

Article

CracksGPT: Exploring the Potential and Limitations of Multimodal AI for Building Crack Analysis

Biyanka Ekanayake ¹, Vishal Thengane ², Johnny Kwok-Wai Wong ^{1,*}, Sara Wilkinson ¹ and Sai Ho Ling ³

¹ School of Built Environment, Faculty of Design and Society, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia; biyanka.ekanayake@uts.edu.au (B.E.); sara.wilkinson@uts.edu.au (S.W.)

² School of Computer Science and Electronic Engineering, Faculty of Engineering and Physical Sciences, Alan Turing Building, Stag Hill Campus, University of Surrey, Guildford GU2 7XH, UK; v.thengane@surrey.ac.uk

³ School of Electrical and Data Engineering, Faculty of Engineering and IT, University of Technology Sydney, 15 Broadway, Ultimo, NSW 2007, Australia; steve.ling@uts.edu.au

* Correspondence: johnny.wong@uts.edu.au

Abstract

Building cracks are among the critical building defects, as they can compromise structural integrity, occupant safety and building sustainability. Traditional laborious building inspection methods are cumbersome and erroneous. Computer vision-based crack detection relies on image recognition and does not analyse the underlying causes or suggest rectification strategies. This study explores the potential and limitations of multimodal AI models, that integrate image and text modalities for building crack analysis. As a proof-of-concept, the vision–language model, CracksGPT was built upon a fine-tuned MiniGPT-v2. It was trained on custom crack images with text descriptions detailing visual features, possible causes, and rectification options. It was tested on crack images from a building site in Sydney. When provided with an image of a wall crack, CracksGPT can classify crack patterns of vertical, horizontal, diagonal, and stair-step and interpret possible underlying causes with potential rectification strategies. The ROUGE metric was used for language generation quality evaluation followed by a performance evaluation by building inspection experts. The model’s performance is sensitive to input image quality and training data limitations, specifically in complex scenarios, reaffirming the value of expert overseeing. The findings highlight the potential and limitations of multimodal AI for integrating vision–language reasoning into building inspections.

Keywords: artificial intelligence; building defects; computer vision; cracks; large language models; multimodal models; vision-language models



Received: 6 October 2025

Revised: 6 November 2025

Accepted: 23 November 2025

Published: 28 November 2025

Citation: Ekanayake, B.; Thengane, V.; Wong, J.K.-W.; Wilkinson, S.; Ling, S.H. CracksGPT: Exploring the Potential and Limitations of Multimodal AI for Building Crack Analysis. *Buildings* **2025**, *15*, 4327. <https://doi.org/10.3390/buildings15234327>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Building defects significantly impair the structure’s ability to function properly, shorten its lifespan, and pose safety risks to occupants [1,2]. Identifying such defects, understanding their root causes, and implementing timely rectifications are critical for ensuring long-term sustainability [3]. Among the most prevalent defects in buildings are cracks, cladding issues, and waterproofing failures [4]. These defects are often costly and time-consuming to detect and repair [5]. A major share of the building maintenance budget is allocated to rectify such building defects [4].

Among the building defects, cracks are particularly critical. If not diagnosed and addressed promptly, cracks can severely impact structural integrity and occupant safety [3]. Additionally, cracks can act as entry points for water, leading to significant waterproofing

problems [6]. Crack severity is commonly classified as minor, medium, or severe, often based on measurable crack width [7]. Traditional manual inspection methods rely heavily on subjective judgement and sometimes hazardous when assessing building cracks on mid-to high-rise structures [8].

Advancements in computer vision and deep learning algorithms have significantly improved the automated detection of surface wall cracks to replicate human vision [9,10]. These methods employ cameras and algorithmic techniques to collect, process, and analyse visual information, substantially reducing reliance on labour-intensive and error-prone manual building inspections [9,11,12]. However, visual data alone is insufficient for a comprehensive analysis of cracks. To fully understand the causes of cracks and recommend effective remediation strategies, it is necessary to integrate text information with images. This need has given rise to multimodal AI models, which combine and interpret data from multiple sources or modalities, such as images (vision) and descriptive text (language) [13].

This study introduces CracksGPT, a vision–language model, which explores the potential and limitations of multimodal AI for analysing building cracks. It leverages the strengths of Large Language Models (LLMs), and computer vision to process image–text combinations [13,14] for detecting cracks based on visual characteristics and interpreting possible causes with potential rectification strategies. The proof-of-concept model of CracksGPT was developed to visually demonstrate its potential functionality, applying design science research (DSR) framework [15]. The main research question guiding this study is: “What are the potential capabilities and current limitations of multimodal AI, particularly vision–language models, in analysing building cracks for building inspection?”

The manuscript is structured as follows: Section 1 introduces the study, followed by a review of relevant literature in Section 2. Section 3 outlines the research methodology, and Section 4 presents the results and discussion. Section 5 concludes the study and outlines future research directions.

2. Literature Review

2.1. Building Cracks

Buildings regularly subject to minor movements due to factors such as foundation settlement, ground condition changes, and alterations in the building fabric [16]. While many of these movements are negligible, if the building is unable to accommodate them, cracking may occur [17]. These cracks, aside from being visually unappealing and concerning for occupants, may escalate into serious structural issues if not identified and addressed in time [18]. Surface cracks on building facade can be attributed to factors such as building age, material degradation, environmental conditions, and construction quality [19]. Cracks in exterior walls allow rainwater to seep into the building, causing corrosion of internal steel reinforcements. Over time, this can compromise the waterproofing membrane and affect the structural performance of the building [20]. The risk of seepage and corrosion increases with building age, highlighting the significance of early detection and rectification through effective defect analysis [19].

High-profile structural failures in Australia such as the evacuation of Sydney’s Opal Tower in December 2018 and Mascot Tower in June 2019 due to the appearance of significant cracks, have drawn global attention [21]. These incidents have raised serious concerns about construction quality, structural integrity, effective maintenance and regulatory oversight in high-rise residential buildings in Australia [22]. These building crack incidents triggered investigations and legislative reforms related to serious defects, particularly in multi-storey apartment buildings in New South Wales (NSW), Australia. According to the [23], 53% of buildings reported serious defects in 2023, a significant increase from 39% in 2021. To address such frequent problems, legislative reforms such as the Residential Apartment

Buildings (RAB) Act 2020 and the Design and Building Practitioners (DBP) Act 2020 were introduced by the NSW Government.

Traditionally, building wall cracks are detected through manual visual inspections by trained, professionally qualified personnel. While this approach provides direct observations, it is highly subjective, labour-intensive, and error-prone, especially when inspecting hard-to-reach areas or subtle micro-cracks [24]. Although tools such as thermal imaging cameras, borescopes, and drones have been introduced to improve accuracy, these methods still rely heavily on the inspector's expertise [25]. Inspectors observe the length, width, and direction of cracks and track this information over a period to assess the condition of structures [26]. However, the manual visual-inspection method cannot completely satisfy the requirements of modern high-rise buildings owing to limitations in terms of access, safety, accuracy, and efficiency. Thus, reliable crack detection methods are urgently required to satisfy the practical requirements [25]. Over the years, detection methods have evolved significantly, moving from manual inspections to advanced AI approaches using computer vision and deep learning to analyse images of building cracks [2,9].

2.2. Role of Computer Vision in Image-Based Crack Detection

There has been a growing adoption of image-based techniques leveraging digital technologies for the detection of cracks in building walls. Various semi-automated crack detection methods have been explored to capture images, including the use of stereoscopic cameras, terrestrial laser scanners, and unmanned aerial vehicles (drones) [27]. These innovations have significantly advanced computer vision-based building inspections based on image processing and analysis [11,28]. To classify cracks effectively during exterior wall inspections, computer vision systems frequently classify them based on visual characteristics, such as shape, orientation, and patterns [1,29] as shown in Figure 1.

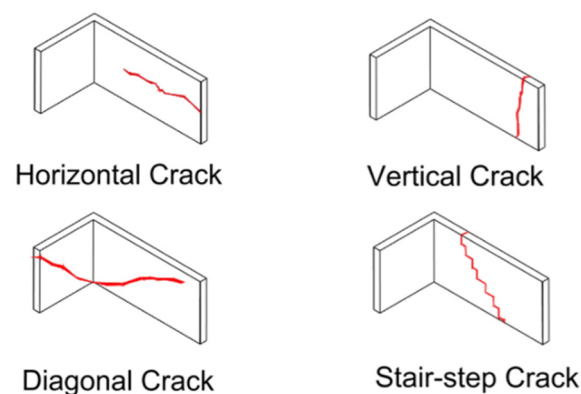


Figure 1. Crack classification for computer vision models. Source: (Lai, 2020 [29]).

With computer vision techniques, the emergence of deep learning, a subset of machine learning, has further transformed image-based crack detection based on the visual characteristics. Deep learning models, particularly convolutional neural networks (CNNs) have demonstrated high accuracy in detecting cracks in building walls [1]. The process starts with collecting and labelling a large image set of cracks, which is then pre-processed to improve model performance. The CNNs learn to detect patterns such as edges and textures that indicate cracks. Once trained, the model can detect, classify, and locate cracks in new images [9,25].

Training deep learning models incurs high computational costs and programming complexities. Their effectiveness depends on access to large and diverse datasets, which are often difficult to collect and analyse [30]. Images with complex and ambiguous crack patterns captured under inconsistent lighting, varying surface textures, and material variability

further complicate accurate crack detection in real-world conditions [25,31]. Although visual data plays a crucial role in detecting building cracks, on its own, image-based analysis is not sufficient to analyse building cracks. To fully understand the underlying causes of building defects and recommend effective remediation strategies, text information should be combined with visual inputs [32]. Multimodal AI models have been designed to process and interpret data from multiple sources, such as images and descriptive text, simultaneously [33].

2.3. Multimodal AI Models and Vision–Language Models

Multimodal AI models are inspired by the way humans naturally perceive the world aiming to process and integrate data from multiple sources such as text, images, audio, and video for a more comprehensive understanding [33]. The capabilities of such models have demonstrated significant success across diverse fields such as healthcare diagnostics, robotics navigation, and environmental monitoring [34,35]. The advancement of Generative AI driven by LLMs and Generative Pre-trained Transformer (GPT) model architecture has significantly accelerated progress in multimodal AI. These innovations laid the groundwork for conversational AI platforms such as ChatGPT [36], which provide more intelligent interactions with multimodal data. Although initially designed for text, GPT model architecture has proven highly effective across other modalities due to their self-attention mechanisms and ability to model long-range dependencies [34].

The success of computer vision and natural language processing models has prompted researchers to combine these techniques in developing vision–language models (VLMs) [37]. VLMs are capable of jointly processing visual data (images, videos) and language data (text, voice) [34]. By integrating visual and language inputs, these models can produce more effective and accurate results in image captioning, visual question answering, and text-to-image generation [33,35]. To train these models efficiently, researchers use vision–language pre-training, where large datasets of image–text or video–text pairs help the model learn shared representations of visual and textual information [38]. As ref. [39] explain, vision is the main input in these models and is often combined with text or audio to create a more complete contextual understanding. To process visual data, VLMs often use CNNs or Vision Transformers (ViTs). For processing text, they rely on language models such as Bidirectional Encoder Representations from Transformers (BERT) [40] or GPT. Transformer-based models (GPTs) are especially popular in multimodal AI because they are scalable, flexible, and can handle different types of data effectively [35].

Several well-known models demonstrate how vision and language work together. For example, Contrastive Language–Image Pretraining (CLIP) learns to align images and their corresponding text within a shared feature space, enabling the model to understand the relationship between visual content and language [35]. ViTs use attention mechanisms to capture global patterns in images, allowing more effective alignment between visual and textual information than CNNs. As a result, ViTs are now widely used as the visual backbone in modern multimodal models such as CLIP, offering enhanced generalisation and interpretability [37].

VLMs prove that it is feasible to build strong multimodal models without large-scale training [39]. The developments in multimodal vision–language models have shown notable progress in aligning visual and textual representations. For example, LLaVA-1.5 [41] and Qwen-VL [42] have demonstrated significant advances in multimodal alignment and large-scale visual–language training. Nevertheless, at the time of their release, these frameworks and associated checkpoints were still undergoing active development, making them less stable and therefore less suitable for integration into a focused, small-scale proof-of-concept study. MiniGPT-v2 [43] integrates a vision transformer with the LLaMA 2 Chat

7B language model (LLM from Meta AI), through a lightweight connector. The input image is first processed by the ViT to extract visual features, which are then mapped into the language model's input space using the connector. These adapted embeddings, along with text prompts, are then processed by the LLaMA 2 model to generate outputs. MiniGPT-v2's well-documented implementation enables stable fine-tuning and effective visual–textual reasoning using modest GPU resources. This makes MiniGPT-v2 faster and easier to train while remaining capable of generating detailed image descriptions [44].

Despite these advances, recent studies have highlighted that developing appropriate evaluation metrics for multimodal AI models is an ongoing research challenge. As ref. [45] note, the current evaluation metrics focus on recognition accuracy and text similarity, overlooking measures of reasoning quality and interpretive accuracy. Similarly, ref. [46] emphasise the importance of complementing text-based evaluation metrics with expert evaluation to assess the factual and contextual validity of VLM outputs.

2.4. Research Gaps

Despite significant advancements in computer vision and deep learning for crack detection, the existing models focus primarily on visual data. Image-based computer vision methods cannot interpret the underlying causes of cracks or suggest rectification strategies, as visual information alone is insufficient for comprehensive crack analysis. While multimodal AI models, such as VLMs, have demonstrated success in domains such as healthcare and robotics, their application in the built environment, specifically in defect diagnosis of buildings, remains limited. Additionally, many VLMs require extensive datasets and computational resources, causing practical challenges for domain-specific implementation in the built environment. Although lightweight models such as MiniGPT-v2 show promise, their application to specialised domains like building crack analysis remains limited. Existing studies have not yet examined the potential and the limitations of such trained models in identifying cracks, interpreting possible causes, or suggesting rectification strategies. Furthermore, limited research exists on how building inspection professionals perceive such models.

3. Methodology

This study employs the DSR framework as the underlying methodology, as it provides a structured approach to solving practical problems through the development of technological artefacts [15]. The DSR process involves five key stages: identifying the problem, defining the objectives for a solution, designing and developing the artefact, demonstrating its functionality, and evaluating its performance to inform further refinement. This process was mapped with the research process of the study.

In this study, the problem was identified through a literature review. This revealed that while advances in computer vision have significantly improved automated building crack detection, both visual and textual information are needed to analyse cracks. The aim was to explore the potential and limitations of VLMs for analysing building cracks. During the design and development stage, a VLM proof-of-concept model named CracksGPT, was built on a fine-tuned MiniGPT-v2 model. It was trained on a custom crack image dataset with text descriptions detailing visual features, possible causes, and rectification options. The functionality of CracksGPT was demonstrated with its ability to process multimodalities of vision (images) and language (text) in the crack analysis context. Overview of the research process is presented in Figure 2.

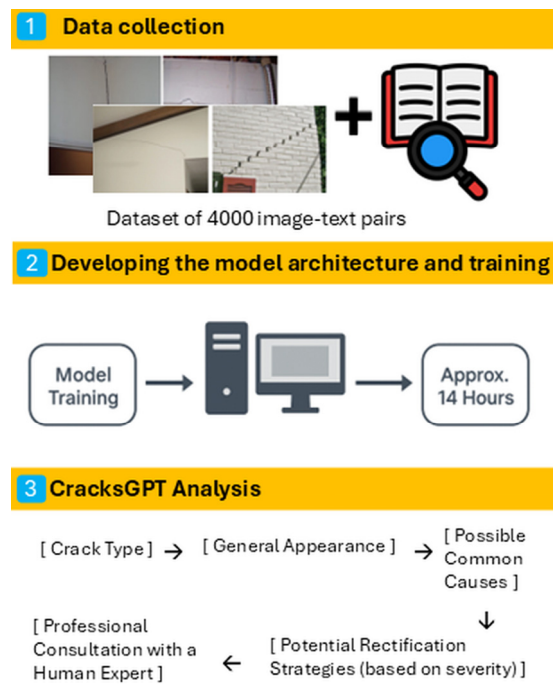


Figure 2. Overview of the research process.

Followed by the data collection and model development, the CracksGPT was tested to demonstrate the functionality. The model was evaluated using ROUGE metric for language generation quality, followed by expert feedback for further refinement. The following sections describe the research process in detail.


3.1. Data Collection

Building wall crack images were collected using published online datasets from previous research [9,47,48]. To ensure representativeness, images were collected to cover diverse crack types (vertical, horizontal, diagonal, and stair-stepped), severity levels, and environmental contexts varying in lighting, materials, and surface finishes. The dataset also contained complex crack patterns, including intersecting, branching, and multi-directional cracks. This visual and contextual diversity provided the foundation for the proof-of-concept model. The corresponding text descriptions related to the nature of the crack, potential common underlying causes and rectification strategies based on severity were collected based on literature [49,50]. From each crack category, 400 images were selected and labelled based on their appearance (vertical, horizontal, diagonal and stair-stepped).

To enhance the scale and diversity of the training data, prompt engineering techniques were used during the training process by systematically combining key attributes (“crack type”, “general appearance”, “possible common causes”, “potential rectification strategies based on severity” and “professional consultation with a human expert”), resulting in a dataset of 4000 image–text pairs. This was mainly achieved using synonym substitution and rewording strategies that preserved the original meaning.

The model was fine-tuned from MiniGPT-v2, which jointly processes image embeddings and textual prompts to generate responses and learns visual–text relationships during training. To demonstrate an example of the training process, a “horizontal crack” type was selected. First, all the images collected under this category were labelled as “horizontal crack”. Then, the text descriptions for training were prepared as per Table 1. The attributes are “crack type”, “general appearance”, “possible common causes”, “potential rectification strategies based on severity” and “professional consultation with a human expert”.

Table 1. Example of the model training process.

Attribute	Text Description Example Used for Training
Crack type	
General appearance	Horizontal cracks in walls are generally aligned parallel to the ground and appear along the length of the wall. They might range up to about 5–10 degrees. These cracks are typically observed at points where the lateral pressure is most effective.
Possible common causes	<p>Check whether the wall is a foundation wall, a basement wall, an internal partition wall or a structural wall.</p> <p>Horizontal cracks are often caused by lateral pressure exerted by the surrounding soil, particularly due to changes in soil moisture content. As the soil expands and contracts, it can exert lateral pressure on the walls, leading to cracking. Another reason would be hydrostatic pressure. Water accumulating in the soil around a building’s foundation can create hydrostatic pressure, pushing against the walls and leading to cracking. Inadequate reinforcement, poor material quality, or improper curing during construction can lead to horizontal cracks. If upper levels are added or modified without adequately reinforcing the lower structures, it can exceed the design capacity of the walls or foundation and cause horizontal cracks. Uneven settling or heaving of the soil, often caused by frost or expansive soil in cold climate regions, can lead to horizontal cracking.</p>
Potential rectification strategies based on severity	<p>To determine the strategies for rectifying the cracks, the severity of the crack needs to be identified. To define the severity of cracks, refer to the guidelines provided by the Building Research Establishment (BRE) Digest 251 (2014) [51].</p> <p>0—Hairline cracks: Less than 0.1 mm in width. 1—Fine cracks: Up to 1 mm in width. 2—Cracks easily filled: Up to 5 mm in width. 3—Cracks that require opening: Widths of 5–15 mm. 4—Extensive damage: Widths of 15–25 mm. 5—Structural damage: Widths greater than 25 mm.</p> <p>In general, categories 0, 1, and 2 with crack widths up to 5 mm can be regarded as minor cracks and ‘aesthetic’ issues that require only redecoration. Categories 3 and 4 can generally be regarded as moderate cracks causing ‘serviceability’ issues, which affect the tightness of the building and the operation of doors and windows. Category 5 presents serious cracks with ‘stability’ issues and is likely to require structural intervention.</p> <p>For minor cracks, epoxy injections can be used to fill the cracks, sealing them and preventing water intrusion. Monitor the cracks for changes and further development over time. For moderate to serious cracks, professional consultation is needed for rectification purposes.</p>
Professional consultation with a human expert	CracksGPT is an AI tool and cannot provide context specific advice. Consulting with a professional (building inspector, building surveyor, general contractor, specialised crack rectification contractor, structural engineer), is crucial to diagnose the exact causes of cracks and select the most appropriate rectification techniques.

3.2. Model Development and Evaluation

The model (CracksGPT) was developed using a fine-tuned version of MiniGPT-v2 [43] selected for its unified interface and robust capability in handling diverse vision–language tasks. The model was trained to analyse building wall cracks from uploaded images using predefined prompt questions. These prompt questions were designed to reflect the typical inquiries a user might have, enabling the model to generate relevant answers regarding the crack’s characteristics, underlying causes, and potential remedies as explained below.

- Type of the crack based on the visual appearance and the appearance and location of the crack in general.

<What is this crack>, <Tell me about this crack>, <Tell me about the image>, < What is the type of this crack>

- Potential causes behind the occurrence based on the type of crack.
<What can be the causes of this crack>, <What are the potential causes>, <What caused this crack>, <How did this crack occur>
- Potential solutions and contacting a human expert based on the recommendations
<How to fix this crack>, <How to rectify this>, <How to repair this crack>, <What are the potential solutions>

The model's pipeline is shown in Figure 3. This pipeline begins by processing the input image through a patch-based image encoder, whose outputs are mapped into the Llama 2 embedding space. These visual features are then combined with a textual query and passed to the language model to generate an answer. During the training process, only the language model parameters were fine-tuned, using the Low-Rank Adaptation of Large Language Models (LoRA) method [52], to adapt the system for crack analysis.

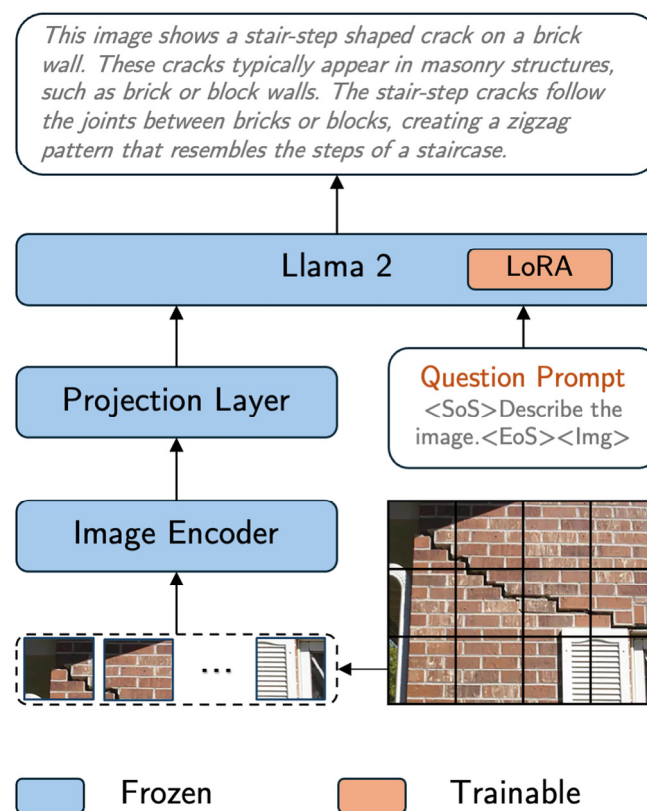


Figure 3. CracksGPT Pipeline.

Given an input image and a question prompt, the model generates a text-based answer derived from the image. The image is first divided into patches, which are processed by the image encoder to produce image embeddings. These embeddings are projected into the Llama 2 embedding space through a projection layer. The projection layer output, combined with the question prompt, is then passed to Llama 2, which produces the final answer. Since MiniGPT-v2 already aligns the image encoder and Llama 2, only the Llama 2 language model was fine-tuned using the LoRA method. In the diagram, "blue colour" indicates unchanged model parameters, while "orange colour" indicates parameters that are updated. The tokens <SoS>, <EoS>, and denote the start of sentence, end of sentence, and image token, respectively.

Model training was conducted on a Linux-based multi-GPU workstation equipped with four NVIDIA RTX 3090 GPUs (24 GB VRAM each) and 64 GB of RAM, with a total training time of approximately 14 h. The model was configured with an input image size

of 448×448 pixels and a maximum text length of 1024 tokens. Training used the AdamW optimiser with linear warm-up (500 steps, learning rate 1×10^{-6}) followed by cosine decay, starting at 1×10^{-5} with a minimum of 1×10^{-6} , and weight decay of 0.05. A per-GPU batch size of 2 was employed. To reduce computing cost, LoRA was applied with a rank of 64 and a scaling factor of 16. In this study, all experiments were conducted under a controlled and consistent environment using a single hardware configuration to ensure reproducibility and to isolate the effects of fine-tuning from hardware variability. The user interface of the trained and developed model is presented in Figure 4.

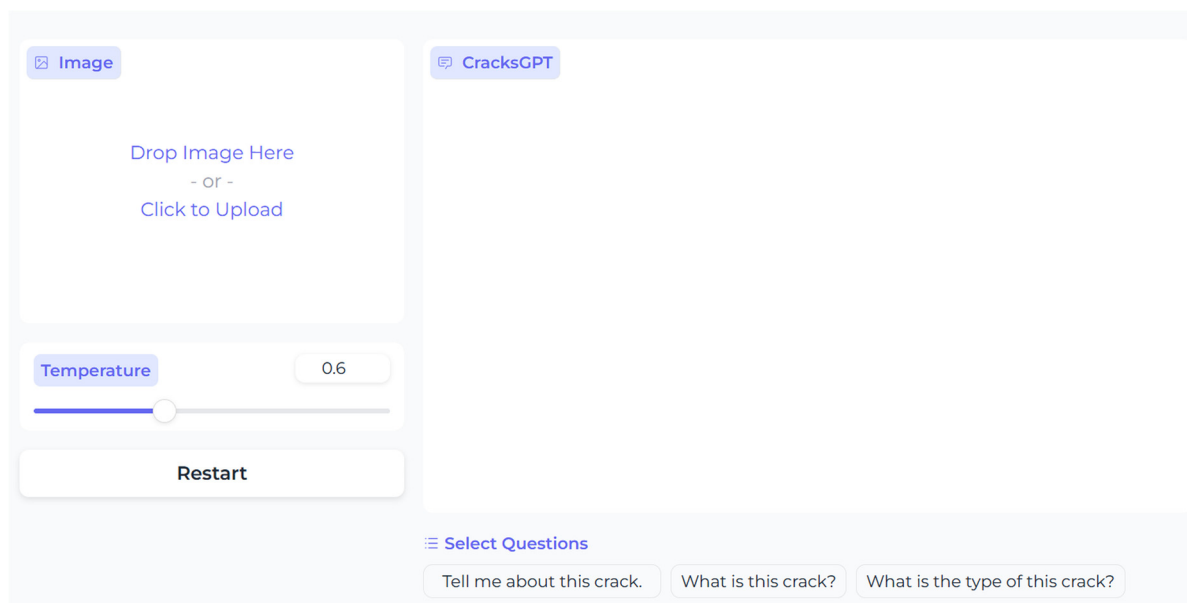


Figure 4. User interface of CracksGPT.

For evaluation, the trained model was tested using real-world images of building wall cracks representing vertical, horizontal, diagonal, and stair-stepped types captured from a building site in Sydney, Australia.

4. Results and Discussion

The initial performance assessment of CracksGPT was conducted internally by the research team using the test images, focusing on evaluating model's potential of the workflow, usability, and reasoning capabilities. Figure 5 illustrates part of a sample interaction with CracksGPT, using a horizontal wall crack.

When a building wall crack image is uploaded, CracksGPT prompts the user to ask a question to identify the crack type based on the shape. It then offers insights into the typical appearance and common locations of that crack type on a wall. The user is subsequently prompted to explore potential causes, supported by a list of frequently associated contributing factors. In the next step, the user is prompted to assess the severity of the crack, upon which the model recommends potential rectification strategies. Through the internal assessment conducted by the research team, it was found that the model successfully identified the crack types from uploaded images, most commonly stair-stepped cracks. The identification of underlying causes and rectification strategies for stair-stepped cracks is aligned with descriptions in existing literature. On average, the model generated an analysis within 10 s.

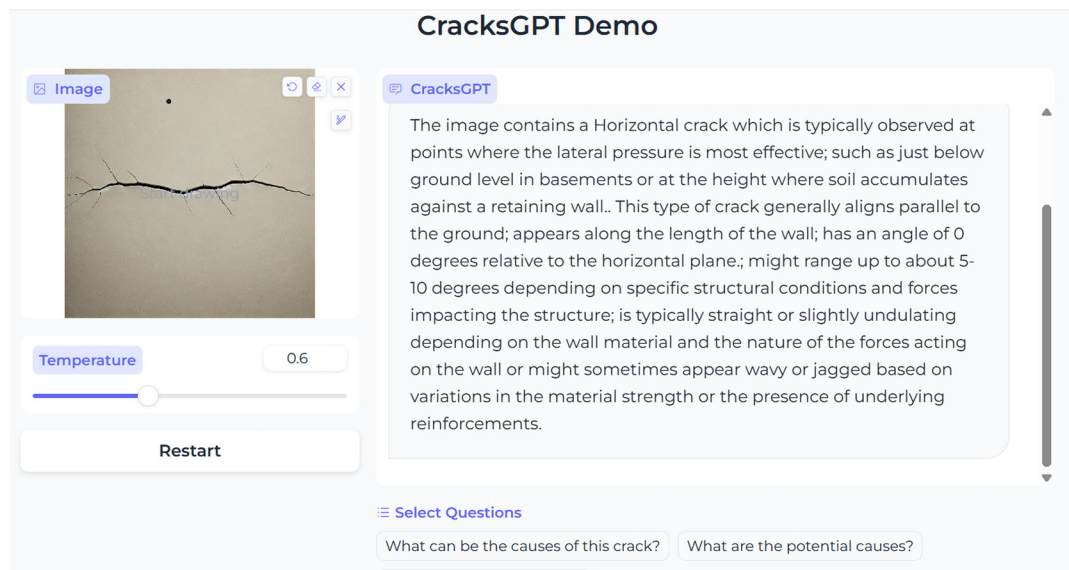


Figure 5. A sample interaction with CracksGPT.

4.1. Internal Trials and Evaluation

The existing evaluation metrics for VLMs measure recognition and text similarity but not reasoning or interpretive accuracy [45,46]. The research team employed the ROUGE (Recall-Oriented Understudy for Gisting Evaluation) metric to assess the performance of CracksGPT model. ROUGE is a widely adopted metric to evaluate generative AI tasks such as text summarisation, content generation, and question answering [53]. ROUGE evaluates the quality of AI-generated text by measuring its similarity to reference (human-written) outputs, focusing on overlapping word sequences and word pairs. Higher ROUGE scores indicate closer alignment with human-written responses, reflecting improved fluency and accuracy. The trials compared the performance of CracksGPT with its base model, MiniGPT-v2 using ROUGE. In this study, the human-written outputs were developed by the research team based on relevant literature. ROUGE was applied to assess the extent to which CracksGPT's generated descriptions, such as crack type classification, identification of potential causes, and suggestion of rectification strategies, corresponded with the human-written outputs. The evaluations were conducted at two levels: (1) per-response, comparing individual model outputs with their corresponding human-written outputs; and (2) full-conversation on a crack analysis, comparing the entire AI-generated dialogue to a complete human-written version. The results of these evaluations are summarised in Table 2.

ROUGE-1 measures the overlap of single words (unigrams) between the generated and human-written texts, providing a basic indication of content similarity. ROUGE-2 evaluates the overlap of word pairs (bigrams), reflecting the fluency and logical ordering of the generated content. ROUGE-L focuses on the longest common subsequence, capturing the sentence-level structure and coherence. Precision indicates the proportion of the AI-generated content that is relevant, while recall reflects the proportion of relevant content from the human-written text that the AI successfully captured. The F-measure (F1 score) represents the harmonic mean of precision and recall, offering a balanced assessment of overall performance.

As shown in Table 2, CracksGPT consistently outperforms the base model, MiniGPT-v2, across both evaluation levels. In the per-response comparison, CracksGPT shows improvements of +24.70% in precision, +27.54% in recall, and 29.81% in F-measure over MiniGPT-v2 (Row 3 vs. Row 6, ROUGE-L). Similarly, in the full conversation comparison, CracksGPT achieves an increase in +24.72% in precision, +24.96% in recall, and +28.47% in

F-measure (Row 9 vs. Row 12, ROUGE-L). These results underscore CracksGPT's ability to generate accurate, coherent, and human-aligned responses. The ROUGE metric score indicates the similarity between the AI outputs generated and human-written (reference) outputs. Higher scores indicate higher similarity between the summary and the reference. However, several limitations emerged during testing.

Table 2. Evaluation of the CracksGPT model using the ROUGE metric.

No.	Model	Metrics	ROUGE Score			
			Rouge	Precision	Recall	F1 Score
1	MiniGPT-v2	ROUGE per response	Rouge-1	0.1786	0.5381	0.2351
2			Rouge-2	0.0350	0.1163	0.0486
3			Rouge-L	0.0978	0.2852	0.1235
4	CracksGPT	ROUGE per response	Rouge-1	0.4347	0.7081	0.5315
5			Rouge-2	0.2922	0.4743	0.3570
6			Rouge-L	0.3448	0.5606	0.4216
7	MiniGPT-v2	ROUGE per conversation	Rouge-1	0.1877	0.7075	0.2967
8			Rouge-2	0.0549	0.2069	0.0868
9			Rouge-L	0.0822	0.3099	0.1299
10	CracksGPT	ROUGE per conversation	Rouge-1	0.4709	0.7999	0.5928
11			Rouge-2	0.3257	0.5532	0.4100
12			Rouge-L	0.3294	0.5595	0.4146

As with many LLMs, instances of AI hallucinations were observed [44], where the model sometimes confidently generated inaccurate or unsupported information. These were particularly evident in the suggested causes and rectification strategies. Such cases emphasise the importance of encouraging users to critically evaluate AI-generated content through independent research and consultation with qualified professionals. Due to these limitations, the model was designed to consistently prompt users to seek expert assessment because building cracks can appear complex, even to the human eye. The current version of the model faces some challenges in detecting cracks in complex scenarios, such as those with subtle angles that make it difficult to categorise them accurately. It also struggles to analyse images containing multiple cracks simultaneously, which can lead to misinterpretations. Furthermore, the model lacks the ability to recognise the underlying building material or wall surface type, which can limit the contextual accuracy and precision of the diagnostic recommendations. These challenges are largely attributed to dataset limitations and inherent constraints in computer vision accuracy [2,9]. The model's recommendations serve as general guidance. Therefore, the internal trials demonstrated the model's potential value in educational and exploratory contexts.

4.2. Expert Evaluation Results Analysis

In addition to the internal performance assessment, five in-person interviews were conducted to obtain expert feedback on CracksGPT's usability and real-world applicability. The approval to conduct interviews with five building inspectors possessing expertise in crack diagnosis and digital technologies was granted by the University Human Research Ethics Committee (reference: ETH25-10718). The experts were provided with access to the CracksGPT model and a test image dataset to facilitate their evaluation. They also had

the option to upload their own crack images to further assess the model's performance on real-world samples. The profile of the respondents is presented in Table 3.

Table 3. Respondents' profile.

Expert ID	Designation	Experience
E1	Managing Director and a licensed Building Inspector of a Sydney based home inspection company	Over 30 years
E2	Lead Building Inspector of a Sydney based property inspection company	Over 23 years
E3	Lead Building Inspector Sydney based home inspection company	Over 20 years
E4	Licensed Building Inspector of a Sydney based property inspection company	Over 17 years
E5	Licensed Property Inspection Specialist of a Sydney based property inspection company	Over 15 years

The qualitative data collected from the interviews were structured around questions related to the user interface, detection accuracy, reasoning and logic, and real-world applicability. These were analysed using content analysis following the approach of ref. [54]. This analysis enabled the extraction of experts' insights into CracksGPT's capabilities and its practical applicability, as discussed below.

(1) User interface

All five experts found the CracksGPT interface to be user-friendly. The model's structured interaction was considered particularly beneficial for users who may lack experience in performing crack analysis. As noted by E3, *"I like how it mimics the way I explain issues to clients during an inspection—question by question, step by step"*. E1 and E4 emphasised the importance of flexibility in the interface, suggesting that users with diagnostic experience should have the option to engage with open-ended queries in addition to the model's prompted questions. E2 and E5 raised concerns about the workflow surrounding the crack severity assessment step and recommended streamlining this component with open-ended prompts by the user to improve the overall user experience.

(2) Detection accuracy

Experts generally agreed that CracksGPT was reliable in identifying common crack types, including vertical, horizontal, and stair-step cracks. However, they noted limitations in distinguishing diagonal cracks from vertical ones, particularly when the angle was not clear. As E1 explained, *"CracksGPT detected most cases well, mostly the stair-stepped cracks but with diagonal cracks especially the ones with a small angle, it generated incorrect classifications"*. All experts also highlighted a significant limitation when analysing images containing multiple cracks. They stated that in such instances, the model generated inaccurate and fabricated responses. E3 and E4 further pointed out that the model has not been trained to detect complex crack patterns, such as those resembling spider webs, and currently relies on the four predefined crack types. However, they acknowledged that cracks with complex patterns are challenging even for experienced inspectors and are prone to misclassification by an AI model. E2 and E5 recommended adding a user prompt that asks users to upload images containing only one crack, corresponding to one of the four types currently detected by CracksGPT.

(3) Reasoning and logic

Most experts (E1, E2, E4) acknowledged that the model provided credible general causes and rectification strategies for most cases. Despite occasional misclassifications, the model's logic was generally viewed as reasonable and consistent with standard inspection practices. As E4 noted, *"The logic behind its classification made sense most of the time. It wasn't just random guesses."* E3 and E5 pointed out that some of the model's recommendations lacked the nuance needed for more complex scenarios. They highlighted the model's inability for users to provide details related to key contextual factors such as building materials, age of the structure, structural changes, environmental changes, and weather changes, all of which are critical for improving contextual accuracy of the diagnosis. They explained that this limitation could result in AI generating inaccurate and fabricated information in complex cases. All experts appreciated the inclusion of a disclaimer that clearly communicated the model's limitations. They stressed the importance of reminding users that CracksGPT, an AI model, is not capable of offering context-specific advice. While the model demonstrates the potential of AI to support crack analysis, users should be encouraged to consult qualified professionals for accurate assessment and tailored rectification strategies.

(4) Real-world applicability

CracksGPT was recognised by all the experts as a promising decision-support tool with practical value in both professional and educational contexts, with more future developments. While not intended to replace human expertise, E2, E4 and E5 emphasised CracksGPT's potential for preliminary screening before a detailed manual review. E2 explained, *"CracksGPT could serve as an initial aid to help residential property owners or less-experienced personnel identify potential defects before a professional inspection"*. E1 and E3 highlighted CracksGPT's usefulness in educational and training environments, particularly for junior inspectors or students learning the fundamentals of building defect diagnosis. E5 noted, *"There is a potential application of CracksGPT as an initial screening tool in regional and remote areas, where access to licensed building inspectors may be limited"*. All respondents agreed that AI-generated outputs should not replace professional judgment in real-world settings. They emphasised the importance of positioning CracksGPT strictly as an assistive tool rather than a definitive diagnostic system.

The analysis of expert insights played a critical role in informing the refinement of CracksGPT for future developments. While there were some discrepancies in experts' opinions, they were minor, natural, and constructive rather than contradictory because experts bring diverse experience and interpretive lenses.

4.3. Potential and the Limitations

By moving beyond manual inspection methods, CracksGPT highlights the potential of multimodal AI for integrating vision–language reasoning into building inspections. The findings demonstrate CracksGPT's potential to enhance building crack awareness and support early identification of structural issues through AI-assisted analysis that extends beyond image-based deep learning models such as YOLO and Mask R-CNN. CracksGPT could serve as an initial screening tool for homeowners, helping them recognise visible building cracks before professional assessment, and as an educational aid for junior building inspectors and students, fostering understanding of crack types, causes, and repair strategies.

Several limitations were identified during the development and evaluation of CracksGPT. The model is sensitive to input image quality. Variations in lighting, resolution, and noise can substantially affect its ability to accurately interpret crack features. The model exhibits limitations in accurately distinguishing visually similar crack orientations and

becomes less reliable when analysing images that feature multiple cracks, intersecting geometries, or intricate morphological patterns. These performance gaps largely reflect limitations in the underlying dataset as well as the constraints inherent to current computer vision techniques. Because CracksGPT was not trained on contextual structural and environmental metadata, it is unable to incorporate important contextual cues such as building materials, surface conditions, construction methods, structural loading, and weather-related effects. The model's susceptibility to generating inaccurate information in complex diagnostic scenarios, results from AI hallucinations inherent in LLMs.

5. Conclusions and Future Directions

This study introduced CracksGPT, a novel vision–language AI model, as a proof-of-concept. The model was designed to integrate visual crack detection using images with text-based natural language guidance. Using a DSR approach, CracksGPT was built on a fine-tuned MiniGPT-v2 model. When provided with an image of a wall crack, the trained model classifies cracks into vertical, horizontal, diagonal, and stair-step categories and interprets possible underlying causes with potential rectification strategies. The model significantly outperformed its base model, MiniGPT-v2, as demonstrated by substantial improvements in ROUGE scores for language generation quality.

The study contributes both methodologically and practically to the growing body of knowledge on AI-assisted building inspections. The study also draws attention to the AI models' limitations in handling complex or highly context-specific scenarios. AI models in this context must be deployed with transparency and supported by clear communication of its constraints. This reinforces the need for human expert verification to ensure that outputs are interpreted appropriately. This study also underscores the importance of positioning AI models strictly as supportive tools rather than definitive diagnostic systems. Therefore, the broader implication is that this study highlights the boundaries between automation and human expertise and calls for the ethical and responsible deployment of AI models within the built environment.

Future research should prioritise advancing CracksGPT's visual detection capabilities through a comprehensive model training process supported by a high-quality, large and diverse image dataset including different crack morphologies captured under various lighting, environmental and material conditions. To further expand dataset diversity, synthetic data with morphology-preserving augmentation, such as mild geometric jitter, photometric variation, could also be used. Validating the model's robustness across different hardware platforms represents an important step toward reliable real-world deployment. Simultaneously, incorporating contextual parameters specific to the building and the surroundings will be essential to enhance natural language diagnostic logic. Additionally, future iterations could expand the model's diagnostic scope to cover other types of building defects beyond cracks through more comprehensive field testing using more advanced models such as LLaVa-NeXT and Qwen-VL-Chat. Integrating expert-in-the-loop feedback mechanisms could enhance the model's diagnostic accuracy and foster greater user trust. Cross-domain validation and large-scale field testing with building inspectors, builders, building managers, facility managers, valuation professionals, and insurance companies will be essential before the integration of CracksGPT into real-world building inspection and maintenance workflows.

Author Contributions: Conceptualization, methodology, analysis, writing—original draft preparation, B.E.; Data curation, software, validation, V.T.; Resources, project administration, writing—review and editing, visualisation, J.K.-W.W.; S.W. and S.H.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The authors would like to acknowledge the internal funding received through the 2024 Faculty Research Strategic Funding Scheme and the 2025 Faculty Research Support Scheme of University of Technology Sydney.

Data Availability Statement: Some or all data and code that support the findings of this study are available from the corresponding author upon request. This is due to the planned future studies by the research team using the dataset and the code.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cheng, M.; Ma, Z.; Xie, J.; Li, Q. Specific defect detection for efficient building maintenance. *J. Build. Eng.* **2025**, *112*, 113710. [[CrossRef](#)]
2. Pandey, V.; Mishra, S.S. A review of image-based deep learning methods for crack detection. *Multimed. Tools Appl.* **2025**, *84*, 35469–35511. [[CrossRef](#)]
3. Mishra, M.; Lourenço, P.B.; Ramana, G.V. Structural health monitoring of civil engineering structures by using the internet of things: A review. *J. Build. Eng.* **2022**, *48*, 103954. [[CrossRef](#)]
4. Hauashdh, A.; Jailani, J.; Abdul Rahman, I.; Al-Fadhali, N. Factors affecting the number of building defects and the approaches to reduce their negative impacts in Malaysian public universities' buildings. *J. Facil. Manag.* **2022**, *20*, 145–171. [[CrossRef](#)]
5. Park, M.; Kwon, N.; Lee, J.; Lee, S.; Ahn, Y. Probabilistic maintenance cost analysis for aged multi-family housing. *Sustainability* **2019**, *11*, 1843. [[CrossRef](#)]
6. Douglas, J.; Ransom, B. *Understanding Building Failures*; Routledge: London, UK, 2007. [[CrossRef](#)]
7. Şimşek, B. Investigation of self-healing ability of hydroxyapatite blended cement paste modified with graphene oxide and silver nanoparticles. *Constr. Build. Mater.* **2022**, *320*, 126250. [[CrossRef](#)]
8. Kong, Q.; Allen, R.M.; Kohler, M.D.; Heaton, T.H.; Bunn, J. Structural health monitoring of buildings using smartphone sensors. *Seismol. Res. Lett.* **2018**, *89*, 594–602. [[CrossRef](#)]
9. Ekanayake, B. A deep learning-based building defects detection tool for sustainability monitoring. In Proceedings of the 10th World Construction Symposium 2022, Colombo, Sri Lanka, 24–26 June 2022; Ceylon Institute of Builders: Colombo, Sri Lanka, 2022; pp. 1–8. Available online: https://ciobwcs.com/downloads/WCS2022_Full_Proceedings.pdf (accessed on 1 August 2025).
10. Kung, R.Y.; Pan, N.H.; Wang, C.C.; Lee, P.C. Application of deep learning and unmanned aerial vehicle on building maintenance. *Adv. Civ. Eng.* **2021**, *2021*, 8836451. [[CrossRef](#)]
11. Dizaji, M.S.; Harris, D.K. 3D InspectionNet: A deep 3D convolutional neural networks-based approach for 3D defect detection on concrete columns. In Proceedings of the Nondestructive Characterization and Monitoring of Advanced Materials, Aerospace, Civil Infrastructure, and Transportation XIII, Denver, CO, USA, 4–7 March 2019; Volume 10971, p. 109710E. [[CrossRef](#)]
12. Munawar, H.S.; Ullah, F.; Heravi, A.; Thaheem, M.J.; Maqsoom, A. Inspecting buildings using drones and computer vision: A machine learning approach to detect cracks and damages. *Drones* **2022**, *6*, 5. [[CrossRef](#)]
13. Wang, S.; Cheng, N.; Hu, Y. Comprehensive environmental monitoring system for industrial and mining enterprises using multimodal deep learning and CLIP model. *IEEE Access* **2025**, *13*, 19964–19978. [[CrossRef](#)]
14. Xuanfan, N.; Piji, L. A systematic evaluation of large language models for natural. In Proceedings of the 22nd Chinese National Conference on Computational Linguistics, Harbin, China, 3–5 August 2023; Volume 2, pp. 40–56. Available online: <https://aclanthology.org/2023.ccl-2.4/> (accessed on 1 August 2025).
15. Hevner, A.R. A three-cycle view of design science research. *Scand. J. Inf. Syst.* **2007**, *19*, 4.
16. Olurotimi, O.J.; Yetunde, O.H.; Akah, U. Assessment of the determinants of wall cracks in buildings: Investigating the consequences and remedial measures for resilience and sustainable development. *Int. J. Adv. Educ. Manag. Sci. Technol.* **2023**, *6*, 121–132.
17. Kacker, R.; Singh, S.K.; Kasar, A.A. Understanding and addressing multi-faceted failures in building structures. *J. Fail. Anal. Prev.* **2024**, *24*, 1542–1558. [[CrossRef](#)]
18. Ajagbe, W.O.; Ojedele, O.S. Structural investigation into the causes of cracks in building and solutions: A case study. *Am. J. Eng. Res.* **2018**, *7*, 152–160.
19. Kang, S.; Kim, S.; Kim, S. Improvement of the defect inspection process of deteriorated buildings with scan to BIM and image-based automatic defect classification. *J. Build. Eng.* **2025**, *99*, 111601. [[CrossRef](#)]
20. Krahmalny, T.A.; Evtushenko, S.I. Typical defects and damage to the industrial buildings' facades. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *775*, 012135. [[CrossRef](#)]
21. Council on Tall Buildings and Urban Habitat. Report on Sydney's Opal Tower Blames Structural Design, Construction Flaws for Cracks. *Council on Tall Buildings and Urban Habitat*. Available online: <https://www.ctbuh.org/news/report-on-sydneys-opal-tower-blames-structural-design-construction-flaws-for-cracks> (accessed on 1 August 2025).

22. Engineering Institute of Technology. *Australia's Apartment Building Cracks Show Corner-Cutting in Civil Engineering*; Engineering Institute of Technology: Lund, Sweden, 2019. Available online: <https://www.eit.edu.au/australias-apartment-building-cracks-show-corner-cutting-in-civil-engineering/> (accessed on 1 August 2025).
23. NSW Government. Strata Defects Survey Report—November 2023. 2023. Available online: <https://www.nsw.gov.au/sites/default/files/noindex/2023-12/strata-defects-survey-report.pdf> (accessed on 1 August 2025).
24. Abdel-Qader, I.; Abudayyeh, O.; Kelly, M.E. Analysis of edge-detection techniques for crack identification in bridges. *J. Comput. Civ. Eng.* **2003**, *17*, 255–263. [CrossRef]
25. Ding, W.; Yang, H.; Yu, K.; Shu, J. Crack detection and quantification for concrete structures using UAV and transformer. *Autom. Constr.* **2023**, *152*, 104929. [CrossRef]
26. Valero, E.; Forster, A.; Bosché, F.; Hyslop, E.; Wilson, L.; Turmel, A. Automated defect detection and classification in ashlar masonry walls using machine learning. *Autom. Constr.* **2019**, *106*, 102846. [CrossRef]
27. Spencer, B.F.; Hoskere, V.; Narazaki, Y. Advances in computer vision-based civil infrastructure inspection and monitoring. *Engineering* **2019**, *5*, 199–222. [CrossRef]
28. Bhowmick, S.; Nagarajaiah, S.; Veeraraghavan, A. Vision and deep learning-based algorithms to detect and quantify cracks on concrete surfaces from UAV videos. *Sensors* **2020**, *20*, 6299. [CrossRef]
29. Lai, J.L.W. Type of Cracks on Buildings. IPM. 24 March 2020. Available online: <https://ipm.my/type-of-cracks-on-buildings/> (accessed on 1 August 2025).
30. Ai, D.; Jiang, G.; Lam, S.K.; He, P.; Li, C. Computer vision framework for crack detection of civil infrastructure—A review. *Eng. Appl. Artif. Intell.* **2022**, *117*, 105478. [CrossRef]
31. Chen, Y.; Zhu, Z.; Lin, Z.; Zhou, Y. Building surface crack detection using deep learning technology. *Buildings* **2023**, *13*, 1814. [CrossRef]
32. Wen, Y.; Chen, K. Autonomous detection and assessment of indoor building defects using multimodal learning and GPT. In Proceedings of the Construction Research Congress 2024, Des Moines, IA, USA, 20–23 March 2024; American Society of Civil Engineers: New York, NY, USA, 2024; pp. 1001–1009. [CrossRef]
33. Baltrusaitis, T.; Ahuja, C.; Morency, L. Multimodal machine learning: A survey and taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 423–443. [CrossRef]
34. Thawakar, O.C.; Shaker, A.M.; Mullappilly, S.S.; Cholakkal, H.; Anwer, R.M.; Khan, S.; Laaksonen, J.; Khan, F.S. XrayGPT: Chest radiographs summarization using large medical vision-language models. In Proceedings of the 23rd Workshop on Biomedical Natural Language Processing, Bangkok, Thailand, 16 August 2024; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024; pp. 440–448. Available online: <https://aclanthology.org/2024.bionlp-1.35/> (accessed on 1 August 2025).
35. Xu, P.; Zhu, X.; Clifton, D.A. Multimodal learning with transformers: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 12113–12132. [CrossRef]
36. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 1 August 2025).
37. Zhang, Y.; Liu, C. Vision-enhanced multi-modal learning framework for non-destructive pavement damage detection. *Autom. Constr.* **2025**, *177*, 106389. [CrossRef]
38. Chen, F.; Zhang, D.; Han, M.; Chen, X.; Shi, J.; Xu, S.; Xu, B. VLP: A survey on vision-language pre-training. *Mach. Intell. Res.* **2023**, *20*, 38–56. [CrossRef]
39. Zhu, D.; Chen, J.; Shen, X.; Li, X.; Elhoseiny, M. MiniGPT-4: Enhancing Vision-Language Understanding with Advanced Large Language Models. *arXiv* **2023**, arXiv:2304.10592. Available online: <https://arxiv.org/abs/2304.10592> (accessed on 1 August 2025).
40. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MN, USA, 2–7 June 2019; pp. 4171–4186. Available online: <https://aclanthology.org/N19-1423/> (accessed on 1 August 2025).
41. Liu, H.; Li, C.; Li, Y.; Lee, Y.J. Improved baselines with visual instruction tuning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (26296–26306), Seattle, WA, USA, 16–22 June 2024.
42. Bai, J.; Bai, S.; Yang, S.; Wang, S.; Tan, S.; Wang, P.; Lin, J.; Zhou, C.; Zhou, J. Qwen-VL: A versatile vision-language model for understanding, localization, text reading, and beyond. *arXiv* **2023**, arXiv:2308.12966.
43. Chen, J.; Zhu, D.; Shen, X.; Li, X.; Liu, Z.; Zhang, P.; Krishnamoorthi, R.; Chandra, V.; Xiong, Y.; Elhoseiny, M. MiniGPT-v2: Large language model as a unified interface for vision-language multi-task learning. *arXiv* **2023**, arXiv:2310.09478. Available online: <https://arxiv.org/abs/2310.09478> (accessed on 1 August 2025).
44. Alsabbagh, A.R.; Mansour, T.; Al-Kharabsheh, M.; Ebdah, A.S.; Al-Emaryeen, R.A.; Al-Nahas, S.; Al-Kadi, O. MiniMedGPT: Efficient large vision–language model for medical visual question answering. *Pattern Recognit. Lett.* **2025**, *189*, 8–16. [CrossRef]

45. Salin, E.; Ayache, S.; Favre, B. Towards an exhaustive evaluation of vision-language foundation models. In Proceedings of the IEEE/CVF International Conference on Computer Vision (339–352), Paris, France, 2–6 October 2023.
46. Areerob, K.; Nguyen, V.Q.; Li, X.; Inadomi, S.; Shimada, T.; Kanasaki, H.; Okatani, T. Multimodal artificial intelligence approaches using large language models for expert-level landslide image analysis. *Comput.-Aided Civ. Infrastruct. Eng.* **2025**, *40*, 2900–2921. [[CrossRef](#)]
47. Elhariri, E.; El-Bendary, N.; Taie, S.A. Historical-crack18-19: A dataset of annotated images for non-invasive surface crack detection in historical buildings. *Data Brief* **2022**, *41*, 107865. [[CrossRef](#)] [[PubMed](#)]
48. Özgenel, Ç.F. *Concrete Crack Images for Classification [Data Set]*; Mendeley Data: Online, 2019; V2. [[CrossRef](#)]
49. Ransom, W.H. *Building Failures: Diagnosis and Avoidance*; Longman: London, UK, 1987.
50. Watt, D.S. *Building Pathology: Principles and Practice*, 2nd ed.; Wiley-Blackwell: Oxford, UK, 2007.
51. Building Research Establishment (BRE). Assessing Cracks in Houses, Digest 251. 2014. Available online: <https://bregroup.com/insights/assessing-cracks-in-houses> (accessed on 5 October 2025).
52. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Chen, W. LoRA: Low-rank adaptation of large language models. In Proceedings of the International Conference on Learning Representations (ICLR), Online, 25–29 April 2022. Available online: <https://openreview.net/forum?id=nZeVKeeFYf9> (accessed on 1 August 2025).
53. Gardner, N.; Khan, H.; Hung, C.-C. Definition modeling: Literature review and dataset analysis. *Appl. Comput. Intell.* **2022**, *2*, 83–98. [[CrossRef](#)]
54. Elo, S.; Kyngäs, H. The qualitative content analysis process. *J. Adv. Nurs.* **2008**, *62*, 107–115. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.