(3), 50.

21. H. T. T. Nguyen, H. Q. Cao, K. V. T. Nguyen, N. D. K. Pham. *Evaluation of explainable artificial intelligence: shap, lime, and cam,* The FPT AI Conference, Hoa Lac, Ha Noi, 2021.

22. E. Tjoa, C. Guan. A survey on explainable artificial intelligence (XAI): toward medical XAI, *IEEE Transactions on Neural Networks and Learning Systems*, **2020**, *32*, 4793-4813.

23. M. T. Ribeiro, S. Singh, C. Guestrin. *"Why should I trust you?" Explaining the predictions of any classifier,* The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 2016.

24. A. Chattopadhay, A. Sarkar, P. Howlader, V. N. Balasubramanian. *Grad-cam++: generalized gradient-based visual explanations for deep convolutional networks,* The IEEE Winter Conference on Applications of Computer Vision, WACV 2018, Lake Tahoe, NV, USA, 2018.

https://doi.org/10.52111/qnjs.2023.17510

25. D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G. Z. Yang. XAI-explainable artificial intelligence, *Science Robotics*, **2019**, *4*(37), 7120.

26. F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, J. Zhu. *Explainable AI: a brief survey on history, research areas, approaches and challenges*, Natural Language Processing and Chinese Computing: 8th CCF International Conference, NLPCC 2019, Dunhuang, China, 2019.

27. B. Min, H. Ross, E. Sulem, A. P. B. Veyseh, T. H. Nguyen, O. Sainz, E. Agirre, I. Heinz, D. Roth. Recent advances in natural language processing via large pre-trained language models: a survey, *ACM Computing Surveys*, **2021**, *56*(2), 30.

28. Y. Lecun, Y. Bengio. *Convolutional networks for images, speech, and time-series,* MIT Press, Cambridge, 1995.

29. X. S. Teng, Z. Y. Zhang. Artificial intelligence identification of ore minerals under microscope based on deep learning algorithm, *Acta Petrologica Sinica*, **2018**, *34*, 3244-3252.

30. P. Theerthagiri, A. U. Ruby, B. Chaithanya, R. R. Patil, S. Jain. D-Resnet: deep residual neural network for exploration, identification, and classification of beach sand minerals, *Multimedia Tools and Applications*, **2023**, *83*(6433), 1-25.

31. G. Yanjun, Z. Zhe, L. Hexun, L. Xiaohui, C. Danqiu, Z. Jiaqi, W. Junqi. The mineral intelligence identification method based on deep learning algorithms, *Earth Science Frontiers*, **2020**, *27*(5), 39-47.

32. W. Ren, M. Zhang, S. Zhang, J. Qiao, J. Huang. *Identifying rock thin section based on convolutional neural networks*, The 9th International Workshop on Computer Science and Engineering, WCSE 2019*,* Hong Kong, China, 2019.

33. T. Miller, D. C. Lech, A. Kisiel, P. Kozlovska, A. Krzemińska, S. Mosiundz, A. Kutsevych, K. Lewita, M. Jawor. Applied statistics and machine learning in earth sciences: choosing the right approach for modern scientific research, *Collection of Scientific Papers*, **2023**, 244–249.

34. Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, B. Guo. *Swin transformer: hierarchical vision transformer using shifted windows*, The IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 2021.

35. D. Luo, W. Zeng, J. Chen, W. Tang. Deep learning for automatic image segmentation in stomatology and its clinical application, *Frontiers in Medical Technology*, **2021**, *3*, 767836.

36. Z. Liu, H. Hu, Y. Lin, A. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, B. Guo. *Swin transformer v2: scaling up capacity and resolution,* The IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, 2022.

37. J. Kang, S. Tariq, H. Oh, S. S. Woo. A survey of deep learning-based object detection methods and datasets for overhead imagery, *IEEE Access*, **2022**, *10*, 20118-20134.

38. A. Binder, G. Montavon, S. Lapuschkin, K. R. Müller, W. Samek. *Layer-wise relevance propagation for neural networks with local renormalization layers*, Artificial Neural Networks and Machine Learning–ICANN 2016: 25th International Conference on Artificial Neural Networks, Barcelona, Spain, 2016.

39. V. Petsiuk*,* R. Jain, V. Manjunatha, V. I. Morariu, A. Mehra, V. Ordonez, K. Saenko. *Black-box explanation of object detectors via saliency maps,* The IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2021, Nashville, TN, USA, 2021.

40. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba. *Learning deep features for discriminative localization,* 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, 2016.

41. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra. *Grad-cam: visual explanations from deep networks via gradient-based localization,* The IEEE International Conference on Computer Vision, Venice, Italy, 2017.

42. P. X. Nguyen, H. Q. Cao, K. V. Nguyen, H. Nguyen, T. Yairi. SeCAM: tightly accelerate the image explanation via region-based segmentation, *IEICE Transactions on Information and Systems*, **2022**, *105*, 1401-1417.

43. H. Wang, Z. Wang, M. Du, F. Yang, Z. Zang, S. Ding, P. Mardziel, X. Hu. *Score-CAM: score-weighted visual explanations for convolutional neural networks,* The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020.

44. T. T. H. Nguyen, V. B. Truong, V. T. K. Nguyen, Q. H. Cao, Q. K. Nguyen. *Towards trust of explainable AI in thynoid nodule diagnosis*, The 7th International Workshop on Health Intelligence, W3PHAI 2023, Washington, DC, USA, 2023.

45. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. H. Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, D. Amodei. Language models are few-shot learners, *Advances in Neural Information Processing Systems*, **2020**, *33*, 1877-1901.

46. OpenAI. GPT-4 Technical Report, California, 2023.

47. H. Huang, O. Zheng, D. Wang, J. Yin, Z. Wang, S. Ding, H. Yin, C. Xu, R. Yang, Q. Zheng, B. Shi. ChatGPT for shaping the future of dentistry: the potential of multi-modal large language model, *International Journal of Oral Science*, **2023**, *15*(29).

48. J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. F. Fei. *Imagenet: a large-scale hierarchical image database,* The IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, Miami, FL, USA, 2009.

https://doi.org/10.52111/qnjs.2023.17510