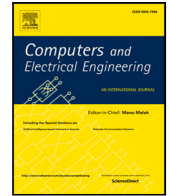


Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

# Computers and Electrical Engineering

journal homepage: [www.elsevier.com/locate/compeleceng](http://www.elsevier.com/locate/compeleceng)

## Artificial intelligence applied to the stone manufacturing industry: A systematic literature review

Alexandre Silva , Carolina Antunes, Rolando Miragaia , Rogério Luís Costa ,  
Fernando Silva , José Ribeiro \*

Computer Science and Communications Research Centre, School of Technology and Management, Polytechnic of Leiria, Leiria, 2411-901, Portugal

### ARTICLE INFO

#### Keywords:

Artificial intelligence  
Machine learning  
Stone industry  
Stone manufacturing  
Natural stone  
Computer vision

### ABSTRACT

Natural stone has long been used in construction, as its properties provide functional and visual value, and the natural stone market currently holds significant importance in the global economy. It is important to consider integrating new technologies in the production chain to aid the industry in moving forward, increasing profit margins and reducing wasted material. This article reviews recent trends in using Artificial Intelligence and Machine Learning techniques in the industry between 2017 and 2024, following a methodology for Systematic Literature Reviews in computer science. It was found that extensive research has been conducted on the subject of tile classification, with solid solutions proposed, achieving results that can be considered robust enough for industrial application. Other subjects comprise tasks regarding stone cutting and defect detection, as well as variable prediction, and quarry activity monitoring. Some authors propose solutions to integrate new technologies into the complete production chain. While more research needs to be done on specific subjects, this review provides a solid first step to future research.

### 1. Introduction

The natural stone industry supplies stone products to the construction sector, which are often used for architectural purposes. The use of stone in construction comes with many advantages: it is a natural material that presents a low environmental impact (when compared to synthetic alternatives), it is strong and durable, meaning that it can withstand heavy loads and last for centuries (as evidenced by ancient constructions), it is fire resistant and is an excellent thermal insulation material.

The natural stone industry encompasses several stages, from extraction to processing and finishing. The extraction process involves removing stone blocks from quarries, which are then transported to processing plants for cutting, shaping, and finishing. The final products are distributed to various markets, including construction companies, architects, and interior designers.

The natural stone production/transformation industry significantly values the global economy. According to Fortune Business Insights [1], the total global market value was at 50.65 billion USD in 2022, being projected to grow to 73.47 billion USD by 2030, and exhibiting a Compound Annual Growth Rate (CAGR) of 4.5%. According to [2], published by the Portuguese Association of the Industry of Mineral Resources, ASSIMAGRA, the industry saw a global value of exports of 487.61 million EUR, marking a product valorisation of 5.24% year-over-year, when compared to 2022. The export market has grown 59% since 2017, when it was valued at 345.84M EUR.

\* Corresponding author.

E-mail addresses: [2230329@my.ipleiria.pt](mailto:2230329@my.ipleiria.pt) (A. Silva), [2230333@my.ipleiria.pt](mailto:2230333@my.ipleiria.pt) (C. Antunes), [rolando.miragaia@ipleiria.pt](mailto:rolando.miragaia@ipleiria.pt) (R. Miragaia), [rogerio.l.costa@ipleiria.pt](mailto:rogerio.l.costa@ipleiria.pt) (R.L. Costa), [fernando.silva@ipleiria.pt](mailto:fernando.silva@ipleiria.pt) (F. Silva), [jose.ribeiro@ipleiria.pt](mailto:jose.ribeiro@ipleiria.pt) (J. Ribeiro).

<https://doi.org/10.1016/j.compeleceng.2025.110702>

Received 10 March 2025; Received in revised form 25 July 2025; Accepted 3 September 2025

Available online 16 September 2025

0045-7906/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

To further valorise natural stone, it is important to consider the use of new technologies. In an industry where so many processes are done manually, employing new technologies could optimise processes, leading to better management of human resources, less material wastage, higher production rate, and overall final product quality increase. By using AI, it is possible to provide better quality control – for example, by better identifying and correcting defects in stone slabs –, improve customer satisfaction, automate processes and predict machine maintenance, as well as understand market shifts at their first signs, all while lowering labour costs, optimising resource allocation, reducing waste and minimising downtime. All of this contributes to a stronger, better-optimised and better-valued industry.

Despite the growing interest in AI, we observed that a comprehensive review of research done in the past few years for the natural stone industry was lacking. We aim to address this gap by providing a systematic literature review (SLR) that aggregates research articles published since 2017.

In this review, which follows the research methodology proposed in [3], various studies published between 2017 and 2024 will serve as the basis for discussion on how AI and ML can be implemented in the natural stone industry. We summarise and analyse articles containing research in the natural stone industry field, identifying benefits, challenges, limitations, trends and opportunities in applying AI to the industry. Overall, 20 studies are presented and discussed.

In summary, the contributions of this review are as follows:

- Summarise research made for the natural stone industry since 2017;
- Identify areas where research has been done, present solutions and outcomes, and discuss how future research can improve on the existing work;
- Analyse and discuss trends verified in research articles over the years;
- Provide a document with detailed information on the research so far, thus presenting a starting point to future researchers.

In Section 2 we outline the methodology followed to obtain the studies that are summarised in Section 3 and discussed in Section 4. In Section 5, we present a conclusion.

## 2. Methodology

The present systematic literary review was performed according to the methodology proposed in [3]. This methodology proposes the following work:

- Planning:
  - Define Population, Intervention, Comparison, Outcome and Context (PICOC) and Synonyms;
  - Formulate Research Questions;
  - Select digital libraries sources;
  - Define inclusion and exclusion criteria;
  - Define quality assessment checklist;
  - Design the data extraction form.
- Conducting:
  - Build digital library search strings;
  - Gather studies;
  - Study selection and refinement;
  - Data extraction;
  - Analysis and report.

The following subsections analyse each step of the review planning process: definition of PICOC and synonyms, formulation of research questions (RQs), selection of digital sources, definition of inclusion/exclusion criteria, and the quality assessment (QA) checklist; design of the data extraction form, build of search strings, and study gathering, selection and refinement.

By following this methodology, we ensure that we provide a comprehensive look into how AI is used in the stone manufacturing industry, while ensuring reproducibility. This methodology has been extensively used in other search topics in literature [4–9].

### 2.1. Defining PICOC and synonyms

PICOC is a framework used to define research questions and guide the literature search in systematic literature reviews. Using PICOC, researchers can better formulate clear and focused research questions, develop search strategies, establish inclusion/exclusion criteria for studies, and synthesise evidence relevant to the research question. Table 1 contains the terms selected in the context of using AI/ML in the natural stone industry.

**Table 1**  
PICOC analysis

	PICOC	Synonyms
Population	Natural Stone Industry	Researchers in the natural stone industry
Intervention	Learning Algorithms	AI, ML
Comparison	Traditional Methods	Manual Processing
Outcome	Productivity Enhancement	Process Automation
Context	Natural Stone Manufacturing	Extraction and Processing of Natural Stone

## 2.2. Formulating research questions

The primary aim of this review is to provide insight into how AI is currently used in the natural stone industry. As such, it is important to understand which problems have been identified in the industry and how they were solved or mitigated by utilising AI. Furthermore, from the solutions already presented, would there be any way of pushing them towards mass adoption, mainly by integrating them with other technologies? These ideas led to the proposal of a list of research questions:

- **RQ1:** How has AI been applied to various aspects of the natural stone industry, and what are the key benefits and challenges associated with these applications?
- **RQ2:** What specific AI techniques and algorithms have been most effectively employed in the natural stone industry, and why?
- **RQ3:** How could AI be integrated with other technologies to enhance value and efficiency in the natural stone industry?
- **RQ4:** What are AI adoption's potential challenges and limitations in the natural stone industry, and how can these be overcome?

## 2.3. Selecting sources and defining inclusion and exclusion criteria

It is of utmost importance to select the correct databases for research. We selected four databases holding scientific papers in the computer science and engineering fields: Scopus, Web of Science, ACM, and IEEEExplore.

Scopus and Web of Science are multidisciplinary databases covering a wide range of research fields. ACM is focused on computer science and related fields, while IEEEExplore is better known for engineering articles, mainly on the electrical and computer engineering fields. All of these databases are used for academic research, particularly in computer engineering. They provide access to peer-reviewed articles, conference papers, and other scholarly publications. They also offer features for citation analysis, author profiling, and advanced searching.

While extensive research on the topic has been conducted in the last decades, it was decided that reviewed studies should be recent. Thus, the publication year of the literature reviewed for this systematic literature review was defined as 2017 and afterwards.

Moreover, while conducting the first searches, it was found that many studies were related to other research areas, such as medical or civil engineering. To better filter the results, it was agreed to accept only articles in computer science and engineering.

It was initially understood that articles should be written in English. However, after further consideration, articles in Portuguese were also accepted since the authors of this review are native Portuguese speakers. Other languages were not considered, as important details might be lost in translation.

## 2.4. Defining the quality assessment checklist

A QA checklist is essential for robust SLRs. It minimises bias by systematically evaluating methodological flaws in included studies, ensuring conclusions are based on reliable evidence. We selected the following criteria:

- **QA1:** Is there a clear definition/description/statement of the aims/ goals/purposes/motivations/objectives/questions of the research?
- **QA2:** Do the researchers discuss any problems (limitations, threats) with the validity of their results (reliability)?
- **QA3:** Are the results presented and supported by the data?

A five-point scale was used for each question, with options for “agree”, “somewhat agree”, “neither agree nor disagree”, “somewhat disagree”, and “disagree”. A weighted scoring system was employed, assigning values between 5 and 1, in steps of 1, from “agree” to “disagree”.

Each selected study was subjected to the questions in the QA checklist.

## 2.5. Designing the data extraction form and building digital library search strings

The data extraction form allows for the systematic synthesis of study information. Each study was subjected to data extraction for the following topics: Subject, Initial Problem, Proposed Solution, Testing Methodology, Outcome and Conclusion.

This information was then used to create the study summary tables and further analyse the proposed solutions in Sections 3.1 to 3.4.

By this point, the scope of the review is defined: the objective is to understand how AI has been employed in the natural stone industry.

Deconstructing this sentence, we find that “Artificial Intelligence” is a required phrase. However, other equivalent terms are often used, such as “Machine Learning” or “Deep Learning”. We also want to know if Large Language Models (LLMs) have been used in the industry.

Furthermore, we want to learn how AI is used in the stone industry. As such, “Stone” can be identified as a core keyword. Often, “Stone” is too broad a term. Authors may prefer identifying the specific type of stone their research is based on. In the natural stone industry, some of the most used types of stone are “Limestone”, “Marble” and “Granite”.

As the SLR is aimed at the industry, other important keywords can be identified: “Manufacturing”, “Production” and “Industry”.

Based on this information, a search string was formulated, combining them with boolean operators.

After an initial search, we found that authors often use phrases such as “Stepping Stone”, “Corner Stone”, or “Foundational Stone” to describe their research as important in its field. A consequence of simply using the term “Stone” was that many articles not related to the desired context were retrieved as they used those phrases in the abstract. We revised the search string to exclude articles containing these phrases in the abstract.

Other terms that did not relate to our goals were “Concrete” and “Cement”. Many articles have been written on the properties of concrete from a civil engineering perspective. While AI was used to accomplish the research topics defined by the authors, the cement production industry is not directly related to the natural stone industry.

The final search string is as such:

**(“Artificial Intelligence” OR “AI” OR “A.I.” OR “Machine Learning” OR “ML” OR “M.L.” OR “Deep Learning” OR “DL” OR “D.L.” OR “Large Language Model” OR “LLM”) AND (“Manufacturing” OR “Industry” OR “Production”) AND (“Stone” OR “Limestone” OR “Marble” OR “Granite”) AND NOT (“Corner Stone” OR “Stepping Stone” OR “Foundational Stone” OR “Concrete” OR “Cement”).**

## 2.6. Study gathering, selection and refinement

First, we searched Scopus. Using the previously identified search string and inclusion/exclusion criteria, 122 results were obtained. At this point, we manually analysed each article’s title and abstract. Articles which did not conform to the specified requirements were rejected. In the end, 30 articles were accepted for review.

With the first articles selected, it was time to move on to other databases. It was clear from the start that there would be duplicated entries. Furthermore, as with the search on the Scopus database, the vast majority of articles would end up being rejected. With a better understanding of what to look for in each article, it was possible to streamline the search process somewhat. For each new source, we performed a search using the previously defined search string and inclusion/exclusion criteria. The results list was exported to BibTeX format whenever possible, except for IEEEXplore, where it is only possible to export results in CSV format. Scripts written in Python were used to create a CSV file and a BibTeX file containing entries accepted for manual review.

The algorithm for the scripts that filter out results from each source can be seen in Algorithm 1.

---

### Algorithm 1 Algorithm used to filter results

---

- 1: Read input file
  - 2: Read comparison file (from a previous database)
  - 3: For each article:
    - 4: Check “DOI” against the list from the comparison file
    - 5: If “DOI” found in comparison file:
      - 6: Add article to “Duplicate Entries” dictionary
    - 7: Check title and abstract for forbidden keywords
    - 8: If mandatory AI-related keywords are not found:
      - 9: Add article to “Lacking AI Keywords” dictionary
    - 10: If mandatory production-related keywords are not found:
      - 11: Add article to “Lacking Production Keywords” dictionary
    - 12: If mandatory stone-related keywords are not found:
      - 13: Add article to “Lacking Stone Keywords” dictionary
  - 14: Write entry to output CSV file
  - 15: Add entry to output BibTeX database
  - 16: Write BibTeX database to output BibTeX file
  - 17: Write statistics to log file
- 

Any rejected entries are removed from the BibTeX file during the manual review, so only accepted entries are further analysed.

Following the first selection of articles, we use the *Parsifal* tool for further analysis. This web platform allows researchers to organise information related to an SLR through every step. From planning to selecting articles and their subsequent evaluation and review, *Parsifal* ensures that the methodology is strictly followed and respected.

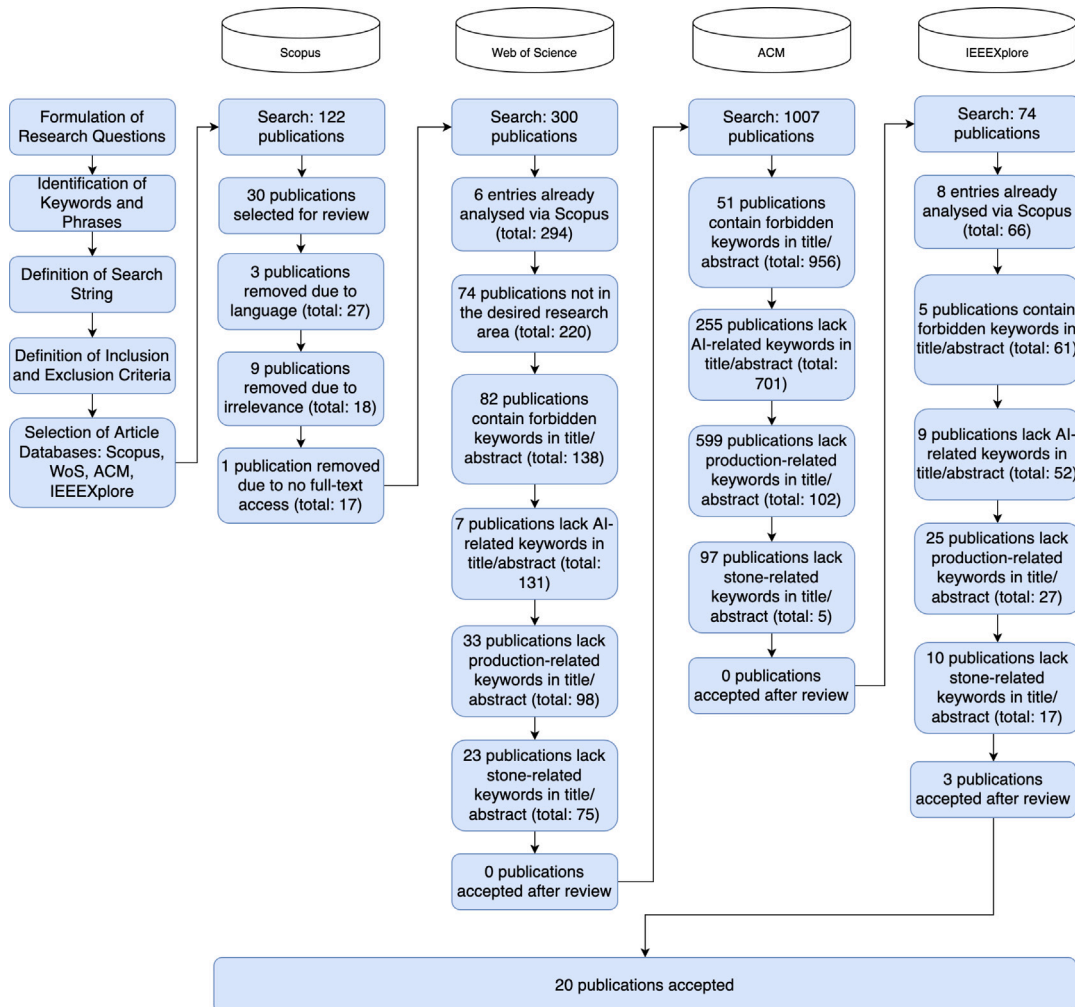


Fig. 1. Search results by following the described methodology.

Furthermore, the list from the log file is manually verified to ensure that no article was incorrectly removed from the review. While all entries are ultimately verified, the scripts allow for a deeper focus on the articles that are more likely to be accepted, saving time in the process.

The summarised process can be seen in Fig. 1.

### 3. Results

This section reviews and summarises each article according to the authors' proposed solutions and outcomes. We describe the subjects identified during the article selection and review process in Sections 3.1 to 3.4. Section 4 presents a discussion on the collected information and answers the RQs.

The articles were first assessed as to their quality, according to the QA checklist defined in Section 2.4. Every article was given a maximum score of 5 in QA1 and QA3 questions. During our first assessment of each article, we determined that all of them clearly defined the research aims, goals, purposes, motivations and objectives, usually in the introduction, as well as the research questions. The same applies to the results being supported by the data, as there was never evidence of data contradicting the results that were achieved by the authors of the studies. As for QA2, we found that some articles did not thoroughly discuss problems with the validity of their results. In those articles, there may not have been a clear discussion on how different variables may affect the results of the proposed solution, such as how the chosen dataset impacts accuracy in a real scenario, but they still highlighted enough limitations to deem this QA point as satisfactory. As such, some articles were given a score of 4 out of 5. The results can be seen in Table 2.

**Table 2**  
QA scores.

References	Score
[10], [11], [12], [13], [14], [15], [16], [17], [18], [19], [20]	Maximum Score in QA1, QA2 and QA3
[21], [22], [23], [24], [25], [26], [27], [28], [29]	Maximum Score in QA1 and QA3, 4 in QA2

### 3.1. Tile/slab classification

The tile/slab classification subject articles were summarised and can be seen in [Table 3](#).

In 2017, [21] reflected on the use of CNNs provided by the VLFEAT package for MATLAB, optimised for the Canadian Institute For Advanced Research (CIFAR) [30] and Modified National Institute of Standards and Technology (MNIST) [31] datasets, applied for granite tile classification. To do so, each image was divided into small patches of resolution  $32 \times 32$ px and majority voting led to the final image classification. A dataset, developed initially by Bianconi et al. used in [32], containing 1000 RGB images of  $1000 \times 1000$ px resolution subdivided into 25 classes, is used, and the models were trained using  $5 \times 2$  cross-validation. The results were highly accurate, reaching, in fact, 100% when using the CIFAR network. Variations of the MNIST network were also tested, with two of them achieving accuracy results above 98%. The authors also proposed custom solutions using three variations of the MNIST network simultaneously, combining the results for the tile classification. One such solution, the early fusion approach, achieved an accuracy of 96.50%. In this approach, the results of each MNIST network are combined into one single vector, from which each patch is classified and, finally, the granite slab as a whole. These networks were trained from scratch. Pre-trained networks led to poor classification results. The authors also evaluated the CIFAR network on the small  $32 \times 32$ px patches, obtaining a result of 87.26%.

In 2018, [22] proposed a solution for marble slab classification involving the use of Local Binary Patterns (LBP) [33,34] and image processing techniques, such as thresholding, Sobel edge detection, and Gabor filtering. Two variants of the solution were tested. In the first, images were processed and, using LBP, colour and texture information was extracted. No image processing was done in the second variant, and texture information was directly extracted using LBP. The extracted information was then fed to three AI algorithms: Support Vector Machine (SVM) [35,36], Random Forest (RF) [37,38] and Decision Tree (DT) [39]. The best overall approach, based on accuracy, Kappa, and execution time, was that in which no image processing was done and the DT algorithm was used.

Also in 2018, [10] proposed a CNN which takes as input small blocks of colour granite images. The original image is divided into patches of  $30 \times 30$ px, each classified by the model. Besides the input layer, the CNN contains 14 other layers: three convolutional layers, three pooling layers, four ReLU layers, two fully-connected layers, one softmax layer, and one classification layer. In the end, majority voting is performed, where the most predicted class is accepted as the class the image belongs to. The authors used the dataset originally developed by Bianconi et al. [32], which was also used by Ferreira et al. [21]. Results were very accurate, at 95.69% accuracy. A similar CNN, taking patches of  $32 \times 32$ px, was also tested, with an overall accuracy score of 97.44%. Compared to the CIFAR network provided in the MatConvNet library by VLFEAT, both approaches significantly exceed this network's performance, which scored 87.26% accuracy. The authors claim that the proposed solution makes the model resolution invariant.

In 2020, a study applied to limestone slab classification, [11], put forward a solution comprising feature extraction using the Inception-V3 model [40] and three different algorithms for decision-making: SVM [35,36], Naive Bayes (NB) [41] and Logistic Regression (LR). The dataset comprised images of limestone slabs from four distinct classes: CADOICO (Cadoico Azul Mónica Silva), SBM (Salgueira Branco do Mar), SBR (Salgueira Branco Real) e VMF (Vidraço Moleanos Farpedra). A comparison between the algorithms was performed, along with an analysis of which image formats were best for the classification tasks — PNG or JPEG. A dataset containing 2500 images of limestone slabs from four categories at a resolution of  $3500 \times 2500$ px, divided into three parts (two for training, one for testing) is used. The training process comprises 10-fold cross-validation. The results showed better accuracy for the SVM and LR algorithms, the former presenting a slight advantage, whilst paired with the PNG format. The differences were negligible, however, and the JPEG format was more beneficial overall. Despite having slightly worse results than SVM, the LR algorithm was deemed the best performing for the use case due to lower inference times.

In another study, [23], published in 2020, the authors proposed a CNN architecture that was able to provide promising results regarding the classification of marble slabs. The authors initially found that overfitting took a heavy toll on the models' performance without employing image processing techniques to the dataset. However, through experimentation, it was found that by using a blur filter, a 2D linear separable filter, and an erosion filter, specific characteristics could be smoothed out while others could be emphasised, thus leading to less overfitting and a significant improvement in terms of the model's accuracy. A dataset containing 2100 images across six different marble quality classes was used, with an initial resolution of  $2400 \times 2400$ px. The images were cropped to  $300 \times 300$ px and, after data augmentation, 6300 images were available, 5400 of which were used for training. The model achieved an accuracy of 92.2% on the test dataset, comprising 900 samples of six different marble classes.

In 2021, [24] proposed a solution that, instead of relying on conventional CNNs, uses a comparison between three images for classification. The Distance Metric Learning (DML) [42] technique relies on the distance between the anchor (i.e. an image) and its positive (i.e. an image of the same class), and simultaneously on the distance to a negative image (i.e. an image from a different class). The closer the anchor is to the positive and furthest from the negative, the more likely it is to be of the same class as the positive image. Through Metric Learning, a high-dimensional embedding vector is extracted. As a Deep Neural Network (DNN) is used to

**Table 3**  
Summary of tile/slab classification articles.

Ref.	Summary	Outcomes
[10]	<ul style="list-style-type: none"> <li>• Colour images of granite blocks are divided into small patches</li> <li>• Each patch is classified by the model</li> <li>• Model is trained on a custom CNN</li> <li>• Majority voting performed, image belongs to most predicted class</li> </ul>	<ul style="list-style-type: none"> <li>• Accuracy: 95.69% for 30 × 30px patches, 97.44% for 32 × 32px patches</li> <li>• CIFAR network achieves 87.26% accuracy, lower than the proposed solution</li> </ul>
[11]	<ul style="list-style-type: none"> <li>• Obtain image characteristics using the Inception-V3 network</li> <li>• Comparison between algorithms: SVM, NB, LR</li> </ul>	<ul style="list-style-type: none"> <li>• Best algorithm: SVM, followed closely by LR</li> <li>• PNG yields better results, JPEG overall more beneficial due to smaller file size</li> </ul>
[12]	<ul style="list-style-type: none"> <li>• Marble tile images are assigned a Quality Value (QV) from 0 to 1, with 0.1 increments</li> <li>• Comparison between 5 CNNs pre-trained on the ImageNet dataset</li> <li>• Changes to the final layers of each network</li> <li>• Best performing CNN is compared with fully trained network</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from MN model: MAPE of 31.20%</li> <li>• Better results from fully trained MN model: MAPE of 21.17%</li> </ul>
[13]	<ul style="list-style-type: none"> <li>• Comparison between 4 CNN architectures</li> <li>• Colour vs. grey-scale images</li> </ul>	<ul style="list-style-type: none"> <li>• 100% accuracy from ResNet-152 and Inception models with grey-scale images, but Inception requires less computational resources</li> <li>• 98.6% accuracy from ResNet-152 and VGGNet models with colour images; ResNet-152 takes less time to train, but VGGNet is less complex</li> <li>• Overall, models perform better with grey-scale images</li> </ul>
[21]	<ul style="list-style-type: none"> <li>• Comparison between MNIST, variations of MNIST and CIFAR architectures</li> <li>• Images are patched into blocks of smaller size</li> <li>• Two custom approaches to results fusion: early and late fusion</li> </ul>	<ul style="list-style-type: none"> <li>• CIFAR network excels, at 100% accuracy</li> <li>• MNIST architecture with nine layers shows 99.32% accuracy</li> <li>• Early fusion approach reaches 96.50% accuracy</li> <li>• Late fusion approach performs poorly, at 26.44%</li> <li>• Pre-trained networks fare poorly</li> </ul>
[22]	<ul style="list-style-type: none"> <li>• Use of image processing techniques: thresholding, sobel edge detection and gabor filtering</li> <li>• Two solutions compared</li> <li>• First solution: images are processed; LBP is used to extract colour and texture information</li> <li>• Second solution: no image processing; LBP is used to extract texture information</li> <li>• Comparison between algorithms: SVM, RF, DT</li> </ul>	<ul style="list-style-type: none"> <li>• Best approach: no image processing, use DT algorithm</li> </ul>
[23]	<ul style="list-style-type: none"> <li>• Use of custom CNN with data augmentation techniques: blur filter, 2D linear separable filter and erosion filter</li> <li>• Comparison with other algorithms</li> </ul>	<ul style="list-style-type: none"> <li>• Average accuracy: 0.922 with test dataset, 0.961 with 10-fold cross-validation</li> <li>• Better results than with other algorithms: best next avg. accuracy at 0.757 from MLP network</li> </ul>
[24]	<ul style="list-style-type: none"> <li>• Deep Distance Metric Learning (DDML)</li> <li>• Uses three images simultaneously: anchor, positive and negative</li> <li>• Objective: minimise the distance between anchor and positive, maximise the distance between anchor and negative</li> <li>• Comparison with other algorithms</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from DenseNet201, with TL: 80.77% accuracy</li> <li>• DDML shows slightly lower accuracy at 79.78%, but lower standard deviation</li> </ul>
[25]	<ul style="list-style-type: none"> <li>• Initial pre-processing of images</li> <li>• Comparison between 15 different CNNs pre-trained on the ImageNet dataset</li> <li>• Output layer is removed, models' weights frozen, new layers added and trained</li> <li>• Networks' performance tested by training layers in increments of 25%</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from DenseNet201: 83.24% accuracy</li> <li>• DenseNet169 showed better precision at 83.97%</li> </ul>
[26]	<ul style="list-style-type: none"> <li>• Use ELM, often used for texture classification</li> <li>• Comparison against ID3, C4.5, SVM and ANN</li> <li>• LBP and histogram algorithms used to extract features</li> <li>• Feature extraction algorithms compared</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from ELM, using both LBP and histogram algorithms for feature extraction: 97.5% accuracy</li> <li>• ANN model shows similar performance but with significant computational cost</li> </ul>

extract the embedding vector; this approach is called Deep Distance Metric Learning (DDML). Tile images go through preprocessing tasks, are resized to a resolution of  $224 \times 224$ px, and their values normalised. A model based on one of seven pre-trained CNNs is trained using triplets of images — anchor, positive, and negative. An embedder converts the output of the last convolutional layer to an embedding vector of 256 values, while the classifier outputs the predicted class of the input image. A dataset initially containing 986 images of marble tiles goes through preprocessing tasks and is trimmed. In the end, class A contains 66 images, class B contains 67 images, and class C contains 96 images. After training the models, it is observed that the best model results from using the DenseNet201 [43] network with TL, at 80.77% accuracy, with the following best result coming from the same network when using the DDML approach, at an accuracy of 79.78%. On average, DDML offers a lower standard deviation between models than TL, which means that while accuracy results were not much improved from the conventional CNNs, a notable increase in stability was observed.

In that same year, in [12], the authors tested five different CNN architectures to assign a quality value ranging from 0 to 1 in increments of 0.1 to images of marble tiles. Five CNNs pre-trained on the ImageNet dataset were compared. In the proposed solution, one neuron is used as the output layer for each architecture. For the MobileNetV2 (MNv2) [44] architecture, a fully connected layer consisting of 256 neurons was added before the output layer. The same was done for both the VGG16 and the VGG19 [45] architectures, with 256 neurons in the former and 128 neurons in the latter. All networks were pre-trained using the same dataset as in [24]. A QV was assigned to each image. The initial dataset was reduced to 812 samples, which went through preprocessing tasks and were downsized to  $224 \times 224$ px or  $299 \times 299$ px, depending on the network used. The best-scoring architecture was the MNv2, with a 31.20% MAPE. A new model was trained on the dataset from scratch using the MNv2 network, which achieved a MAPE of 21.17%. While there was room for improvement, the approach chosen by the authors of assigning a QV instead of a class had not yet been proposed. Furthermore, the study showed that a fully-trained model could achieve a higher accuracy than pre-trained models in this scenario. The authors concluded that regression in tile sorting is an approach worth examining in depth, as the fully-trained CNN achieved a significantly superior result compared to pre-trained networks. As the authors obtained performance metrics which differ from the ones obtained in other studies, it is harder to compare this solution with others.

The same authors also explore, in another study, [25], the accuracy of 15 pre-trained CNNs using the ImageNet [46] database when retrained to classify marble tile images. As in [12,24], the authors propose an initial preprocessing of images, with a downsizing to  $224 \times 224$ px. Each model's output layer was removed, its weights frozen, and new layers added. The networks' performance was tested by training 25% of the layers, 50%, 75%, and 100%. The 10-fold cross-validation technique was used for training the new layers. Gradient-weighted Class Activation Mapping (Grad-CAM) [47] was used to study the heat maps output. The results show that pre-trained CNNs perform better than conventional classifiers. The best-performing networks were DenseNet201 and DenseNet169 [43] with an accuracy of approximately 83%. The worst-performing network, XC, still maintained an accuracy of 78.73%. It could have been possible to achieve better results by training the CNNs from scratch, as supported by a previous study.

In a study published in 2021, [26], the authors propose a model based on Extreme Learning Machine (ELM) [48] in conjunction with LBP [33,34] and histogram feature extraction methods. With feature extraction methods, information from the input images is stored in text files to be used with classifiers such as SVM [35,36], Artificial Neural Network (ANN) and ID3 [49–51]. The authors use a single-layer feed-forward ANN, with ELM as its learning algorithm. The ELM algorithm is tested against a typical ANN model, as well as SVM, C4.5 [52], and ID3 classifiers. The authors use a dataset containing 200 microscopically scaled marble images uniformly distributed across four classes, at  $320 \times 240$ px. The classifiers are evaluated using 10-fold cross-validation. The ELM presents a high accuracy result at 97.5%. The ANN model can achieve similar results, but the lightweight nature of the ELM model makes it more suitable for marble classification tasks. The authors state that due to the ELM model's lightweight nature, it could be implemented in a mobile platform.

A study from 2024, [13], compares four CNN models trained on a dataset containing four classes of marble, classified according to characteristics such as texture, cracks, thickness, cut directions, and density. The authors also compare the results of the architectures for greyscale and colour images. A dataset containing 441 images of marble slabs, with a resolution of  $400 \times 700$ px, uniformly distributed across three classes based on texture, cracks, thickness, cut direction, and density, is used. Each of the four CNNs was trained once for coloured images and another time for greyscale images. The conclusion is that ResNet-152 achieves a higher accuracy score with greyscale images, at 100%, against 98.6% with coloured images. The Inception model also achieved 100% accuracy for greyscale images while being more efficient in execution time. The CNNs would have to be trained using other datasets for different marble classes. However, the results are surprising, and custom solutions can be easily derived after obtaining new data.

All relevant metrics provided by the authors are available in Table 4. In studies where several models were tested, the most relevant were selected to be displayed in the table.

### 3.2. Stone processing

Some stages of the stone transformation process can adopt new technologies. This applies to stone cutting and crack detection, as illustrated in Table 5.

#### 3.2.1. Crack detection

In [14], the authors propose a solution able to detect a large percentage of cracks. The authors analyse the performance of four deep CNN models, each with 28 different backbones, on a dataset comprising 264 images of marble slabs with cracks labelled, with a size of  $256 \times 256$ px. This dataset was built upon the Marble Surface Anomaly Detection 2 dataset, initially published in [54]. The model/backbone combinations were compared to a pre-trained model, DeepCrack [55], which had been trained on cracks

**Table 4**  
Results for tile/slab classification articles.

Ref.	Dataset	Method	Accuracy	Precision	Recall	MAPE
[10]	Bianconi et al. [32]	Modified CNN, 32 × 32px patches	97.44%	–	–	–
		Modified CNN, 30 × 30px patches	95.69%	–	–	–
[11]	2260 sample images of limestone slabs, divided by four classes, in PNG format	Inception-V3 feature extraction with LR	98.1%	98.1%	98.1%	–
	2260 sample images of limestone slabs, divided by four classes, in JPG format		97.8%	97.8%	97.8%	–
[13]	D-DOM (Grey-scale) [53]	ResNet-152, Inception	100%	–	–	–
	D-DOM [53]	ResNet-152, VGGNet	98.6%	–	–	–
[21]	Bianconi et al. [32]	CIFAR (fully trained, w/majority voting)	100%	–	–	–
		CIFAR (fully trained, no majority voting)	87.26%	–	–	–
		CIFAR (pre-trained)	94.84% ± 0.52	–	–	–
[22]	195 sample images of marble slabs, equally divided by three classes	LBP feature extraction with DT, no image processing	86.67%	–	–	–
[23]	2100 sample images of marble slabs, equally divided by six classes; data augmentation was applied, resulting in 6300 samples	Custom CNN	92.2% (on test dataset) 96.1% (10-fold cross-validation)	–	–	–
[24]	489 sample images divided by three classes	DenseNet201 DDML	79.78%	80.78%	79.30%	–
		DenseNet201 TL	80.77%	83.34%	79.78%	–
[12]	812 sample images, each assigned a QV between 0 and 1, in increments of 0.1	MobileNetV2 TL	–	–	–	31.20%
		MobileNetV2 Fully Trained	–	–	–	21.17%
[25]	The same as in [24]	DenseNet201	83.24%	82.98%	82%	–
		DenseNet169	82.84%	83.97%	81.56%	–
[26]	200 sample images equally divided by four classes	Histogram + LBP feature extraction with ELM classifier	97.5%	–	–	–

on other surfaces, such as asphalt and concrete. The images go through preprocessing tasks, namely Contrast-Limited Adaptive Histogram Equalization (CLAHE) [56], and other specific tasks are applied depending on the backbone. The DeepCrack model was also retrained on the marble crack dataset, using 5-fold cross-validation. The best results came from the combination of the Feature Pyramid Network (FPN) [57] model with SE-ResNet [58] family backbones, namely seresnext50, which achieved an IoU score of 71.35%. More performance metrics are available in Table 6. The authors claim that noise in the images affected the results of the

**Table 5**

Summary of articles related to stone processing. Abbreviations: CD = Crack Detection, SC = Stone Cutting.

Ref.	Sub.	Summary	Outcomes
[14]	CD	<ul style="list-style-type: none"> <li>• Comparison between four deep CNN models and model pre-trained on DeepCrack dataset</li> <li>• All models are combined with 28 feature extraction networks pre-trained on the ImageNet dataset to identify the best combination for the scenario</li> <li>• Images are pre-processed using CLAHE, other specific tasks are applied depending on the backbone</li> </ul>	<ul style="list-style-type: none"> <li>• The most efficient model is FPN: 71.35% IoU</li> <li>• Best results from Se-ResNet family of backbones: seresnext50, followed by seresnet101, seresnet50 and seresnext101</li> </ul>
[27]	SC	<ul style="list-style-type: none"> <li>• Improved Adaptive Genetic Algorithm</li> <li>• Can dynamically adapt crossover and mutation probability values throughout the searching process</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from proposed solution: fitness average of 1.420 vs. 1.344 (with CP of 0.7 and MP of 0.05)</li> <li>• Prevents premature convergence of local optimal solution</li> </ul>

**Table 6**

Performance metrics for crack segmentation experiments.

Ref.	Dataset	Method	mIoU	Precision	Recall	F1
[14]	Marble Crack Segmentation (MCS) Dataset (available in [59], based on [54]) - 246 sample images with cracks annotated	FPN network with various backbones	71.35%	91.54%	78.71%	N/A
		CNN with seresnext-50 backbone	77.30%	89.91%	86.08%	82.75%

models, as well as the presence of irregular cracks. Furthermore, as it is unavailable, the dataset lacked enough data to produce more robust models.

The authors point out that the research conducted for this study is a starting point for developing a robotic marble resin application system capable of identifying cracks in the slabs and applying resin. Such a solution would drastically reduce the time spent manually detecting and filling cracks, leading to higher production rates.

As the authors concluded, there is a limited amount of data for training marble crack detection models. More data would need to be collected to produce more robust models. Using thermal images could help better identify cracks to further increase the model's performance. An automatic resin application system would require robust models to increase confidence in the system and phase out manual inspection.

### 3.2.2. Stone cutting

The authors of [27] propose an improved adaptive genetic algorithm. The genetic algorithm (GA) [60] is an AI technique based on the processes of natural selection in biological evolution. With the GA, a population of candidate solutions is iteratively improved. The adaptive GA proposed by the authors can find optimal values for mutation and crossover probabilities throughout the searching process, thus mitigating the possibility of early maturity of the model. After testing, the proposed algorithm achieved a better fitness average (1.420) when compared to the best traditional GA result (1.344, with a Crossover Probability (CP) of 0.7 and a Mutation Probability (MP) of 0.05). The complete set of results obtained by the authors is available in Table 7. Furthermore, the improved algorithm proved to prevent premature convergence of a locally optimal solution.

No other studies on the stone slab cutting problem were found. However, this study's results are quite impressive. The solution proposed by the authors can find a cutting layout that minimises material waste while significantly reducing the time it takes to find the layout compared to manual planning.

One possible improvement over the proposed solution would be to employ a system capable of detecting the dimensions of the slab and the largest rectangle available instead of having the user manually input them. Computer vision could aid in implementing this proposal. This way, the process would be nearly completely streamlined.

Using a solution such as the one proposed by the authors of this study would allow experts to focus on other parts of the production process, thus increasing productivity. As such, for this specific problem, possible challenges that might arise do not seem to overshadow the benefits.

### 3.3. Prediction tasks

Table 8 summarises prediction task articles.

**Table 7**  
Performance metrics for Stone Cutting experiments.

Method	Crossover probability	Mutation probability	Avg. fitness
Traditional GA	0.6	0.05	1.339
		0.1	1.334
	0.7	0.05	1.344
		0.1	1.342
	0.8	0.05	1.282
0.1		1.308	
0.9	0.05	1.290	
	0.1	1.326	
Improved GA	Adaptive	Adaptive	1.420

**Table 8**

Summary of prediction task articles. Abbreviations: VP = Value Prediction, PP = Profitability Prediction, OCP = Operation Costs Prediction, NP = Noise Prediction, PRP = Production Rate Prediction.

Ref.	Sub.	Summary	Outcomes
[15]	VP	<ul style="list-style-type: none"> <li>• Comparison between RNN and ARIMA models</li> </ul>	<ul style="list-style-type: none"> <li>• Best accuracy from ARIMA model: 95.7%</li> <li>• RNN model presents an accuracy of 91.8%</li> <li>• Prices of limestone are most affected by cement and gold prices, as well as interest rates</li> </ul>
[16]	PP	<ul style="list-style-type: none"> <li>• Comparison between ANN and MVR models</li> <li>• WipFrag software to evaluate blast fragmentation efficiency and optimise blast design</li> </ul>	<ul style="list-style-type: none"> <li>• Best performance from ANN model: coefficient of determination (<math>R^2</math>) at 0.996 and lower error metrics</li> <li>• MVR achieved <math>R^2</math> of 0.985</li> </ul>
[17]	OCP	<ul style="list-style-type: none"> <li>• Comparison between DT, RF, LR and LGBM Regressor models</li> <li>• Predict machine hours and fuel consumption for hauling, loading, and dozing operations</li> </ul>	<ul style="list-style-type: none"> <li>• DT model performed best for predicting machine hours for dump trucks, with the lowest MAE at 15 h</li> <li>• For excavators and dozers, the LGBM Regressor showed superior accuracy</li> <li>• As for the fuel consumption prediction, the LGBM Regressor was the most accurate for dump trucks, while the Decision Tree excelled for excavators, and the Random Forest model performed best for dozers</li> <li>• The minimum MAE for fuel consumption predictions was approximately 450 litres</li> </ul>
[28]	NP	<ul style="list-style-type: none"> <li>• Prediction of noise produced during drilling of granite samples</li> <li>• Comparison between three different methods</li> <li>• Best performing model used to examine the factors that influence the noise value during drilling activities in multiple scenarios</li> </ul>	<ul style="list-style-type: none"> <li>• Best performance from ANFIS-FCM model</li> <li>• Drill bit hardness is found to be the factor that most influences noise level</li> </ul>
[29]	PRP	<ul style="list-style-type: none"> <li>• Development of PI-BANN model that predicts the blast production rate (BPR) in limestone mining</li> <li>• Comparison between PI-BANN and MVR models</li> <li>• PI-BANN model uses back-propagation and is trained using Bayesian regularisation and Levenberg–Marquardt algorithms</li> <li>• Optimisation of blast design parameters</li> </ul>	<ul style="list-style-type: none"> <li>• Best results from PI-BANN: coefficient of correlation (<math>R^2</math>), at 89.0%, and lower error rates</li> <li>• MVR showed <math>R^2</math> at 81.8%</li> </ul>

### 3.3.1. Value prediction

While [15] focuses more on the variation of the value of limestone as a raw material for other industries, such as the cement industry, it provides valuable insight into the factors that could be considered for an analysis of the price variation of both raw and finished material from the natural stone industry. The models trained by the authors could accurately predict price variations in the short term, given factors such as the price of gold, the production rate of limestone, and interest rates.

To predict variations in the price of raw limestone, the authors suggest two approaches: a model using a Recurrent Neural Network (RNN) based on Long Short-Term Memory (LSTM) [61] and another model using the AutoRegression Integrated Moving Average (ARIMA) [62] technique. RNNs are helpful in prediction problems as they explore the concept of time series. The output of a given time series can be used as input for the next time series. The problem of Vanishing Gradient Descent, however, arises. As such, gradients become very low, impeding the network from learning long-term dependencies. LSTM aims to fix that issue by introducing gates that control the flow of information into and out of the hidden state of the RNN, allowing it to remember important

**Table 9**  
Performance metrics for stone value prediction tasks.

Ref.	Dataset	Method	Accuracy
[15]	Limestone prices from January 2014 to April 2020, obtained from Yahoo Finance, Eoddata, Marketwatch, Indexmundi, and Statista	RNN	91.8%
		ARIMA	95.7%

**Table 10**  
Performance metrics for profitability prediction tasks.

Ref.	Dataset	Method	$R^2$	RMSE	MSE
[16]	Total production and its cost, royalty, other expenses and total revenue for two mines in Nigeria, collected over 8 years	ANN	99.6%	4.355	474.0668
		MVR	98.5%	N/A	N/A

information for long periods. ARIMA is another AI tool for time series forecasting. It can provide predictions for a wide range of data, but results are best for linear data. It assumes that data is linear, stationary, and normally distributed. Outliers can have an adverse and impactful effect on the model's performance.

The models are trained on a dataset containing 1575 points representing data collected daily over six years, where the values were normalised to the range of (0, 1). Through correlation analysis, it is found that cement and gold prices, as well as interest rates, directly affect limestone prices, showing a positive correlation. In this study, the ARIMA model provided more accurate results than the LSTM model (95.7% vs 91.8%). The performance metrics are also available in Table 9. The authors note that the ARIMA model can fine-tune its hyper-parameters automatically, while the LSTM model was tuned manually.

The approach suggested and tested by the authors could be an example of developing a price prediction system for products from the natural stone industry, such as tiles made from marble, granite, or other types of stone. By identifying possible external factors relevant to the market value of stone tiles, it could help companies better understand when to invest or uphold further investment and how to control the production rate of tiles.

Trusting a system's prediction can sometimes be hard for entrepreneurs who have been in the industry for a long time, especially when the system predicts something that goes against their beliefs. To give an example, if the industry sentiment is that investment should be upheld due to market uncertainty and a model predicts that investment should be increased, the prediction would most likely be ignored. To gain the confidence of its users, these models will need to prove their worth through years of testing.

### 3.3.2. Profitability prediction

Similarly to product value prediction, profitability prediction could help managers better understand economic indicators, giving some guidance for more efficient resource management.

One study shows how AI can predict a quarry's profitability, considering variables such as production rate, costs, revenue, royalty, and other operational costs.

In [16], the authors compare two types of models for quarry profitability prediction: ANN and MVR [63]. ANNs are used for problems requiring time series, usually for predicting information given past data. The models take as input production rates, costs, revenue, blast efficiency, and uniformity index, which, according to the authors, traditional models fail to capture effectively. The authors also use WipFrag, a software for evaluating blast fragmentation efficiency and optimising blast design. A dataset containing records obtained over eight years was used to train the models. It was found that the ANN model produced better results, achieving higher coefficient of determination ( $R^2$ ) and lower error metrics when compared to the MVR model ( $R^2$  of 0.996 vs 0.985). More performance metrics are available in Table 10.

Integrating the model with an LLM would be a way of combining this solution with other technologies. A model could also be trained to assess a natural stone transformation company's profits, taking into account the complete process from extraction to processing and selling/shipping.

Quarry managers would need to gain confidence in the system, a process that could take a long time. However, if it is proven to be effective, it would be a beneficial tool for the industry.

### 3.3.3. Operation costs prediction

By predicting machine usage in hours and fuel consumption for a set period in a quarry, it becomes easier to keep up with the maintenance schedule for machines and restock on fuel.

In a study from 2023, [17], four ML algorithms are tested: DT [39], RF [37,38], and Light Gradient Boosting Machine (LGBM) Regressor [64]. After identifying production tonnage as the variable that most influences the time machines spend running, and both those variables influence their fuel consumption, models were trained. A dataset containing records obtained over three years

**Table 11**  
Performance metrics for operation costs prediction tasks.

Ref.	Dataset	Method	MAE (hours of use)	MAE (fuel consumption, in litres)
[17]	Dataset for hauling operations, containing data on fuel consumption, hours of use and production tonnage	DT	15	~2600
		LGBM	~35	~700
	Dataset for loading operations, containing data on fuel consumption, hours of use and production tonnage	DT	~45	450
		LGBM	30	~3700
	Dataset for dozing operations, containing data on fuel consumption, hours of use and production tonnage	LGBM	30	~900
		RF	~45	~500

**Table 12**  
Performance metrics for noise prediction tasks.

Ref.	Dataset	Method	$R^2$	RMSE	VAF
[28]	135 laboratory tests on 5 types of rock	RBF ANN	0.9	0.93	90
		ANFIS-FCM	0.93	0.15	92
		ANFIS-SCM	0.89	0.17	89

in a quarry, including machine hours and fuel consumption for hauling, loading, and dozing operations, was used for that task. The conclusion is that DT was the best model for predicting machine hours for dump trucks, while LGBM was better at predicting machine hours for excavators and dozers. For fuel consumption prediction, LGBM was best for dump trucks, DT for excavators, and RF for dozers. The results were compelling, with a minimum MAE of 15 h for machine hours prediction and a minimum MAE of 450 litres for fuel consumption prediction. The best models' performance metrics are available in Table 11.

A system for predicting and reporting monthly machine hours and fuel consumption could be developed. A nice addition would be to integrate an LLM and advise on maintenance scheduling.

Relying solely on a model may not be enough to run operations at a quarry. External factors must be considered when making important decisions, such as how much fuel to purchase each month. It would take a while for managers to gain confidence in the models.

### 3.3.4. Noise prediction

In [28], three different AI algorithms are analysed: two based on Adaptive-Network-based Fuzzy Inference System (ANFIS) [65] algorithms – ANFIS-Subtractive Clustering Method (ANFIS-SCM) [66] and ANFIS-Fuzzy c-Means (ANFIS-FCM) [67] – and Radial Basis Function (RBF) ANN [68–71]. Given a dataset containing 135 laboratory tests on five rock types, seven factors are considered for the noise prediction. The models are trained and evaluated. The best-performing algorithm in this scenario was ANFIS-FCM, as can be seen in Table 12. Analysing the results, the authors conclude that drill bit hardness has the highest impact on noise level.

Using the model developed by the study authors, a safety recommendation system could be created. One such approach would be to train an LLM to take information on the task to be performed, pass it through the noise prediction model, and, according to its output, recommend safety procedures, such as what equipment to use.

The authors state that while the developed models are reasonably accurate for predicting drilling noise levels for hard rocks, they cannot be accurately used with other kinds of rocks. As such, it would be necessary to expand the dataset and retrain the model to develop a broader safety recommendation system.

### 3.3.5. Production rate prediction

In [29], the authors analyse how AI can be used in quarries to predict blast production rates, given some important parameters. According to the authors, this can help optimise mining performance, minimise costs, and improve efficiency.

The authors compare two different models trained on a dataset containing information on eight parameters for each blast round: Permutation Important-based Bayesian Artificial Neural Network (PI-BANN) and Multivariate Regression (MVR) [63]. A dataset containing 219 entries from blast rounds, each containing eight blast design parameters from limestone blasting, is used to train the models. The PI-BANN model uses back-propagation, and it is trained using two algorithms: Bayesian regularisation and Levenberg–Marquardt. It was concluded that, compared to the MVR model, the PI-BANN model offers higher accuracy, presenting a coefficient

**Table 13**  
Performance metrics for stone production rate prediction tasks.

Ref.	Dataset	Method	$R^2$
[29]	8 measurements for each of 219 blasts conducted in a limestone mine in India	PI-BANN	89.0%
		MVR	81.8%

**Table 14**  
Summary of articles showcasing alternative tasks. Abbreviations: QAM = Quarry Activity Monitoring, PR = Product Recommendations.

Ref.	Sub.	Summary	Outcomes
[18]	QAM	<ul style="list-style-type: none"> <li>• Uses remote sensing data from UAVs, LiDAR and optical sensors</li> <li>• DSMs are derived from aerial LiDAR and satellite images for volume change assessment</li> </ul>	<ul style="list-style-type: none"> <li>• Generated indicators include production rate, natural soil loss, and in-situ MCW/MQW removals and disposals</li> <li>• Indicators are proposed to integrate a DSS for sustainability and environmental impact assessment</li> </ul>
[19]	PR	<ul style="list-style-type: none"> <li>• Chatbot for marble product recommendation</li> <li>• Solution encompasses conversational interface, Flask API to facilitate POST requests, Dialogflow API for conversation processing</li> </ul>	<ul style="list-style-type: none"> <li>• Users answered to a survey containing six questions</li> <li>• Overall positive experience using the chatbot</li> <li>• The chatbot was able to make accurate recommendations in 83% of the cases</li> <li>• If the marble's name consists of multiple words, the system might not be able to fully extract the data from user input, which affects the recommendations</li> </ul>

of correlation of 89.0%, while MVR shows a coefficient of 81.8% and better error metrics. The performance metrics are also available in Table 13. Its performance makes it suitable for optimising mining operations.

The models could be integrated with an LLM, which, given some information on the blast parameters, could output the predicted blast production rate.

The models performed very well using a dataset obtained from a limestone mining quarry. For other kinds of rocks, obtaining new datasets and training new models would be necessary. Furthermore, specific geological characteristics varying from location to location could affect performance. As such, further research would need to be done to understand how much a model's performance is affected between quarries where the same kind of rock is mined.

### 3.4. Other tasks

Besides the tasks analysed thus far, two others were studied: quarry and mining activity monitoring, and product recommendation. The summarised articles are presented in Table 14.

#### 3.4.1. Quarry and mining activity monitoring

Quarrying and mining activities have an environmental impact, related not only to the extraction of the stone but also to the waste created in the process. There are two types of waste in the extraction of marble stone: Marble Cutting Waste (MCW) and Marble Quarry Waste (MQW). This waste can harm a region's landscape and water quality. Therefore, quarries are only allowed temporary MCW/MQW in-situ disposals. MCW is committed to external plants or used in quarry areas for service roads and wall management.

In the Carrara region, where the study takes place, a large industrial basin is home to over 100 active marble quarries. Every year, extraction activities are monitored utilising remote sensing techniques. Unmanned Aerial Vehicle (UAV) and/or terrestrial Light Detection and Ranging (LiDAR) are often used for single quarry surveys, while aerial/satellite high-resolution images and aerial LiDAR are used for basin-wide surveys. Furthermore, quarry owners report production data to local and regional bodies every year, including production volumes of ornamental stone, secondary products, and debris.

In [18], data obtained between 2009 and 2020 is used to assess land cover changes and estimate production volumes. For land cover changes, high-resolution aerial and satellite photos are used to digitise changes in this timeframe, identifying natural soil loss, MCW/MQW removal, and MCW/MQW fills. It is found that land cover changes occurred mainly between 2013 and 2016, areas subjected to natural soil loss and MCW/MQW removals between 2009 and 2020 are higher than the ones interested by MCW/MQW

fills, being respectively the 29% and 35% of total land cover changes' surface; however, between 2017 and 2020 natural soil loss has been significantly lowered concerning years 2009–2016.

To assess 3D volume change precision, a comparison between reported production volumes with computed volume changes between 2017 and 2020 for the top 15 quarries with the highest ornamental stone production rates was made. Upon further investigation, four were dismissed as 2020 terrain models could not be used due to matching errors in complex morphology areas. In most cases, the extracted volume obtained by semiautomatic and manual methods shows little error compared to the reported volume. It was only in three samples that the error was high, with extracted volumes in a range of 30% to 75% lower than reported volumes. While for two of the samples, the discrepancy comes from in-situ displacement of extracted volumes, the 2020 Digital Surface Model (DSM) was inaccurate for one of the quarries. In-situ MCW/MQW volumes obtained from manual methods are far lower than the ones obtained via semiautomatic methods, which highlights that the waste is mainly composed of low-height disposals that could be discarded in manual digitising, besides the issue with DSM precision.

Regarding MCW production, the ratio between reported and estimated values is, on average, around 25%. This does not necessarily imply that the estimation is wrong, as in-situ disposals can cause a discrepancy. However, the accuracy was more satisfactory for quarries with low production rates, given their reduced capability in MCW dismission for industrial purposes.

The research was conducted on in-situ MQW disposal sites, taking into account information from the Regional Ornamental and Building Stones dataset (BDPIOR) made by the LAMMA Consortium for the Tuscany Regional Administration. Using zonal statistics GIS techniques, the authors were able to identify areas subjected to extraction and disposal activities and compute both area and volume changes between 2018 and 2020.

The authors propose including several indicators in a Decision Support System (DSS) for sustainability and environmental impact assessment. In the first category, the proposed indicators include reported and measured production rates, MCW transferred to the recycling industry, and natural soil loss measured in area and volume. As for environmental impact assessment, the proposed indicators include old in-situ MCW/MQW removals, new in-situ MCW/MQW in both area and volume and restored quarries. The DSS could compute a risk score related to waste management performances. However, the terrain model's availability must be improved "by integrating aerial/satellite stereo acquisition with periodic local UAV surveys in quarries with complex morphology unfit for stereo extraction of terrain models".

#### 3.4.2. Product recommendation chatbot

With the popularisation of chatbots to assist customers while shopping, a chatbot for marble recommendation was developed and tested. Chatbots can provide helpful information to customers, often eliminating the need to forward information requests to a human assistant.

In [19], the authors propose a chatbot based on DialogFlow, a platform for creating and integrating application conversational interfaces. A conversational interface was created, consisting of an agent capable of processing human language and an intent — a collection of similar questions matched against the user input. A Flask API was developed to facilitate communication with DialogFlow's API, which provides answers in the JSON format to the chatbot. The chatbot was made public, and a survey asked for user feedback. Users were satisfied with the chatbot's performance, especially in terms of ease of use and ease of understanding. The only complaint was that the recommendations could be affected when a marble class contained more than one word. Still, the chatbot was able to accurately make recommendations in 83% of the cases.

Chatbots are often integrated into companies' web platforms to provide recommendations to customers and help them with frequently asked questions. They serve as a first line of assistance for customers. While they cannot answer every question a customer asks, they can reduce the number of information requests sent for human verification.

To provide robust chatbots, many questions and answers must be planned. The chatbot will be more valuable if it can answer a broader spectrum of questions. Ultimately, there will always be a question without an answer, so it is important to maintain a team of people capable of assisting customers.

Recently, multimodal LLMs (MLLMs) have been researched as to their effectiveness as recommendation systems, with the ability to process images and even sound to provide better informed recommendations. These models could be a good option for stone recommendation systems, as they could search for visual cues given by the user, thus further improving results [72,73]. This technology is yet to be implemented in the stone industry, which reveals an opportunity for future research.

#### 3.5. New technologies

Some studies present new approaches to natural stone production, integrating new technologies into the production chain.

The authors of [20] discuss how the Zero Defect Manufacturing Process (ZDMP) approach can impact natural stone production. ZDMP aims to improve production processes and product quality so that only products without defects are delivered to consumers.

The production of natural stone tiles involves a process comprising four stages. First, the stone is extracted in blocks, from which rough slabs are cut. The slabs are then assessed as to their thickness and, if they do not conform to the requirements, are disposed of as waste. Slabs that conform to the requirements are polished. The pre-cutting stage is next in line, and silhouette moulds are projected onto the slabs, which are then cut to size. After finalising the process, the slabs are visually analysed to confirm their quality.

The ZDM approach proposed by the authors implements new technologies in the process. Table 15 summarises the new process.

By employing the suggested approach, the authors predict a 35% automation of the detection of natural defects, an improvement in production defect detection from 80% to 100%, a waste reduction from 8 to 6 in every 100 parts with a 25% reduction per production stage, and the automation of the detection of worn-out blades with 85% reliability.

**Table 15**  
Proposed ZDMP applications and their purpose.

App	Purpose
Cutting Monitoring	<ul style="list-style-type: none"> <li>• Monitors stone dimensions via computer vision</li> <li>• Detects and locates defects</li> </ul>
zWiresMonitor	<ul style="list-style-type: none"> <li>• Checks the position of the cutting wires</li> <li>• Detects broken wires and irregular wire movements</li> </ul>
zThicknessMonitor	<ul style="list-style-type: none"> <li>• Verifies slab thickness</li> </ul>
zDetectDefects	<ul style="list-style-type: none"> <li>• Detects defects by using CV and AI</li> <li>• Assigns moulds avoiding defects</li> </ul>
zWornOutBladeDetection	<ul style="list-style-type: none"> <li>• Detects worn blades during the cutting process</li> </ul>
zTilesConformity	<ul style="list-style-type: none"> <li>• Using data collected during the stone transformation process, verifies compliance of the final product</li> </ul>

**Table 16**  
Summary of identified problems and possible solutions involving new technologies.

Stage	Problems	Solutions
Crane Lifting	<ul style="list-style-type: none"> <li>• Stone blocks are lifted using overhead cranes</li> <li>• Most injuries and deaths occur in this stage</li> </ul>	<ul style="list-style-type: none"> <li>• Use of robotised overhead gantry crane, remotely operated</li> <li>• It can be caged by a steel fence</li> </ul>
Rock Slicing	<ul style="list-style-type: none"> <li>• The stone is sliced using large-diameter circular saws driven by high-power electric motors</li> <li>• Workers are in risk of injury due to their proximity to the saws</li> </ul>	<ul style="list-style-type: none"> <li>• Use of safety shields on the saws</li> <li>• Laser measurement can increase product quality</li> </ul>
Square Cutting	<ul style="list-style-type: none"> <li>• The stone is cut into smaller slabs, according to the pretended size</li> <li>• This process is done manually</li> </ul>	<ul style="list-style-type: none"> <li>• Adoption of modern machinery, such as the multi-wire stone cutting machine</li> </ul>
Surface Finishing/Engraving/Shaping	<ul style="list-style-type: none"> <li>• Stone slabs must be polished and smoothed</li> <li>• Accuracy and precision are necessary to ensure quality</li> </ul>	<ul style="list-style-type: none"> <li>• Adoption of automatic machines</li> <li>• Use of CV and AI to detect defects</li> </ul>
Sorting	<ul style="list-style-type: none"> <li>• Critical process, usually done by experts</li> <li>• Sorting criteria is not clear and varies according to the expert</li> </ul>	<ul style="list-style-type: none"> <li>• Use of ML to classify slabs</li> <li>• Employ robotic cranes/ industrial arms for automatic sorting according to the model's classification</li> </ul>
Internal Moving	<ul style="list-style-type: none"> <li>• The final product is moved inside the factory using big trucks</li> </ul>	<ul style="list-style-type: none"> <li>• Move products using automated vehicles/forklifts</li> </ul>

Another study, [13], previously analysed in Section 3.1 in the context of tile classification, also proposes a framework for integrating new technologies within the stone processing workflow. The authors highlight the importance of providing products of higher quality and reducing material waste, as well as electricity and water. Table 16 presents the solutions identified for each stage of the stone transformation process.

Although improving the quality of products helps reduce material waste and, by extension, electricity and water, sustainability issues remain. The authors identify a set of issues and propose solutions involving AI. Table 17 summarises the solutions.

#### 4. Discussion

The present study has thus far provided an overview of articles that focus on the use of AI techniques applied to the natural stone industry. In the following pages, we will answer the RQs according to our findings.

**RQ1 — How has AI been applied to various aspects of the natural stone industry, and what are the key benefits and challenges associated with these applications?**

As seen in Fig. 2, most of the AI studies for the natural stone industry published since 2017 focus on tile/slab classification.

**Table 17**  
Proposed solutions to mitigate sustainability issues.

Objective	Solutions
Electricity and Fuel Saving	<ul style="list-style-type: none"> <li>• Install photovoltaic solar panels</li> <li>• Automate machinery</li> </ul>
Water Saving	<ul style="list-style-type: none"> <li>• Install solenoid valves</li> <li>• Implement water flow rate control systems</li> </ul>
Wastewater Treatment	<ul style="list-style-type: none"> <li>• Install smart slurry presses for water/solid separation</li> <li>• Recycle water for reuse</li> </ul>
Rock Waste Circulation	<ul style="list-style-type: none"> <li>• Crush rock waste, which can be used for different products</li> <li>• Manufacture street tiles</li> </ul>
Air/Sound Pollution	<ul style="list-style-type: none"> <li>• Use dust filter for air filtration</li> <li>• Install sound isolators</li> </ul>

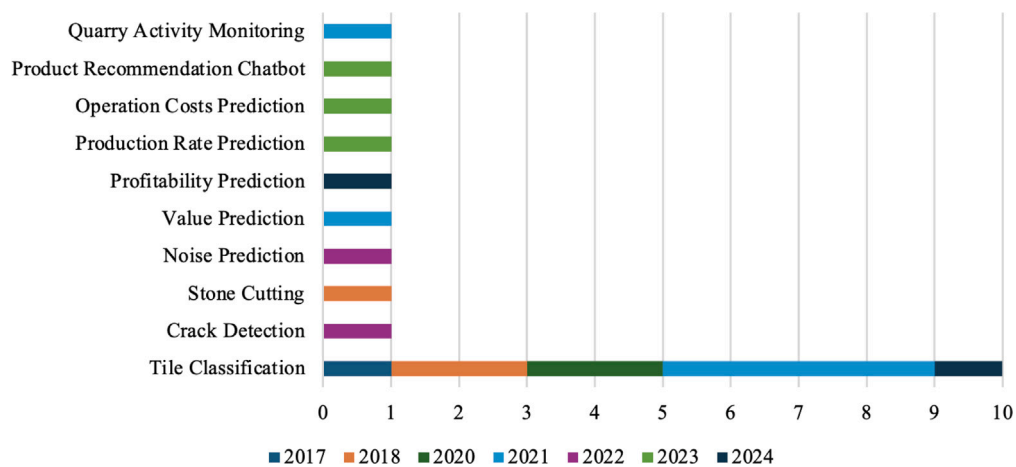


Fig. 2. Number of articles published by subject and year.

Tile Classification is a task that is usually done manually by experts. Each tile must be classified according to its visual characteristics, e.g. colour and texture; and its quality, usually determined by the quantity, size, and orientation of veins and cracks. Given the sheer number of implied characteristics, misclassification often takes place. Upon receiving a tile that does not conform to the original requirements, a customer who ordered a tile with a given set of characteristics may feel discontent; this is only aggravated when an order comprises several tiles that should be visually similar to each other, as wrong tiles may affect the quality of the finished product/construction. Tile classification can be made using computer vision allied with AI. A model trained to identify a given set of characteristics can rapidly and successfully classify tiles. Usually, supervised learning is recommended, as the model uses a custom dataset. While the datasets used for research often contain a variety of image samples representing different varieties of stone, which leads to more robust models, the models are expected to behave differently in a real use-case scenario. Most datasets contain clean images, obtained in a standardised way, lacking noise which can impact the models' real accuracy, as well as lighting variations. There is a need for further research on the impact of noise and lighting conditions on classification models' accuracy. Either the models need to be tested in an industrial environment, or the datasets must be able to account for environment noise and varied lighting conditions. To improve model robustness in industrial settings, some pre-processing tasks can be made to mitigate noise in the images fed into the classification model, including the detection and isolation of the relevant area of classification. Furthermore, the datasets used for model training can be augmented by using techniques which vary image characteristics such as brightness or saturation, as well as by introducing some noise or blur.

While the tile classification subject reveals abundant existing research, other subjects featured only one study, implying many research opportunities.

One problem in the industry is finding the optimal layout for cutting stone slabs. Sometimes, a slab must be cut into several pieces of varied dimensions. An expert usually determines the layout, and stone material is often wasted. By using an AI algorithm, it becomes possible to find more efficient layouts. While in most cases, there is no way to avoid wasted material, this can be largely reduced.

Often, marble slabs contain cracks that may affect their structural integrity. Usually, slabs are treated with resin over their complete surface to fill as many cracks as possible, thus avoiding slab-by-slab verification, a time-consuming process. However, this

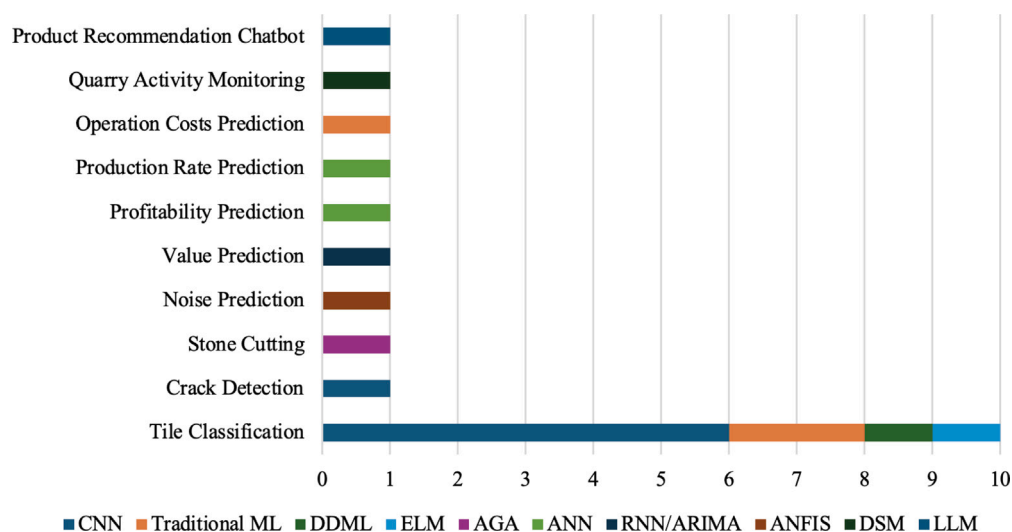


Fig. 3. Number of articles published per main AI technique and subject.

process may not fill deeper cracks, and these must be manually fixed. AI can be employed to avoid manual inspection and speed up the resin application process.

Noise exposure can lead to several health issues, the most common being hearing loss. AI can be used to predict noise levels during drilling activities.

Another area where AI can be employed is in estimating variations in product value, where it could help manage production rates and investment.

Furthermore, AI can be used in quarries to predict blast production rates, which can help optimise mining performance, minimise costs and improve efficiency [29].

Models can also assist in predicting machine usage in hours and fuel consumption for a set period in a quarry, making it easier to keep up on the maintenance schedule for machines and restock on fuel.

Similarly to product value prediction, profitability prediction could help managers better understand economic indicators, giving some guidance for more efficient resource management. One study shows how AI can predict a quarry's profitability, considering variables such as production rate, costs, revenue, royalty, and other operational costs.

Lastly, turning the focus to customers, AI solutions can be employed for assistance. Chatbots can provide helpful information to customers, often eliminating the need to forward information requests to a human assistant. A chatbot for marble recommendation was developed and tested.

## RQ2 — What specific AI techniques and algorithms have been most effectively employed in the natural stone industry, and why?

As seen in Fig. 3, the AI technique that has been most employed in the natural stone industry by far is CNNs. However, these are most used in image classification/segmentation. Seven out of eleven studies in this context use CNNs as the primary AI technique, which represents approximately 63.63% of the total number of studies. When it comes to other problems, CNNs are not used.

While traditional ML can still be used with good results in specific scenarios, its usage has been on a decline. In fact, in this specific context, only three studies look into traditional ML: two for tile classification and one for predicting operation costs. In the case of tile classification, as seen in Fig. 4, the last article exploring traditional ML was published in 2020. Since then, all other studies have focused on using DDML, ELM, and CNNs.

For tile classification, the results indicate that both CNN and feature extraction networks combined with traditional ML algorithms can achieve high accuracy scores. While it is impossible to directly compare all studies with each other, as the authors use various datasets with different characteristics, we can reach some conclusions. Two older studies focused on the use of simpler CNNs, such as CIFAR, for image classification. Both used the dataset provided by Bianconi et al. [32] and divided each image in smaller blocks. In [21], the accuracy of the CIFAR network, when trained from scratch, reached 87.26%. In [10], a modified CNN showed an improvement of over 10% in accuracy, having reached a score of 97.44%. In [21], after performing majority voting, all images were correctly classified. More complex CNNs, such as ResNet-152, VGGNet and Inception show great accuracy results on high-resolution images. In fact, ResNet-152 and Inception networks reached a 100% accuracy result [13] on the greyscale D-DOM dataset [53]. Feature extraction networks, when combined with traditional ML algorithms also show great results. In [11], the Inception-V3 feature extraction network was used with several ML algorithms. The best-performing algorithm was LR, and results indicated an accuracy of 98.1%, with similar values for precision and recall metrics. In [26], the authors used histogram and LBP to extract features from marble tile images, and proceeded to use the ELM classifier. The results indicate a 97.5% accuracy. Despite these good results, CNNs are generally regarded as being the better choice for image classification. In traditional approaches, the feature

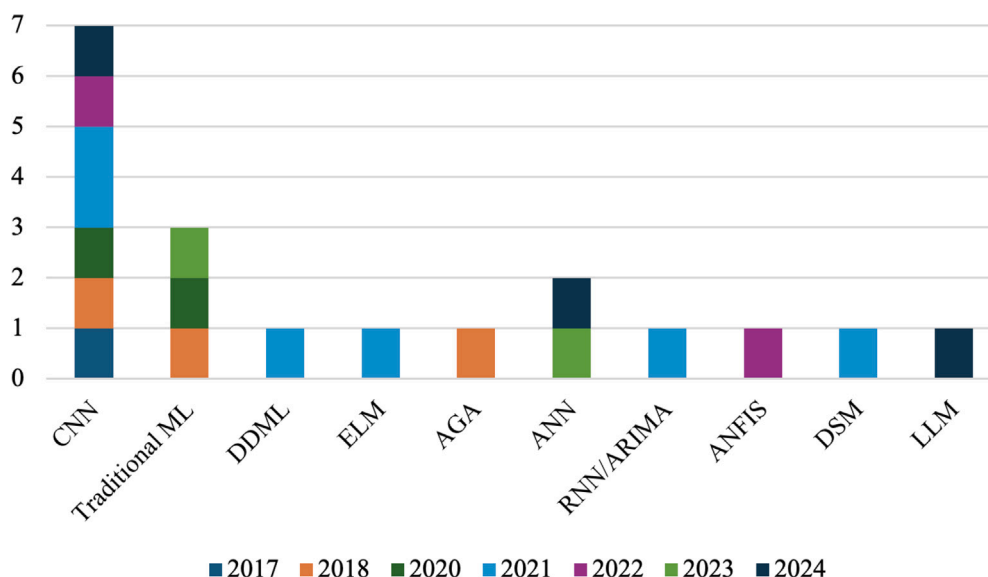


Fig. 4. Number of articles published per main AI technique and year.

extraction and ML algorithm work independently. That means that the ML algorithm will only produce good results if the extracted features allow it to correctly classify an image. While this approach can be very effective in small datasets, or datasets where images share similar features, deep learning approaches typically show better results in larger, more diverse datasets with higher resolution images [74].

Overall, for classification and segmentation tasks, CNNs are the preferred AI technique. In fact, seven out of eleven articles explore the use of CNNs for these tasks, all showing good results — in classification tasks, it is common for accuracy values to be superior to 90%. Their ability to extract and learn spatial features from images makes them an ideal solution for image classification and segmentation.

For prediction tasks, it is more difficult to generalise which algorithm is best due to the limited number of available articles and their focus on different prediction tasks. In essence, there is not one single algorithm that is ideal for all prediction tasks, rather the choice of algorithm depends on the task at hand.

For problems involving the prediction of values over time, such as costs and profits, other algorithms based on ANNs or RNNs are preferred. This is because these kinds of algorithms are typically better for problems involving time dependencies. While some problems exist, namely overfitting, they can make accurate predictions and show good robustness when correctly trained. They can also quickly adapt to changes and learn from new data. In regards to the prediction of production rates, as well as the prediction of quarry profitability, ANNs proved to be more accurate than MVR algorithms (see Tables 10 and 13). For the prediction of limestone prices, ARIMA was more accurate than an RNN based on LSTM (see Table 9). The authors state that RNNs typically struggle in learning long-term dependencies, which is why they used LSTM to avoid the Vanishing Gradient Descent problem. However, ARIMA was more accurate than both a traditional RNN and an LSTM-based RNN. The authors note that the ARIMA model fine-tunes its hyperparameters automatically, whereas the RNN was tuned manually, and this could have affected the results.

For the stone cutting problem, where it is required to find an optimal layout given the number of cuts that must be made to a slab, GAs are the technique of choice, specifically AGA. These algorithms mimic the process of natural selection and evolution, starting with a population of potential solutions, each given a fitness value. The fittest candidates are likelier to produce a new generation of potential solutions. There is a probability for crossover, where information is exchanged between “parents”. Random mutations are also introduced, which allows the exploration of different solutions and prevents premature convergence. Adjusting the probabilities for crossover and mutation to occur through the searching process distinguishes an AGA from a traditional GA. In [27], the authors tested a traditional GA with various CP and MP value combinations, but obtained a better fitness average score when using the proposed AGA. The best fitness average score obtained with the traditional GA was of 1.344, while the AGA achieved a score of 1.420.

In the case of prediction of noise levels, a selection of ANFIS algorithms were compared, with the best being ANFIS-FCM [28]. These algorithms are known to be able to handle imprecise information and are well-suited for nonlinear problems. The results were very good and allowed for an analysis of which parameters most influence noise levels. They are also a good option for time series predictions, such as for financial forecasting [75–77]. The authors also compared the ANFIS models with an RBF ANN model. While the  $R^2$  value was similar to the values obtained using the ANFIS models, the RMSE was significantly greater. The metrics are available in Table 12.

One recent study shows how AI can be used in conversations with customers to recommend a product given a set of conditions [19]. In this case, a chatbot was trained to understand user input and answer accordingly, with a high success rate.

Customers stated that the overall experience was enjoyable, and the recommendations were mostly on point. While the use of MLLMs for stone recommendation systems is yet to be studied, they might be a good solution, given their ability to process images and other data formats.

### **RQ3 — How could AI be integrated with other technologies to enhance value and efficiency in the natural stone industry?**

Any good AI solution must be easy to use by industry workers. As such, an AI model cannot exist without some sort of user interface.

For tile classification, two possibilities come to mind, depending on the workers' needs and the stage in which classification is required. The first solution involves classification in the production line, where a camera would be permanently mounted. A device could be assigned to the production line with the sole purpose of providing tile classification results. The other solution would require the use of mobile devices, i.e. smartphones and/or tablets. An application could allow workers to perform classification outside the production line, for example, before shipping. This solution allows for more flexibility, but new challenges arise. In this scenario, the background would need to be ignored so that only the tile/slab can be classified. In a production line, where a camera is mounted in a fixed position, and tiles always move through the same place, it is easier to fine-tune parameters to allow only the stone itself to be fed into the classification model. In a mobile solution, one way to do this would be to perform segmentation before classification so that the tile/slab is extracted from the complete scenario.

Certain prediction problems involving time series, where it is important to receive alerts and reports, such as profitability/material value prediction, could be solved by developing a notification system. It could even be integrated within a management web platform or mobile application, where statistics and predictions could be presented in charts. Alerts could be triggered given certain parameters specified by the user and shown via notifications or emails.

Other types of predictions could also be integrated within a web platform or a mobile application used by workers. One such example is noise prediction. A page asking for specific parameters regarding the specifications of the stone to be drilled and the equipment used would allow for the model to predict noise levels during drilling operations. According to the results, safety recommendations could be given.

On the topic of chatbots, research is still heavily lacking. One study, [19], where a chatbot was implemented for marble recommendation, shows good results. However, LLMs could be a more advanced solution for recommendation systems, especially given their ability to process natural language, and MLLMs could provide even more advantages, as they are able to process other data formats, such as images and videos. In fact, MLLMs could be applied to a multitude of areas in the stone industry, from the production chain to customer interaction. In the production environment, they could be a valuable tool for workers to quickly access information which would otherwise require the use of specific, standalone tools. The MLLM could be accessed from mobile devices at all times, saving workers time. It could also integrate other models, such as ones for noise prediction, tile classification or crack detection. Managers could have access to several prediction algorithms and obtain information on operation costs, production rates, quarry profitability, product value and more through queries to the model. On the customer side, it could be useful to advise on a variety of situations, and also to track order status. Their flexibility makes them a simple to use, easily accessible interface to obtain data anywhere.

### **RQ4 — What are AI adoption's potential challenges and limitations in the natural stone industry, and how can these be overcome?**

Using AI to conduct tile classification tasks can carry some limitations. The main limitation would be to ensure accurate classification under varying light conditions. In a production line, using a fixed camera and artificial lighting allows for a consistent environment in which the model operates. Furthermore, a model trained on a dataset taken from the production line it operates in is more likely to produce better results. Should a mobile solution be developed, it would be important to assess the accuracy of the results under different scenarios. In such a critical task, where misclassification may lead to heavy profit losses, it is important to ensure reliability and accuracy under varying scenarios.

In the case of crack detection, the authors state that a robotic resin application system could be developed. Given that an expert usually has to manually inspect the tile for cracks and apply resin to fill them, an automatic system would benefit the industry, increasing production rates and profit.

Another problem exists regarding profitability/material value prediction: how can managers trust a new prediction system, especially when it sometimes seems to contradict their intuition? Unfortunately, there are no simple answers or solutions. The model will have to prove itself reliable and robust throughout its operation. Even then, there will probably always be some hesitance towards such a solution. The same goes for operation cost prediction.

## **5. Conclusions**

This SLR summarises and discusses recent research on applying AI to the stone manufacturing industry, which will encourage further research.

Throughout this SLR, which adopted the methodology proposed in [3], we gathered and analysed 20 articles encompassing several fields of study. In an industry that still relies on so many traditional manufacturing processes, it is clear that research is still lacking.

From eligible articles, we found that the study fields encompass tile classification, optimal slab-cutting layout prediction, crack detection, several value prediction tasks, quarry monitoring, and chatbots. CNNs were the preferred AI technique for image classification tasks, while other RNN-based techniques were used more often for value prediction tasks. Crack detection required the

use of CNNs for segmentation. An adaptive genetic algorithm was used to find the optimal slab-cutting layout with excellent results. Quarry activity monitoring requires high-resolution aerial images from UAV surveys and satellites for land cover changes, which can positively impact sustainability in the natural stone industry. As for the product recommendation chatbot, an LLM was adapted to parse questions regarding the choice of different marble types according to their use. Two studies addressed the production chain from rock extraction to processing and finishing, proposing new technologies to improve the quality of stone products while reducing the industry's environmental impacts and improving workers' safety conditions.

We used a semi-automated process to filter over 1500 articles from four scientific databases down to 20 to achieve this goal. The selection process was rigorous, as required by the methodology followed; strict selection criteria were defined, including a quality assessment stage, to ensure transparency and rigour in the process. Scripts were created to automate the article selection process according to the search string defined in the planning stage of this review. This approach accelerated the selection process by automatically discarding over 200 articles from Web of Science, more than 1000 from ACM, and over 50 from IEEE Xplore. However, each of these articles was briefly analysed manually to review their titles and, in some cases, abstracts to ensure that no relevant articles were wrongfully discarded. A thorough discussion of the results obtained from the analysis of the articles allowed for a better understanding of the technologies used for different tasks. Four research questions regarding the benefits, challenges, and limitations of adopting AI in the different stages of the production process, the effectiveness of AI techniques, and possible integrations with other technologies were answered according to the data we obtained.

This paper introduces new researchers to past studies and provides some paths for future research. By combining existing information into one document, we aim to provide a solid foundation and incentivise further research in the industry.

### CRedit authorship contribution statement

**Alexandre Silva:** Software, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization. **Carolina Antunes:** Investigation, Writing - original draft. **Rolando Miragaia:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Rogério Luís Costa:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Fernando Silva:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition. **José Ribeiro:** Conceptualization, Methodology, Validation, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

### Abbreviations

**AI** Artificial Intelligence

**ANFIS** Adaptive-Network-based Fuzzy Inference System

**ANFIS-FCM** ANFIS-Fuzzy c-Means

**ANFIS-SCM** ANFIS-Subtractive Clustering Method

**ANN** Artificial Neural Network

**ARIMA** AutoRegressive Integrated Moving Average

**CADOICO** Cadoico Azul Mónica Silva

**CAGR** Compound Annual Growth Rate

**CLAHE** Contrast-Limited Adaptive Histogram Equalization

**CIFAR** Canadian Institute For Advanced Research

**CNN** Convolutional Neural Network

**CP** Crossover Probability

**DDML** Deep Distance Metric Learning

**DL** Deep Learning

**DML** Distance Metric Learning

**DNN** Deep Neural Network

**DSM** Digital Surface Model

**DT** Decision Tree

**DSS** Decision Support System

**ELM** Extreme Learning Machine

**EUR** Euro

**FPN** Feature Pyramid Network

**Grad-CAM** Gradient-weighted Class Activation Mapping

**ID3** Iterative Dichotomiser 3

**LBP** Local Binary Patterns

**LGBM** Light Gradient Boosting Machine

**LiDAR** Light Detection and Ranging

**LLM** Large Language Model

**LSTM** Long Short-Term Memory

**MAE** Mean Absolute Error

**MCW** Marble Cutting Waste

**ML** Machine Learning

**MNIST** Modified National Institute of Standards and Technology

**MP** Mutation Probability

**MQW** Marble Quarry Waste

**MNV2** MobileNetV2

**MLLM** Multimodal Large Language Model

**MVR** Multivariate Regression

**PI-BANN** Permutation Important-based Bayesian Artificial Neural Network

**PICOC** Population, Intervention, Comparison, Outcome, Context

**QA** Quality Assessment

**QV** Quality Value

**RF** Random Forest

**RNN** Recurrent Neural Network

**RQs** Research Questions

**SBM** Salgueira Branco do Mar

**SBR** Salgueira Branco Real

**SE-ResNet** Squeeze-and-Excitation ResNet

**SLR** Systematic Literature Review

**SVM** Support Vector Machine

**TL** Transfer Learning

**UAV** Unmanned Aerial Vehicle

**USD** United States Dollar

**VMF** Vidraço Moleanos Farpedra

**ZDMP** Zero Defect Manufacturing Process

**ZDM** Zero Defect Manufacturing

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

This work is partially funded by national funds through FCT – Fundação para a Ciência e a Tecnologia, I.P., in the context of the project UIDB/04524/2020, and under the Scientific Employment Stimulus - CEECINS/00051/2018. The authors also gratefully acknowledge the European Union - Portuguese Recovery and Resilience Plan, Grant Contract number: Sustainable Stone by Portugal, Call: 2021-C05i0101-02—agendas/alianças mobilizadoras para a reindustrialização - PRR, Proposal: C632482988-00467016.

### Data availability

Data will be made available on request.

### References

- [1] Fortune Business Insights, Natural Stone and Marble Market Size, Share & COVID-19 Impact Analysis, By Type (Granite, Limestone, Marble, Sandstone, Slate, and Others), By Application (Building & Construction, Monumental, and Others), and Regional Forecast, (2022) 2023-2030. <https://www.fortunebusinessinsights.com/natural-stone-and-marble-market-104040>.
- [2] ASSIMAGRA. Dados estatísticos exportações - fevereiro 2024. 2024, [https://assimagra.pt/wp-content/uploads/ESTATISTICA\\_SECTOR\\_FEV2024\\_ANUAL\\_2023.pdf](https://assimagra.pt/wp-content/uploads/ESTATISTICA_SECTOR_FEV2024_ANUAL_2023.pdf).
- [3] Carrera-Rivera A, Ochoa W, Larrinaga F, Lasa G. How-to conduct a systematic literature review: A quick guide for computer science research. *MethodsX* 2022;9:101895. <http://dx.doi.org/10.1016/j.mex.2022.101895>.
- [4] Zio E, Miquelès L. Digital twins in safety analysis, risk assessment and emergency management. *Reliab Eng Syst Saf* 2024;246:110040. <http://dx.doi.org/10.1016/j.res.2024.110040>.
- [5] Guendouz BS, Ouchani S, Assaad HEL, Zaher MEL. A systematic review of federated learning: Challenges, aggregation methods, and development tools. *J Netw Comput Appl* 2023;220:103714. <http://dx.doi.org/10.1016/j.jnca.2023.103714>.
- [6] Jarrah AM, Wardat Y, Fidalgo P. Using ChatGPT in academic writing is (not) a form of plagiarism: What does the literature say. *Online J Commun Media Technol* 2023;13(4):e202346.
- [7] Cornejo J, Barrera S, Ruiz CAH, Gutiérrez F, Casanovas MO, Kot L, Solís MA, Larenas R, Castro-Nieny F, Saavedra MRA, et al. Industrial, collaborative and mobile robotics in latin america: Review of mechatronic technologies for advanced automation. *Emerg Sci J* 2023;7(4):1430–58.
- [8] Rustemi A, Dalipi F, Atanasovski V, Risteski A. A systematic literature review on blockchain-based systems for academic certificate verification. *IEEE Access* 2023;11:64679–96. <http://dx.doi.org/10.1109/ACCESS.2023.3289598>.
- [9] Payawal JMG, Kim D-K. Image-based structural health monitoring: A systematic review. *Appl Sci* 2023;13(2):968. <http://dx.doi.org/10.3390/app13020968>.
- [10] Singh N, SinghSaini B. An improved convolutional neural network for classification of small patches of granite tiles. In: 2018 international conference on advances in computing, communication control and networking. (ICACCCN), IEEE; 2018, p. 1125–30.
- [11] Tereso M, Rato L, Goncalves T. Automatic classification of ornamental stones using machine learning techniques a study applied to limestone. In: Iberian conference on information systems and technologies, CISTI. 2020, p. 1–6. <http://dx.doi.org/10.23919/CISTI49556.2020.9140872>.
- [12] Ouzounis AG, Taxopoulos G, Papakostas GA, Sarafis IT, Stamkos A, Solakis G. Marble quality assessment with deep learning regression. In: 021 fifth international conference on intelligent computing in data sciences. ICDS, IEEE; 2021, p. 1–5.
- [13] Hanieh AA, Hasan AA, AbdElall S, Alhanjouri M. Investigating artificial intelligence and modern technologies enhancement in stone and marble cutting in palestine. *Jordan J Mech Ind Eng* 2024;18(1):207–18. <http://dx.doi.org/10.59038/jjmie/180116>.
- [14] Vrochidou E, Sidiropoulos GK, Ouzounis AG, Lampoglou A, Tsimperidis I, Papakostas GA, Sarafis IT, Kalpakis V, Stamkos A. Towards robotic marble resin application: Crack detection on marble using deep learning. *Electron (Switzerland)* 2022;11(20):3289. <http://dx.doi.org/10.3390/electronics11203289>.
- [15] Mbah TJ, Ye H, Zhang J, Long M. Sing LSTM and ARIMA to simulate and predict limestone price variations, mining. *Met Explor* 2021;38(2):913–26. <http://dx.doi.org/10.1007/s42461-020-00362-y>.
- [16] Taiwo BO, Famobuwa OV, Mata MM, Sazid M, Fissah Y, Jebutu VA, Akinlabi AA, Ogunyemi OB, Ozigi A. Granite downstream production dependent size and profitability assessment: an application of mathematical-based artificial intelligence model and WipFrag software. *J Min Environ* 2024;15(2):497–515. <http://dx.doi.org/10.22044/jme.2023.13731.2543>.
- [17] Surangani RKH, Dawalagala HS, Reval SS, Rodrigo MAJ, Wijerathnayake WMNC, Wickrama MADMG, Wedage WN. Machine learning-based prediction of machine hours and fuel consumption: A case study in aruwakkalu limestone quarry, Sri Lanka. In: Moratuwa engineering research conference, mERCon. 2023, p. 368–72. <http://dx.doi.org/10.1109/MERCon60487.2023.10355468>.
- [18] Licciardello C, Di Marco A, Biagini S, Tayeh K, Palazzuoli D. From remote sensing to decision support system for industrial quarry basins. In: E. Borgogno-Mondino P Zamperlin, editor. *Geomatics and geospatial technologies*. Cham: Springer International Publishing; 2022, p. 385–404.
- [19] Jain A, Soni S. MarbleBot: A conversational recommender and assistance chatbot for marble selection based on dialogflow. In: 2023 international conference on electrical, electronics, communication and computers. (ELEXCOM), IEEE; 2023, p. 1–6.

- [20] Sousa J, Nazarenko AA, Ferreira J, Antunes H, Jesus E, Sarraipa J. Zero-defect manufacturing using data-driven technologies to support the natural stone industry. In: 2021 IEEE international conference on engineering, technology and innovation (ICE/ITMC). 2021, p. 1–7. <http://dx.doi.org/10.1109/ICE/ITMC52061.2021.9570260>.
- [21] Ferreira A, Giraldo G. Convolutional neural network approaches to granite tiles classification. *Expert Syst Appl* 2017;84:1–11. <http://dx.doi.org/10.1016/j.eswa.2017.04.053>.
- [22] Forcado MRG, Estrada JE. Model development of marble quality identification using thresholding, sobel edge detection and gabor filter in a mobile platform. In: 2018 IEEE 10th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management, HNICEM 2018. 2018, p. 1–6. <http://dx.doi.org/10.1109/HNICEM.2018.8666359>.
- [23] Karaali I, Eminagaoglu M. A convolutional neural network model for marble quality classification. *SN Appl Sci* 2020;2(10):1733. <http://dx.doi.org/10.1007/s42452-020-03520-5>.
- [24] Sidiropoulos GK, Ouzounis AG, Papakostas GA, Sarafis IT, Stamkos A, Kalpakis V, Solakis G. Exploiting deep metric learning for marble quality assessment with small and imbalanced image data. In: 2021 IEEE 12th annual information technology, electronics and mobile communication conference, IEMCON 2021. 2021, p. 266–9. <http://dx.doi.org/10.1109/IEMCON53756.2021.9623255>.
- [25] Ouzounis AG, Sidiropoulos GK, Papakostas GA, Sarafis IT, Stamkos A, Solakis G. Interpretable deep learning for marble tiles sorting. In: Proceedings of the 2nd international conference on deep learning theory and applications, deLTA 2021. 2021, p. 101–8. <http://dx.doi.org/10.5220/0010517001010108>.
- [26] Turan E, Ucar F, Dandil B. A novel marble recognition system using extreme learning machine with LBP and histogram features. *Concurr Comput: Pr Exp* 2021;33(21):e6428. <http://dx.doi.org/10.1002/cpe.6428>.
- [27] Dong Y, Chong M, Bai C, Feng M, Li J. Optimization of rectangular layout based on an improved adaptive genetic algorithm in stone cutting. In: Proceedings - 2018 11th international symposium on computational intelligence and design, ISCID 2018. vol. 2, 2018, p. 226–30. <http://dx.doi.org/10.1109/ISCID.2018.10153>.
- [28] Mikaeil R, Piri M, Shaffiee Haghsheenas S, Careddu N, Hashemolhosseini H. An experimental-intelligent method to predict noise value of drilling in dimension stone industry. *J Min Environ* 2022;13(3):693–713. <http://dx.doi.org/10.22044/jme.2022.12092.2206>.
- [29] Taiwo BO, Gebretsadik A, Fissaha Y, Kide Y, Li E, Haile K, Oni OA. Artificial neural network modeling as an approach to limestone blast production rate prediction: a comparison of PI-BANN and MVR models. *J Min Environ* 2023;14(2):375–88. <http://dx.doi.org/10.22044/jme.2023.12489.2266>.
- [30] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images. Technical report, Toronto, Ontario: University of Toronto; 2009, <https://www.cs.toronto.edu/kriz/learning-features-2009-TR.pdf>.
- [31] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. *Proc IEEE* 1998;86(11):2278–324. <http://dx.doi.org/10.1109/5.726791>.
- [32] Bianconi F, Bello R, Fernández A, González E. On comparing colour spaces from a performance perspective: Application to automated classification of polished natural stones. In: V. Murino D Sona MCristiani C Sansone, editor. *New trends in image analysis and processing – ICIAP 2015 workshops*. Cham: Springer International Publishing; 2015, p. 71–8.
- [33] Ojala T, Pietikainen M, Harwood D. Performance evaluation of texture measures with classification based on Kullback discrimination of distributions. In: Proceedings of 12th international conference on pattern recognition. 1, 1994, p. 582–5. <http://dx.doi.org/10.1109/ICPR.1994.576366>.
- [34] Ojala T, Pietikainen M, Harwood D. A comparative study of texture measures with classification based on featured distributions. *Pattern Recognit* 1996;29(1):51–9. [http://dx.doi.org/10.1016/0031-3203\(95\)00067-4](http://dx.doi.org/10.1016/0031-3203(95)00067-4).
- [35] Vapnik VN. The support vector method. In: W. Gerstner MHasler J-D Nicoud, editor. *Artificial neural networks — iCANN'97*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1997, p. 261–71.
- [36] Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20(3):273–97. <http://dx.doi.org/10.1007/BF00994018>.
- [37] Ho TK. Random decision forests. In: Proceedings of 3rd international conference on document analysis and recognition. vol. 1, 1995, p. 278–82. <http://dx.doi.org/10.1109/ICDAR.1995.598994>.
- [38] Amit Y, Geman D. Shape quantization and recognition with randomized trees. *Neural Comput* 1997;9(7):1545–88. <http://dx.doi.org/10.1162/neco.1997.9.7.1545>.
- [39] von Winterfeldt D, Edwards W. *Decision analysis and behavioral research*. Cambridge University Press; 1986.
- [40] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. 2015, arXiv preprint [arXiv:1512.00567](https://arxiv.org/abs/1512.00567), <https://arxiv.org/abs/1512.00567>.
- [41] Kononenko I. Comparison of inductive and naive Bayesian learning approaches to automatic knowledge acquisition. *Curr Trends Knowl Acquis* 1990;8:190.
- [42] Suárez-Díaz JL, García S, Herrera F. A tutorial on distance metric learning: Mathematical foundations, algorithms, experimental analysis, prospects and challenges (with appendices on mathematical background and detailed algorithms explanation). 2018, arXiv preprint [arXiv:1812.05944](https://arxiv.org/abs/1812.05944).
- [43] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 4700–8.
- [44] Howard AG. Mobilenets: Efficient convolutional neural networks for mobile vision applications. 2017, arXiv preprint [arXiv:1704.04861](https://arxiv.org/abs/1704.04861).
- [45] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. 2014, arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556).
- [46] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. ImageNet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009, p. 248–55. <http://dx.doi.org/10.1109/CVPR.2009.5206848>.
- [47] Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual explanations from deep networks via gradient-based localization (iccv). 2017, p. 618–26. <http://dx.doi.org/10.1109/ICCV.2017.74>.
- [48] Huang G-B, Zhu Q-Y, Siew C-K. Extreme learning machine: Theory and applications. *Neurocomputing* 2006;70(1):489–501. <http://dx.doi.org/10.1016/j.neucom.2005.12.126>.
- [49] Quinlan JR. *Discovering rules by induction from large collections of examples, expert systems in the micro electronics age*. Edinburgh University Press; 1979.
- [50] Quinlan JR. Learning efficient classification procedures and their application to chess end games. *Mach Learn* 1983;4:63–82.
- [51] Quinlan JR. Induction of decision trees. In: *Machine learning*. vol. 1, (1):1986, p. 81–106. <http://dx.doi.org/10.1007/BF00116251>.
- [52] Quinlan JR. *C4. 5: programs for machine learning*. Elsevier; 2014.
- [53] Sidiropoulos GK, Ouzounis AG, Papakostas GA, Lampoglou A, Sarafis IT, Stamkos A, Solakis G. Hand-crafted and learned feature aggregation for visual marble tiles screening. *J Imaging* 2022;8(7):191. <http://dx.doi.org/10.3390/jimaging8070191>.
- [54] Rastogi A. Marble surface anomaly detection - 2 available. online, kaggle,. 2021, <https://www.kaggle.com/datasets/wardaddy24/marble-surface-anomaly-detection-2>.
- [55] Liu Y, Yao J, Lu X, Xie R, Li L. DeepCrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing* 2019;338:139–53.
- [56] Zuiderveld K. Contrast limited adaptive histogram equalization. In: *Graphics gems IV*. IEEE; 1994, p. 474–85.
- [57] Lin T-Y, Dollár P, Girshick R, He K, Hariharan B, Belongie S. Feature pyramid networks for object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 2117–25.
- [58] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7132–41.
- [59] Vrochidou E, Sidiropoulos GK, Ouzounis AG, Lampoglou A, Tsimperidis I, Papakostas GA, Sarafis IT, Kalpakis V, Stamkos A. MachineLearningVisionRG/mcs-dataset: marble crack segmentation (MCS) dataset. 2022, <https://github.com/MachineLearningVisionRG/mcs-dataset>.

- [60] Michalewicz Z. Genetic algorithms+ data structures= evolution programs,. Springer Science & Business Media; 2013.
- [61] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput* 1997;9:1735–80. <http://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- [62] Hamilton JD. Time series analysis. Princeton University Press; 1994.
- [63] Alexopoulos E. Introduction to multivariate regression analysis. *Hippokratia* 2010;14:23–8.
- [64] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y. Lightgbm: A highly efficient gradient boosting decision tree. *Adv Neural Inf Process Syst* 2017;30.
- [65] Jang J-SR. ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 1993;23(3):665–85. <http://dx.doi.org/10.1109/21.256541>.
- [66] Chiu Stephen L. Fuzzy model identification based on cluster estimation. *J Intell Fuzzy Syst* 1994;2(3):267–78. <http://dx.doi.org/10.3233/IFS-1994-2306>.
- [67] Rajab Sharifa, Sharma Vinod. A new approach to ANFIS modeling using kernel based FCM clustering. 2016.
- [68] Powell MJD. Radial basis functions for multivariable interpolation: a review. *Algorithms Approx* 1987;143–67.
- [69] Powell JD. Radial basis function approximations to polynomials. In: *Numerical Analysis 1987*. Longman Publishing Group, USA; 1989, p. 223–41.
- [70] Powell MJD, Light WA. The theory of radial basis function approximation in 1990. In: *Advances in numerical analysis, II: wavelets, subdivision algorithms, and radial basis functions*. Clarendon Press; 1992, p. 105–210.
- [71] Powell MJD. Recent research at cambridge on radial basis functions. *New Dev Approx Theory* 1999;132:215–32.
- [72] Liu Y, Wang Y, Sun L, Yu PS. Rec-GPT4v: Multimodal recommendation with large vision-language models, arxiv preprint arxiv:2402.08670. 2024, <https://arxiv.org/abs/2402.08670>.
- [73] Ye Y, Zheng Z, Shen Y, Wang T, Zhang H, Zhu P, Yu R, Zhang K, Xiong H. Harnessing multimodal large language models for multimodal sequential recommendation. *Proc the AAAI Conf Artif Intell* 2025;39(12):13069–77. <http://dx.doi.org/10.1609/aaai.v39i12.33426>.
- [74] Wang P, Fan E, Wang P. Comparative analysis of image classification algorithms based on traditional machine learning and deep learning. *Pattern Recognit Lett* 2021;141:61–7. <http://dx.doi.org/10.1016/j.patrec.2020.07.042>.
- [75] Boyacioglu MA, Avci D. An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: The case of the Istanbul stock exchange. *Expert Syst Appl* 2010;37(12):7908–12. <http://dx.doi.org/10.1016/j.eswa.2010.04.045>.
- [76] Tan Z, Quek C, Cheng PYK. Stock trading with cycles: A financial application of ANFIS and reinforcement learning. *Expert Syst Appl* 2011;38(5):4741–55. <http://dx.doi.org/10.1016/j.eswa.2010.09.001>.
- [77] Ghashami F, Kamyar K. Performance evaluation of ANFIS and GA-ANFIS for predicting stock market indices. *Int J Econ Financ* 2021;13:1.