INTEGRATING MACHINE LEARNING AND COMPUTATIONAL PHYSICS TO ASSESS CRACK PATTERN SIMILARITY IN MASONRY BUILDINGS

Krishna Ajithkumar Pillai

TU Delft Delft University of Technology

TNO innovation for life

INTEGRATING MACHINE LEARNING AND COMPUTATIONAL PHYSICS TO ASSESS CRACK PATTERN SIMILARITY IN MASONRY BUILDINGS

A thesis submitted to the Delft University of Technology in partial fulfilment of the requirements for the degree of

Master of Science in Civil Engineering

Structural Engineering Track
Specialisation: Structural Mechanics

by

Krishn Ajithkumar Pillai
(5070929)

February 2022

TU Delft Delft University of Technology

TNO innovation for life

This thesis was done at:

Applied Mechanics
Department of Materials, Mechanics, Management & Design (3MD)
Faculty of Civil Engineering & Geosciences
Delft University of Technology

and

Expertise Group for Structural Reliability
TNO

Graduation Committee:

Giorgia Giardina
Committee Chair
Department of Geoscience & Engineering
Faculty of Civil Engineering and Geosciences
Email: G.Giardina@tudelft.nl

Arthur Slobbe
Expertise Group for Structural Reliability
TNO
Email: arthur.slobbe@tno.nl

Árpád Rózsás
Expertise Group for Structural Reliability
TNO
Email: arpad.rozsas@tno.nl

Anjali Mehrotra
Department of Materials, Mechanics, Management & Design (3MD)
Faculty of Civil Engineering and Geosciences
Email: A.A.Mehrotra@tudelft.nl

Jan Rots
Department of Materials, Mechanics, Management & Design (3MD)
Faculty of Civil Engineering and Geosciences
Email: J.G.Rots@tudelft.nl

ABSTRACT

Cracks in masonry structures are a cause for concern as they signal a potential lack of functionality and/or aesthetics. It thus becomes important to identify the cause of damage in order to mitigate it and to prevent its occurrence in the future. Similarities in crack patterns may correlate to similarities in the damage cause. Currently, the assessment of similarities in crack patterns and their corresponding damage causes is done by masonry experts and structural engineers. This process is often expensive and subjective. The use of a Convolutional Neural Network (CNN) may offer an alternate robust and dependable means to automate the assessment of masonry crack patterns by processing their images.

The main research goal of this MSc thesis is to answer how accurately can the CNN – fitted to data generated from finite element models – estimate masonry crack pattern similarities. To develop a neural network that can perform such an automated assessment of masonry crack patterns with a high degree of accuracy, a large number of crack patterns with similarity ratings given by human experts are required. This data is collected in increasing complexity, first from a statistics-based approach by generating synthetic crack patterns from Markov walks. This is followed by a computational physics-based approach, such as the Finite Element Method (FEM), that generates crack patterns on 2D masonry façades subjected to differential settlements and out-of-plane loads. Finally, real-world data is also collected. This data is used to fit and test a convolutional neural network developed by Kleijn [1]. Continuing along the previous line of research done at TNO (where 12 crack patterns were chosen and developed using the statistics-based approach), this thesis focuses on developing parametric finite element models of 8 out of these 12 Pattern IDs. Additionally, real-world images are also collected from Gouda in The Netherlands. This data is then used to form crack pattern image pairs that can be assessed for their similarities by 28 raters using three similarity label categories: crack pattern similarity label, damage severity label, and the overall similarity label. Using these labels, the raters assessed 2587 image pairs generated from the statistics-based approach, 500 image pairs from the computational physics-based approach, and 50 image pairs from the combination of images from the statistics-based approach, computational physics-based approach, and the real-world cases.

An inter-rater agreement analysis is performed on the similarity assessments using the Krippendorff's alpha measure. Additionally, the agreement of each rater with a chosen standard rater is studied using the Lin's Concordance Correlation Coefficient (CCC). Using the Lin's CCC, the intra-rater agreement is also assessed for the standard rater to see how consistent a rater is with their own annotations. These labelled image pairs are then used to fit and test the regression neural network to evaluate its accuracy in predicting the similarity labels. The neural network is also fitted to and tested with various combinations of labelled data to study its generalisability.

It is found that in all three sets of data, the Krippendorff's alpha is less than 0.80 for all the labels, which indicates an insufficient agreement among the raters. It is also seen that, in general, agreement among the raters increases with their experience level, i.e. the descending order of agreement within the rater group is: industry experts, PhD students, and MSc students. Studying the Lin's CCC of each rater's performance compared to that of the standard rater helps to choose the raters who can be considered as reliable as the standard rater. Additionally, the intra-rater agree-

ment analysis of the chosen standard rater shows that the highest self-consistency (agreement) is achieved for the crack pattern similarity label, followed by the overall similarity label and finally the damage severity label, with corresponding Lin's CCC values of 0.96, 0.86 and 0.72, respectively.

The neural network is tasked to predict the similarity level in each similarity rating for each image pair in the test sample. The ground truth of this neural network is established by averaging the similarity ratings given to each image pair by multiple raters. It is found that the neural network is able to achieve a sufficiently high degree of accuracy when fitted to and tested with all the image pairs generated from the computational physics-based approach. The crack pattern similarity label, the damage severity label, and the overall similarity label achieve an accuracy of 87%, 82%, and 69%, respectively. However, the generalisability experiments on the neural network that consist of predicting the similarity of a type of crack pattern image pair that is not included in the fitting data set, show very poor performance with respect to the prediction accuracy of the similarity labels. When the neural network attempts to predict the similarity of Pattern ID or a façade geometry that it did not see in the fitting procedure, it predicts all the three labels with an accuracy that varies from 40% to 50%. Additionally, the neural network is also fitted to images generated from the computational physics-based approach and then tested with a pool of image pairs generated from the statistics-based approach, computational physics-based approach, and real-world images. The average accuracy with which the three similarity labels are predicted is even lower, lying between 25% and 40%.

This MSc thesis concludes that the neural network fitted to data generated from the computational physics-based approach and assessed by all the raters is able to predict the crack pattern similarity label, the damage severity label and the overall similarity label with sufficiently high degrees of accuracy. However, the generalisability experiments on the neural network show very poor results. This indicates that in order to achieve a greater prediction accuracy, the neural network may need to be fitted to a considerably larger sample of crack patterns that covers all of the relevant situations. Furthermore, the substantial inter-rater variability in the labelling of crack pattern image pairs suggests that even an ideal neural network architecture may not be able to overcome the inconsistencies in the fitting data.

ACKNOWLEDGEMENTS

This graduation thesis is a part of an ongoing effort at TNO to automate the assessment of crack pattern similarities in masonry structures. It has been a very rewarding opportunity to be a part of this research project. I am incredibly thankful to my chair, Dr Giorgia Giardina, and my supervisors, Dr Arthur Slobbe and Dr Árpád Rózsás, for introducing me to this project and for the infinite patience with which they have guided me. Their mentorship has helped me improve my technical proficiency, scientific writing and critical thinking; lessons that I am sure will stay with me long after my time in Delft is over. I am sincerely grateful to Dr Anjali Mehrotra and Dr Jan Rots, for being a part of my graduation committee, and for providing invaluable feedback on how to conduct the assessment of the similarity in masonry crack patterns.

I would also like to thank Kelvin Kleijn, Dr Wyke Pereboom-Huizinga and Dr Hao Wang, for the numerous conversations that gave me insight into the field of machine learning that was completely new to me. Kelvin was a huge help all throughout this thesis, from generating the data to analysing the results from the neural network, all the while keeping me motivated. I would also like to thank Wyke, for her additional efforts in developing the neural network used in this thesis. She patiently helped me tackle the challenges that came up when trying to work with the network. My gratitude also extends to the masonry experts who consulted on this thesis and the large group of raters who volunteered to provide similarity labels to the masonry crack pattern image pairs. Their enthusiasm and dedication were not only instrumental to this thesis, but highly encouraging to me as well.

I am grateful to my parents and to my sister, for the love and trust that they have always shown me, especially during my time in Delft. I would also like to thank Achachan and Ammoomma for being my unflinching pillars of support. Many thanks to my friends in Delft, for their valued feedback and countless coffee breaks that helped me get to the finish line. Finally, I would like to thank my friends in India, especially Renuka, Nandha, and Jayadev, for always believing in me even when I found it difficult to. Thank you all, truly.

Krishna Ajithkumar Pillai
Delft, February 2022

CONTENTS

ACRONYMS

LIST OF FIGURES

LIST OF TABLES

1 INTRODUCTION

1.1 BACKGROUND AND MOTIVATION

Masonry buildings are prone to damage as a result of subsidence or other environmental factors. Cracks are a type of damage in masonry structures that can severely weaken them. In order to mitigate and prevent the occurrence of these types of cracks, it is important to know what caused them. An important step in the masonry damage evaluation by masonry experts includes attempting to relate the observed crack patterns to earlier observations from cases with known damage cause. However, this process is largely dependent on the availability and experience of these experts, their comprehension of global mechanisms and local conditions. All of these limiting factors make this process expensive, time-consuming and subjective.

Convolutional neural networks (CNNs) may offer a robust and dependable means to overcome these limitations by automating the assessment of masonry crack patterns by processing their images. However, there is only limited literature available on this front. Although efforts have been made previously to study crack pattern classification using machine learning in concrete [2] and even in masonry [3], no attempt was made to correlate these crack patterns to relevant causes for damage. In 2019, an autonomous means of concrete crack detection was attempted using a CNN with a success rate of 90% [4]. Another study conducted in 2019, used photogrammetry and distinct element modelling to identify the most probable cause of existing cracks in masonry [5].

In 2019, researchers at TNO used a statistics-based approach to generate crack patterns on 2D masonry façades using Markov walks [6]. A Deep Neural Network (DNN) embedding was developed to classify these images into damage cause classes which achieved a high performance accuracy. However, this method was unable to account for dissimilarities within the same class. In 2020, a regression of crack pattern similarity assessment was attempted by fitting a function that takes two crack pattern images as input and outputs a single scalar, the similarity between the patterns. A Siamese neural network was fitted to crack pattern image pairs and their similarity was predicted using the labels provided by a masonry expert with high accuracy. It was also able to capture similarities within a class. The performance of this network was also tested with image pairs that had more reliable similarity labels given by a masonry expert, yielding good results. Continuing this line of research, this thesis attempts to generalise the model to automate the similarity assessment of crack pattern image pairs using neural networks by fitting them to different varieties of crack pattern image pairs.

In order to perform this automation, large amounts of data are required for fitting these neural networks so as to achieve high degree of accuracy in assessment. This data is difficult and expensive to obtain. In this project, finite element models are used to generate image pairs of crack patterns in masonry façades, which are then assessed by experts to determine the similarity between them. Training the neural networks using these crack pattern image pairs annotated by the experts helps in overcoming the constraint of the large data requirement. Hence, the goal of this

MSc thesis is to automate the process of similarity assessment of masonry crack patterns using a CNN fitted to the data obtained from finite element modelling and annotated by experts.

1.2 RESEARCH QUESTIONS

The aim of this project is to develop a workflow to fit a neural network for an automated similarity assessment of masonry crack patterns using data obtained from detailed computational physics-based (i.e. finite element) models. Therefore, the main research question for this MSc thesis is:

How accurately can a neural network – fitted to data generated from finite element models – predict masonry crack pattern similarities?

The following sub-questions will also be addressed during the course of this project:

- What criteria are used by masonry experts to assess the similarity in crack pattern image pairs? What is the agreement between the masonry experts in using these criteria?
- What are the parameters that should be varied in the finite element models to generate different crack patterns (e.g.: geometry, loading, material properties)?
- How generalisable is the fitted neural network in dealing with different masonry façades?
- How do the fitted neural networks perform on real world crack patterns?

1.3 SCOPE OF THE THESIS

The automation of crack pattern similarity assessment is a potential tool to aid masonry experts in their judgement. The similarity assessments of crack pattern image pairs made by volunteer raters were used to establish the ground truth of the neural network. Thus, the subjectivity in the assessments made by the raters were passed on to the neural networks. Therefore, the accuracy of the neural networks developed with the help of this thesis has the following limitations:

- This thesis mainly explores masonry damage due to soil subsidence along with a few instances of damage due out-of-plane loads. Masonry damage that may be caused by seismic loads, transient loads or due to stresses from temperature changes, shrinkage, or carbonation were excluded.
- Masonry experts conduct the damage assessments and crack pattern similarity assessments by evaluating numerous factors such as the type of masonry bricks used, the brick layout, the type of foundation of the structure, weather conditions, and geographical location to make these decisions. This information is sometimes lost when the similarity assessment is made solely based on images of crack patterns on façades.
- The neural networks is fitted to the opinion of volunteer raters who assessed the similarity of masonry crack patterns. This research is limited to the expertise of these 28 raters. The differences and biases in their judgements will affect the prediction accuracy of the neural network. A larger sample of raters

will help the neural network to become more representative of the assessment of masonry experts.

- The images of the crack patterns simulated were composed of only rectangular 2D shapes and of four aspect ratios. In reality, the masonry façades are of varying shapes and aspect ratios, and belong to a 3D structure. In the façades with openings, only one type of door and two types of windows were considered as openings. This sample of openings was not reflective of the façades seen in real life. This affected the generalisability of the neural networks.
- The finite element models were limited to the masonry material properties of clay bricks and a soil-structure interface that simulated an interface produced by a structure with a foundation on clayey soil. Other types of masonry bricks with different material properties and other types of soil-structure interfaces may result in different crack patterns and may affect the generalisability of the neural networks.

1.4 METHODOLOGY

In the step towards the automation of the crack assessment by using a neural network, large quantities of data were required to fit the neural networks. In this MSc project, the generalisability of the simulated models were tested using façades of different crack pattern types, geometries and method of generation. Given below is the step-by-step approach that is used during the course of the project:

1. A literature review of masonry damage assessment, crack pattern similarity, computational modelling of masonry damage, use of deep neural networks in structural damage assessment, and a recap of the work done in 2019-20 at TNO was carried out.
2. Parametric finite element models were developed to get crack pattern images and their similarity labels. From the 60 crack patterns introduced by [7], eight crack pattern archetypes were chosen to be modelled in the statistics-based approach. Additionally, four more crack patterns seen on masonry façades with openings were chosen. Eight out of these twelve crack patterns were parameterised and generated in large quantities using the finite element models. These crack patterns were used to fit and test the neural network in order to perform the automated masonry damage assessment.
3. Kleijn [1] developed and implemented the mathematical model of the deep neural network. This neural network was fitted to and tested with crack pattern image pairs that were synthetically generated. Kleijn's research focused on generating image pairs using a statistics-based approach, while this thesis focused on generating image pairs using the computational physics-based approach [1]. Additionally, real-world images were also collected jointly for this purpose. These images were then assigned into image pairs so that their similarity can be assessed.
4. The image pairs were annotated by masonry experts using three pre-defined labels to measure their similarity. These labels were chosen such that the assessment of damage similarity by the masonry experts can be quantified. An inter-rater agreement analysis of these labels and raters was done as well.
5. Around 75% of the annotated data was used to fit the neural network by Kleijn [1]. The prediction accuracy of the neural networks was tested with the remaining 25% of the annotated data. The generalisability of this neural network was also tested by seeing how well it can predict a crack pattern type that it was not fitted to.

1.5 OUTLINE OF THE THESIS

This thesis starts with a comprehensive literature review in Chapter 2. This chapter intends to provide the reader with insight into how masonry damage assessment is conventionally done, how masonry cracks have been historically quantified and analysed, and how the relevant aspects of these have been incorporated into the thesis. Chapter 2 also includes some background on how computational modelling has been used to model and assess damage in masonry structures. The chapter also deals with the use of deep neural networks in assessing structural damage, particularly with judging similarity between different crack patterns. A summary of the ongoing work at TNO is also provided.

In Chapter 3, the tools and methodology used to conduct the research work are presented. The chapter explains how the data collection process for the statistically generated data was improved, after talking to a few more masonry experts. It then delves into how the new data generation and collection process was done for the computational physics-based approach and for the real-world images. How this data was then used to establish the ground truth for neural networks is explained as well. Additionally, the chapter explores how the data from the statistics-based approach, computational physics-based approach, and the real-world images were processed into a matrix form that could be fed into the neural networks. The metrics used to establish the inter-rater agreement is explained in detail as well. The last section includes the details of the neural network used in this thesis, along with an explanation of the metrics used to assess the performance of the neural network.

Chapter 4 presents the parametric finite element models used to simulate the various crack pattern archetypes in masonry façades of various dimensions. The models are created and analysed in DIANA FEA 10.4 and Python. It discusses what specific loading and boundary conditions were used to simulate each crack pattern archetype, what analysis methods were used and what the resulting crack patterns looked like. A sensitivity study of the finite element models and the limitations of the models are also provided.

In Chapter 5, the similarity judgements by the raters and their agreement with each other regarding the similarity labels are presented. This chapter also includes the results of the similarity assessment of the neural network fitted to and tested with the image pairs generated using the computational physics-based approach. The results of the generalisability experiments of the neural network are also discussed here.

Finally, in Chapter 6, the conclusions of the research are presented. The chapter highlights the key findings and gives recommendations for further research.

2 LITERATURE REVIEW

2.1 INTRODUCTION

The definition of damage in buildings is closely tied to people's expectations of a building's performance, such as safety, health, utility, and aesthetic appeal [7]. Consequently, the perception and resulting assessment of damage are oftentimes very subjective. Although there are numerous symptoms of damage in buildings, the three main symptoms of damage are: crack development, deformation and tilt [7]. Given the scope of this thesis, the literature review focuses on the assessment of damage in masonry structures using crack patterns, the computational modelling of these types of damage in masonry and the use of neural networks to help automation of the assessment of masonry crack patterns. This literature review also contains a summary of past work at TNO to develop a means to automate the assessment of masonry crack pattern similarity.

2.2 DAMAGE ASSESSMENT OF MASONRY STRUCTURES

Masonry is a widely used construction material in the Netherlands and all over the world. It has a relatively long service life, and is comparably inexpensive to maintain owing to its ability to be incrementally repaired. This is due to the modular nature of masonry which consists of bricks joined together by mortar, unlike monolithic materials like concrete. There are various types of loads that can cause damage in masonry structures. These include damage from soil subsidence, landslides, seismic loads, temperature loads, and transient loads.

A crack in a masonry structure can potentially be the first warning of a significant issue that in time might jeopardise its serviceability or stability. To thoroughly and methodologically study cracks, a crack pattern survey can be employed. Such a survey gives valuable information on local weak points and stiffness changes in the structure. Furthermore, it typically involves classification of cracks based on their direction (horizontal, vertical, diagonal, or curved), their dimensions (length and width) and geometry (whether it passes through the wall or not) [8]. These surveys also usually include information on possible previous repairs and re-opening of retrofitted cracks.

Ground-level inspections, tactile inspections utilising scaffolding or cherry pickers, drone-based inspections, or a combination of these approaches are generally used for masonry façade inspections. Due to their time and resource intensive nature, these procedures are expensive. This has led to a growing interest in the need to automate image-based or visual damage assessment [9].

2.2.1 Visual inspection of cracks

Visual inspection of masonry structures is used as a first step of assessment for damage in masonry structures [8]. It can provide a basis for systematic data collection providing key information of the building characteristics detectable to the naked eye such as the size and texture of the masonry and the mortar, and discontinuities

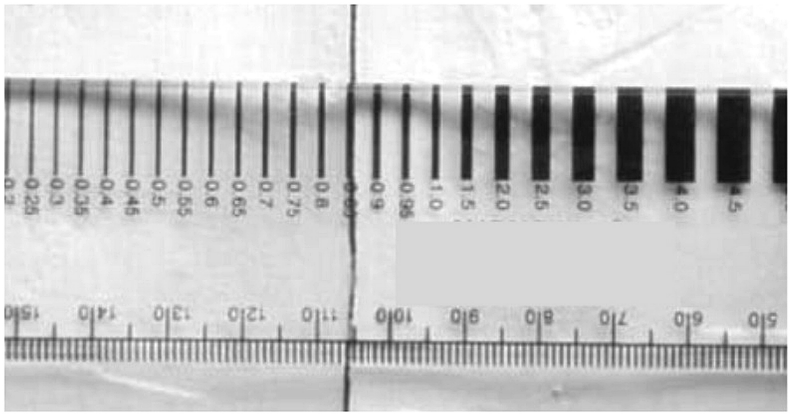


Figure 2.1: A crack width gauge for measuring the crack width [12].

(i.e. cracks and deformations) in the structure. This information is important in the diagnostic procedure [8].

Engineers or masonry experts typically investigate damaged structures by studying the length and width of crack patterns, the area of the cracked zone and the presence of crushing [10]. Crack length is the distance through which cracks propagate on the masonry surface. Usually, the brick-mortar interface provides the path of least resistance for crack propagation. One study attempted to define the laws of evolution of the crack length with the damage progression in masonry structures of various sizes and discovered that the stiffness of the brick-mortar interfaces varied inversely to the third power of the ultimate crack length [11].

Measuring the crack width is an important step in the crack monitoring process. This is usually done using special gauges as shown in Fig. 2.1 [12]. There have been many attempts to link observed crack patterns and their widths in masonry structures with possible interventions needed [13–15]. Overall, masonry cracks can be categorised into three classes [13]:

- Cracks affecting the aesthetics: These only affect the appearance and do not affect the functioning or stability of the structure.
- Cracks affecting the serviceability: These cracks affect the regular function of the structure. For example, the structure is no longer impermeable to water.
- Cracks affecting the stability and safety: Cracks that affect the load carrying capacity of the structure and are at increased risk of collapse.

These three categories of crack patterns are further divided into six categories as summarised in Table 2.1. An attempt was also made to distinguish and quantify *visible* masonry crack patterns [16]. A dimensionless damage assessment parameter ψ was introduced which related to the number of visible cracks. Only cracks with widths $>$ 0.1 mm were considered for this. The lower limit was kept at 0.1 mm because it was seen that once a crack width achieves a width of 0.1 mm and a length of 100 mm, the crack is visible with the naked eye. This was based on the observations of both laboratory and real-world damage specimens. ψ is calculated using the formula in Eq. 2.1 [16]. This damage parameter was used to divide visible masonry cracks as shown in Table 2.2.

$$\Psi = 2 \cdot n_c^{0.15} \cdot \hat{c}_w^{0.3} \tag{2.1}$$

Table 2.1: Damage categories of masonry structures dependingon crack width and ease of repair [13–15].

Category	Type of crack	Crack width (mm)	Type of repair required
0	Aesthetic issues	Hairline cracks 0.1 mm	No action required.
1		Fine cracks of 0.1 - 1 mm	Usually repairs limited to internal wall finishes. Can be treated with normal wall finishing.
2		Cracks between 1 mm - 5 mm	Doors and windows may stick out, requiring intervention. Cracks are usually filled. Some external repointing may be done to ensure water tightness of the structure.
3	Serviceability issues	Cracks between 5 mm - 15 mm or several smaller cracks	Doors and windows may stick out, service pipes may fracture, water tightness is affected. Repairs include repointing of external bricks. Small areas around the repointing may need replacement of bricks.
4		Extensive damage, cracks between 5 mm - 25 mm or several smaller cracks	Windows and doors distorted significantly. Walls leaning in or out of the plane noticeably. Repairs may include breaking out the wall and replacing in sections.
5	Stability issues	Structural damage, cracks 25 mm or several smaller cracks	Beams lose bearing and masonry walls require shoring. High danger of instability of the structure. Major repairs including partial or full rebuilding of the structure.

Where,

n_c = Number of cracks in the masonry specimen.
$\hat{c}_w$ = Crack width calculated with the weighted width and average lengths of all the cracks present (in mm). It is calculated as

$$\hat{c}_w = \frac{\sum_{i=1}^{n_c} c_{w,i}^2 \cdot c_{L,i}}{\sum_{i=1}^{n_c} c_{w,i} \cdot c_{L,i}}$$

Where, $c_{w,i}$ is the maximum crack width along the i^{th} crack (in mm) and $c_{L,i}$ is its length (in mm).

Table 2.2: Damage categories based on damage parameter ψ [16].

Damage state	Damage level	Damage parameter value	Crack width (mm)
DS0	DL0	ψ <1	Not visually detectable
DS1	DL1	1 <ψ <1.4	Up to 0.1 mm
	DL2	1.5 <ψ<2.4	Up to 1 mm
	DL3	2.5 <ψ <3.4	Up to 5 mm
DS2	DL4	ψ >3.5	5 to 15 mm

DS: Refers to damage states [13]
DL: Refers to damage levels [15]

However, visual inspection has its disadvantages. It is expensive, and more significantly, its accuracy is hampered by a number of factors, including visual access to the damaged structure, competence of the human assessor, and human error [12]. Image based visual inspection of crack patterns has been gaining popularity due to its speed, larger area of coverage of the structure, and repeatability. Additionally, there have been several studies on the use of Unmanned Aerial Vehicles (UAV) or drones for an image-based recognition/detection of cracks [17]. Automated image based visual inspection, which analyses digital photographs of damaged structures using computer algorithms, has the ability to speed up and objectify damage inspections. Such methods can assist in detecting structural deterioration, classifying it, and can aid in making repair decisions [10].

2.2.2 Digital Image Correlation

Digital Image Correlation (DIC) is a non-destructive and non-contact technology that employs high-resolution cameras to obtain digital photographs of specimens undergoing damage progression. These photographs are then processed using an image processing tool and the data is analysed to acquire meaningful information such as deformations, surface shapes, and in-plane strains [18]. DIC is usually used for smaller test specimens, although it is now used for larger structures as well, even for full sized walls or façades. It was seen that the accuracy of the results was not compromised, provided that the images were recorded in sufficiently high resolution [19].

The DIC systems can provide information about the absolute size and scale of the test specimen and damage cracks using a reference length of the specimen [20]. Additionally, a correlation algorithm involving the stochastic intensity pattern (speckle) on the masonry surface can be used to determine the relative position of each point in the image by computing the entire field of displacements and strains. DIC, when combined with extremely high-resolution cameras, can even detect crack initiation before it is visible to the naked eye and can identify areas of stress concentration at very low load levels [20].

In an attempt to quantify masonry damages, two damage scores were defined after conducting experiments on plastered unreinforced masonry and taking its photographs: the Normalised Crack Area (NCA) and the Normalised Crack Length (NCL) [21]. The required images were taken in three damage stages of the masonry: no crack, visible crack and plaster fall-off. The NCA and the NCL were determined based on an expert survey and a multivariate logistic regression. With this approach, it was possible to calculate the damage scores of the masonry wall displacement due to various causes (such as induced ground motion), assess the likelihood of seeing a specific damage state, and measure the subsequent risk of collapse. The damage scores were calculated as given below in Eq. 2.2 and 2.3 [21].

1. Normalised Crack Area (NCA):

$$\text{NCA} = \frac{\textit{Damaged area of the plaster}}{\textit{Total area of the plaster}} \tag{2.2}$$

   The NCAwas calculated from the photographs of the masonry cracks after the images were processed using a software called Vic2D to obtain von Mises strain maps that were then processed in MATLAB.

2. Normalised Crack Length (NCL):

$$\text{NCL} = \frac{\textit{Sum of lengths of all crack}}{\textit{Length of wall diagonal}} \tag{2.3}$$

Therefore, a crack along the whole wall diagonal has an NCL score of 1. The NCL was also calculated from the von Mises strain maps obtained from Vic2D. The crack length was estimated by using the edge detection method on each wall's cumulative von Mises strain map.

2.2.3 Damage and crack pattern similarity metrics

One of the most important aspects of masonry damage assessment is determining whether crack patterns are comparable because the degree of similarity across patterns frequently correlates with the likelihood of similar origins. Experts usually assess similarity 'manually' based on their experience, comparing an observed pattern to hundreds of patterns they've seen before and remembered. In practice, this similarity is often determined qualitatively rather than quantitatively.

A few *simple* measures used to quantify the similarity in cracks in synthetically generated cracks [22] are given below:

1. Sum of vector norms (SVN) distance: The distance between crack patterns on the finite element model is the difference between their SVN distances from the same origin modelled with the same mesh [23].

2. 2-norm (M2N) distance: It is computed the same way as is done for the SVN distance measure by evaluating the absolute difference between their 2-norm (M2N) matrix distances [23].

3. Frobenius norm (FRO) distance: As the FRO can be read as a special type of the matrix norm, this similarity distance measure is closely connected to the M2N distance [23].

4. First Wasserstein (W1-1D) distance: This crack pattern similarity assessment measure is calculated using the first Wasserstein (W1-1D) distance [24].

5. Jaccard (JAC) distance: The Jaccard index is a numerical estimation of how much two objects overlap with each other [25].

It was seen that these similarity measures were not suitable to assess the crack pattern similarity [22] due to their sensitivity to size and shape of the finite element mesh, as well as due to their inability to give unique solutions for cracks caused by simple settlement profiles [23] [24] [25]. It was seen that there was no crack pattern similarity measure readily available in the current literature that can overcome these limitations.

Additionally, there are also several shape similarity and open curve similarity metrics available for pattern recognition applications. The term *shape* has no uniform definition. Colour or texture can impart shape impressions, from which a geometrical representation might be constructed [26]. Curve matching or curve similarity assessment attempts have been limited to finding reliable methods for matching shapes or closed curves because they have object boundaries that can be matched to other shapes by finding similar curve areas. However, most of these methods are not suitable for the matching open curves, as they do not have a well defined boundary or enclosed area [27]. Often, it is necessary to first define what is meant by a shape, before a similarity measure can be made for it. For example, an open curve was defined as a finite sequence of equispaced points on a curve. The shape of this open curve was defined independent of origin and orientation of the curve and this definition was used to establish a few distance-based open curve similarity measures [28]. Designing a similarity measure based on the characteristics of a curve can be achieved through numerous means such as remapping,

normalisation and by extension with an empty set [29]. In the context of crack pattern similarity assessment, partial matching of curves is a difficult and important problem. When two shapes contain smaller similar portions, the dissimilarity must be smaller, and the measure should not excessively penalise regions that do not match [29].

However, these generic open curve similarity measures are not suitable for assessing crack pattern similarity as these measures define the *shape* as independent of the origin, location, and orientation of the curves under assessment. These parameters are critical in evaluating the severity and similarity of the crack patterns and thereby identifying their damage cause. These measures also fail to account for the mechanical behaviour of brittle materials such as masonry.

2.3 COMPUTATIONAL MODELLING OF MASONRY STRUCTURES

Masonry refers to a composite material composed of numerous smaller individual units bound together with mortar in a variety of configurations [30]. Masonry has been successfully modelled using numerical methods, such as the finite element method. This has enabled the analysis of masonry structures without (or limited) experimental testing [31]. Numerical methods have also been successful in accommodating the complex geometries and material nonlinearities- of masonry buildings [32]. An appropriate numerical model must be chosen such that it is capable of predicting the behaviour of the masonry from the linear elastic stage, through cracking and finally until the complete loss of strength [33].

Computational modelling strategies of masonry can also be broadly classified as discrete models and continuum models [15]. Discrete models simulate the bricks, mortar and brick mortar interfaces separately. However, the larger the number of elements, the more computationally expensive the model is. The failure criteria in the discrete model is related to tension, compression and shear failure modes [15]. In the case of continuum models, masonry is treated as a homogeneous material. The constitutive law assumes that the material behaviour is related to the average stress and strain. Additionally, this model assumes that the cracks are smeared over an area, and that the crack widths are calculated from the crack strains and bandwidths [34]. These models often require additional research to obtain necessary material properties and can suffer from convergence problems [35]. Continuum models also allow for mesh discretisation that is not controlled by the presence of heterogeneities in the masonry. This also means that the mesh size can be of comparable or larger size than the brick size, which can additionally reduce the computational demand [36]. Continuum models can be generally divided into three categories: the plasticity-based approach, the damage-based approach and the smeared cracked approach [15]. In the smeared cracked model, the material is assumed to have anisotropic properties and the cracking is assumed to be distributed over an area. This method has proven to be effective in modelling brittle cracks in the tension zones, provided that the model adopts adequate compressive and tensile strengths for the masonry to account for the global response of the structure [37]. The use of a smeared crack model with tension softening has also shown to be successful in modelling the progressive failure of masonry [15].

Sensitivity studies on numerical models of masonry have shown that the material parameters such as fracture energy and tensile strength significantly affected the load carrying capacity and crack width of the continuum models [38]. Addition-

ally, the presence of openings in façades leads to crack localisation around these openings [38]. The response of masonry structures to damage-inducing loads is also sensitive to the initial loads acting on the structure. The higher the initial load (such as self-weight, pre-compression load) on the structure, the more will be the stress concentration in the structure, leading to more extensive crack formation [39].

The behaviour of masonry structures to settlement induced loads also depend on the nature of the soil and the soil-structure interface. Light weight masonry structures subjected to expansive soil movements experienced cracks in masonry wall panels and led to crack separation and frictional sliding of the structure [40]. The extent of the damage and shape of the crack patterns are also dependent on the properties of the soil-structure interface. This was verified by comparing analytically calculated strains and numerically obtained crack widths of the same model [38].

2.4 THE USE OF DEEP NEURAL NETWORKS IN DAMAGE ASSESSMENT

Deep learning is a branch of machine learning that uses hierarchical architectures to learn high-level abstractions from data [41]. Deep learning algorithms can be used to solve complex problems by generating appropriate mapping functions provided that adequate labelled training data sets and acceptable models are available[42]. There is significant research available in the area of automated crack pattern detection and classification based on characteristics such as length, width, and severity of the crack [2, 43] [2] . However, there is a dearth of research to connect these crack patterns to their respective damage causes [6]. The advantage of using neural networks to analyse crack patterns in masonry structures is their ability to make predictions that are not easily possible through analytical means due to the spacial variation of its material properties. For example, in a study conducted on predicting the crack growth in masonry wallets, using a single masonry wallet base or a portion of the wallet base, it was seen that the neural networks provided an accurate prediction of the crack growth [44].

One of the ways in which neural networks can be used to perform crack pattern damage assessment is by classifying crack patterns on the basis of their damage cause. A study successfully used an Artificial Neural Network (ANN) based classification model for predicting the cracking loads of laterally loaded masonry wall panels of different dimensions based on their corresponding crack patterns derived from laboratory experiments. In order to do this, the crack patterns were converted to a numerical matrix and transformation functions were used to convert these matrices into matrices of the same dimensions. [45]. In another study, an automated pattern recognition in concrete crack patterns was attempted by using neural networks to perform a binary classification of cracks as structural and non-structural damage cause. The Convolutional Neural Network (CNN) used the Digital Image Structure and Texture Similarity (DISTS) index, an image quality index that assesses the human perception of image quality to compare the texture similarity between a reference image and a target image, and performed this classification with an accuracy of 97% [46].

Additionally, neural network based regression analysis can be used to establish a relationship between crack patterns and their respective damage cause. There is a study where an ANN model was used to successfully evaluate the remaining fatigue life of in-situ reinforced concrete bridge road decks from the crack patterns on their bottom surface. This study also showed that rather than using a huge number of a

single crack type, incorporating samples from multiple crack types aids in building a much more reliable and accurate ANN model for crack pattern assessment [47]. Another investigation used to effectively predict the curved fatigue crack propagation in brittle structures under varied amplitude loads, taking into account the underlying physical mechanisms of the damaged structure. This method was also used to conduct a parametric study to predict the crack growth by varying crack size, position, and orientation [43].

2.5 SUMMARY OF ONGOING WORK FROM TNO

In 2019, TNO attempted to create a fast and cheap automated assessment for the structural integrity of masonry structures [6]. This is a data intensive procedure that requires at least a few hundred crack pattern images with their corresponding damage causes in order to generate a neural network that can give sufficiently accurate results. Using real-world data to do this would be ideal, but this would be expensive and time consuming to collect, especially when the reliable damage cause labels are also required. Alternatively, synthetic data that is produced from a fully known model that attempts to depict a real system can be used. A distinction is made between the different types of synthetic data used in this project:

- Statistics-based data: Data generated from a mathematical model that does not explicitly incorporate any physics. Randomness can be incorporated into these models to increase its resemblance to reality.
- Computational physics-based data: Data that is simulated using a mathematical model that attempts to recreate the real world data by adhering to selected laws of physics. It can also attempt to incorporate randomness by using appropriate random realisations from a probabilistic distribution.

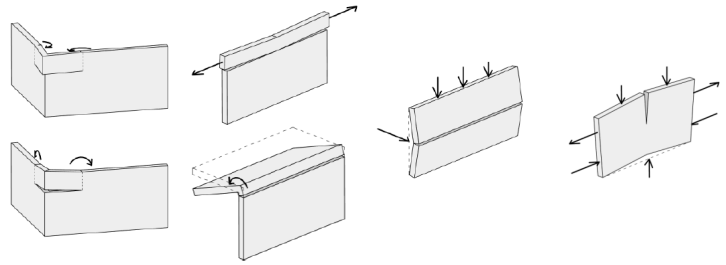
The statistics-based approach is capable of generating synthetic data orders of magnitude faster than the computational physics-based simulation approach. However, the data quality is inferior and hence less indicative of reality.

2.5.1 Classification of the crack pattern images

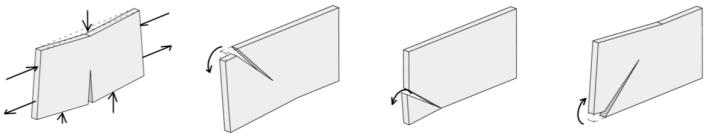
In 2019, crack patterns were generated from the statistics-based approach from 12 damage causes in façades with openings (Case 1) and without openings (Case 2) as shown in Section 2.5.1 and Section 2.5.1. The statistics-based simulation included an algorithm that generates lines in a predefined bounding box (i.e. the geometry of a façade) using Markov-walks. This algorithm was generalised to handle cracked façades of varying parameters such as façade dimensions, the number, dimensions and positions of the openings, the number of cracks, the crack initiation point, the crack orientation angle, the crack length, and how jagged a crack is.

Case 1: Masonry façades without openings

Ilse de Vent [7] introduced 60 different types of crack patterns along with their possible causes of damage. In this case, eight crack pattern archetypes were chosen from these 60 which share some common characteristics and was successfully parametrised. All of the crack patterns in this case were generated on façades of aspect ratio 3:2. The illustrations of these chosen eight crack archetypes and their realisations in the statistics-based approach are shown in Fig. 2.2 and Fig. 2.3.

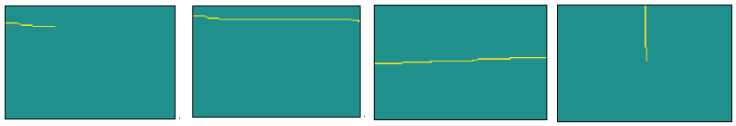


(a) Pattern ID 18 (b) Pattern ID 20 (c) Pattern ID 21 (d) Pattern ID 23

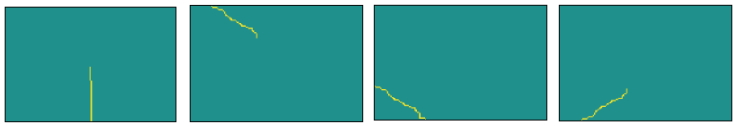


(e) Pattern ID 24 (f) Pattern ID 30 (g) Pattern ID 31 (h) Pattern ID 32

Figure 2.2: Illustrations of the chosen crack patterns for Façades without openings [7].



(a) Pattern ID 18 (b) Pattern ID 20 (c) Pattern ID 21 (d) Pattern ID 23



(e) Pattern ID 24 (f) Pattern ID 30 (g) Pattern ID 31 (h) Pattern ID 32

Figure 2.3: Example realisations of the chosen crack types on façades without openings using the statistics-based method [6].

(a) Pattern ID 101 (b) Pattern ID 102

(c) Pattern ID 103 (d) Pattern ID 201

Figure 2.4: Examples of the realisation of the chosen crack types on façades with openings using the statistics-based method [6]

.

Case 2: Masonry façades with openings

Random crack patterns were generated in the longitudinal masonry façade of typical Dutch dwellings for four different geometries. The crack pattern realisations are made in a similar fashion as in Case 1. All of the crack patterns in this case were generated on façades of aspect ratio 10:3, as shown in Fig. 2.4.

Results of the classification task

A CNN was used for the classification of the crack patterns. A CNN takes a crack pattern image as its input and applies convolutional filters to these images. The cosine similarity of the different crack images was calculated, which was used as a proxy for crack pattern similarity.

Results for Case 1: A total of 4000 cracks pattern images were simulated (2000 for fitting the model and the remaining 2000 for testing), with 500 images per class. With a classification accuracy of greater than 99%, the computed pairwise similarity between the simulated crack patterns was high between cracks of the same class and low between crack patterns of different classes.

Results for Case 2: A total of 8000 crack pattern images were simulated (4000 for fitting the model and the remaining 4000 for testing), with 2000 images per class and 500 for each unique façade-crack pattern archetype pair. With a classification accuracy of greater than 99%, the CNN almost perfectly classified the simulated patterns in their corresponding classes similar to the model in Case 1.

2.5.2 Regression of crack pattern image pair similarity using a proxy assessment

In 2020, it was attempted to develop a function that takes two crack pattern images as input and outputs a single scalar: the similarity between the patterns. As mentioned in Section 2.2.3, no crack pattern similarity assessment was available in the current literature. A simple function was devised to estimate it based on the parameters that was used to generate the patterns in the statistics-based approach. This similarity function assigned a score to each image pair in the range [0, 1], with 1 indicating highest similarity and 0 indicating maximum dissimilarity. Although this function is not representative of real-world masonry crack similarity assessments, it was the best performing function used to annotate the image pairs with similarity labels. The regression described above was attempted with the help of a Siamese neural network that was fitted to crack pattern image pairs having similarity labels. In order to decide the weights that is to be used in the neural network, two loss functions were devised: 1. The difference between a model-estimated similarity label and the ground truth, expressed as a mean squared error, 2. A triplet loss using the Pattern ID to minimise the distance between similar cracks while maximising the distance between dissimilar cracks.

Results of the regression task

4000 image pairs of Case 1 (façades without openings) were generated, of which 2000 was used for fitting and 2000 was used for testing. The Siamese neural network was able to achieve excellent regression with a similarity prediction accuracy of >99% and a similarity loss of 100%. In order to get more reliable similarity labels, a masonry expert manually annotated 395 image pairs with a slightly modified similarity label. This new label judges the similarity in ranges. A value in the range of [0, 0.4] was assigned if the image pairs were maximally dissimilar, a value between (0.4, 0.96] was assigned if there was a degree of similarity and a score of

(0.96, 1] indicated that the images were maximally similar. Out of these 395 images, 296 (75% of the annotated images) were used for fitting the neural network and 99 (25% of annotated images) for testing. The results showed that the neural network was able to predict the similarity with an accuracy of around 92%.

Shortcomings of the fitted regression neural network

After further discussions with other masonry experts, it was concluded that this procedure has some shortcomings:

- The sample of annotated image pairs used for fitting the neural networks was relatively small and contained annotations from only one rater. This neural network need not perform as well when tested with another rater's similarity assessments.
- The masonry expert manually judged the similarity of each image pair. In order to do this, the expert assigned a score in the ranges of numbers based on their assessment. A value in the range of [0, 0.4] was assigned if the image pairs were maximally dissimilar, a value between (0.4, 0.96] was assigned if there was a degree of similarity and a score of (0.96, 1] indicated that the images were maximally similar.

  The disadvantage of this method was that each expert would assess 'similarity' based on their experiences and biases. A wide range of factors affect the judgement of similarity.
- The images only contained information on the size of the façade, and on the location and length of the cracks. The crack width is an important parameter that greatly affects the assessment of the structural damage that was missing in the previous study.

2.6 CONCLUDING REMARKS

The first step in masonry crack pattern assessment is a visual inspection of the damage to determine its underlying cause. Similarity in damage patterns may indicate a similarity in the damage cause as well. However, a study of existing literature shows that there is no consensus on how experts interpret the similarity in crack patterns. Neural networks can be used to attempt an automation of the similarity assessment performed by these masonry experts. The data needed to fit and test these neural networks can be generated synthetically (e.g. statistics-based and computational physics-based approaches) or by collecting the real-world images of cracks. The former is a faster and cheaper approach while the latter could produce the most accurate results, representative of real-world damage in masonry. The computational physics-based approach, which is used in this research, aims to simulate crack patterns using realistic material properties, loads and boundary conditions.

Although there is sufficient literature documented in the area of automated crack pattern detection and characterisation, there is a lack of research to establish a connection between these crack patterns and their respective damage causes. Neural networks can be used to determine the damage cause of crack patterns by classifying the the crack patterns into classes of damage causes or by using regression analysis which outputs the similarity between any given pair of cracks using the images of these crack patterns as input. In the research conducted by TNO in 2019, a CNN was used to perform this similarity assessment by classifying the crack patterns generated using the statistics-based approach. In the work done in 2020, a Siamese neural network was developed to obtain the similarity between the crack

patterns from the crack pattern images. In this project, the similarity of crack patterns is assessed by studying crack patterns that were obtained through computational physics-based simulations as well as from real-world crack images.

3 TOOLS AND METHODS

3.1 INTRODUCTION

This chapter provides an overview of the tools and methods used to gather data from the statistics-based approach, the computational physics-based approach, and the real-world images. The procedure to give similarity labels to the crack pattern image pairs is also discussed. Additionally, the tools used to establish the inter-rater agreement are explained. Furthermore, the data from all the three approaches need to be converted into a similar format that can be used to feed the neural networks. To facilitate this, the chapter explores how masonry façades from different sources and of different dimensions can be handled by the same neural network. Finally, the details of the neural network, developed by Kleijn [1], are explained. The metrics used to evaluate the performance of this neural network are also discussed.

3.2 DATA GENERATION FROM THE STATISTICS-BASED APPROACH

The regression neural network developed, fitted and tested with the rater's similarity assessment of crack patter image pairs at TNO has certain shortcomings. These are discussed in Section 2.5.2. In an attempt to overcome these shortcomings, a modified approach was taken to generate the data from the statistics-based approach and annotate theme with similarity labels. The details of this modified approach used in this thesis are:

- The images generated from the statistics-based approach were modified to contain information on the crack width. Using the information provided in Section 2.1, it was decided to incorporate crack widths in the range of 1 mm to 30 mm in the models. Thus, three categories of maximum crack widths corresponding to different degrees of damage were defined as given below.

  - **Category A:** 5 mm to 15 mm
  - **Category B:** 15 mm to 25 mm
  - **Category C:** 25 mm to 30 mm

  Before generating each crack pattern model, a width category was assigned to it. The maximum crack width was randomly chosen from the range of values associated with the width category. The width reduced linearly to zero along the crack length.

- Three separate similarity labels were defined to aid the raters with their similarity assessment of the the crack pattern image pairs, instead of using just one label as was done in the previous experiment at TNO. The three labels used are:

  - **Crack pattern similarity label**: Using this label, the rater assesses the similarity of the crack patterns by considering the geometry of the crack

Table 3.1: Distribution of image pairs in the sample of statistically generated images pairs annotated by the raters.

Image pair selection criteria	Percentage of the sample (%)
Same Pattern ID and same category of crack width	20
Same Pattern ID and any category of crack width	25
Similar Pattern ID and same category of crack width	20
Similar Pattern ID and any category of crack width	25
Randomly chosen image pairs	10

(i.e. the crack initiation point and crack orientation) and the cracking mechanism.

- **Damage severity label**: This label aids the rater to assess the severity of the crack patterns by considering the crack length and the crack width.
- **Overall similarity label**: The rater uses the overall similarity label to assess the overall similarity of the crack pattern image pair by combining the previous two labels, however they feel fit.

Each label has five ratings that each rater can choose from: *very dissimilar*, *dissimilar*, *similar*, *very similar*, and *I cannot say*.

- The sample of image pairs shown to the raters were chosen such that it contained enough pairs of images for each of the five rating classes. The input parameters used to generate and parametrise the images in this approach include the façade dimensions, crack length, crack width, initiation point, orientation and jaggedness. As the crack width contains critical information that influences the similarity assessment, it was used as a criteria to choose the image pairs, in addition to the crack pattern type. A detailed description of this is provided in Table 3.1. 3000 image pairs following this criteria were generated for the similarity assessment.

- Each image pair was labelled at least three times. This gives information on how each individual rater perceives the labels and how they chose the similarity rating for each label. Out of the 3000 image pairs prepared, 2587 image pairs were annotated at least three times.

- A total of 28 raters attempted this masonry similarity assessment task, including 7 industry experts (including both masonry experts and structural engineering experts), 7 PhD students and 14 MSc students. The details of all the raters are provided in Appendix A.

3.3 DATA GENERATION FROM THE COMPUTATIONAL PHYSICS–BASED APPROACH

To synthetically generate cracks that are more of a realistic representation of the real-world masonry crack patterns, a computational physics-based approach was proposed. The computational physics-based approach employs the laws of physics. Such models can be parametrised by using appropriate random realisations from a probabilistic distribution. In this thesis, these physics-based computational models are generated using the finite element software DIANA FEA 10.4. A more detailed description of the image pair selection and similarity assessment procedure for the images generated using the finite element analysis is given below:

Table 3.2: Distribution of image pairs in the sample of images generated from the computational physics-based approach annotated by the raters.

Image pair selection criteria	Percentage of the sample (%)
Same Pattern ID and same aspect ratio of the façade	10
Similar Pattern ID and same aspect ratio of the façade	12.5
Similar Pattern ID and different aspect ratio of the façade	25
Different Pattern ID and same aspect ratio of the façade	10
Different Pattern ID and same aspect ratio of the façade	12.5
Different Pattern ID and different aspect ratio of the façade	25
Randomly chosen image pairs	10

- In total, 8 out of the 12 crack patterns used in statistics-based approach are used to the generate the crack pattern images in the computational physics-based approach. By parametrising these 8 models, 500 image pairs are generated. The parametrisation of these finite element models is achieved by varying the material properties of the masonry, the dimensions of the façade, the number and positions of the openings and the location and magnitudes of the external loads, as explained in Chapter 4.

- These 500 image pairs were annotated with three similarity labels, similar to the image pairs from the statistics-based approach (Section 3.2). However, these three labels are rated by the raters using only four similarity ratings: *very dissimilar*, *dissimilar*, *similar*, and *very similar*. The *I cannot say* rating, used in the similarity assessment of the images from the statistics-based approach, was avoided [1].

- Each image pair is annotated at least three times, similar to the assessment procedure used for the images in the statistics-based approach.

- The sample of image pairs shown to the raters were chosen such that it contained enough pairs of images from each of the five rating classes. This is important for the fitting of the neural network in order to achieve a similar prediction accuracy for each of the rating class. Table 3.2 shows the used criteria for the selection of the image pairs shown to the raters. The parametrisation of the façade geometry and crack pattern archetype was achieved by changing the input parameters of the finite element model. A detailed description of this is given in Section 4.3.1 and Appendix C.

- A total of 7 raters annotated 500 image pairs at least thrice. These raters included 3 industry experts and 4 MSc students. The details of these raters are provided in Appendix A.

3.4 DATA FROM REAL-WORLD CRACK PATTERNS

As mentioned in Section 2.5, the data collected to fit and test the neural networks are of increasing complexity. Real-world masonry crack patterns are the most complex, time-consuming, and expensive data to collect.

1 The *I cannot say* appears in only approximately 0.7% of the overall sample of annotations of the statistically generated image pairs. Such an unbalanced data set can impair the neural network's performance and diminish its prediction accuracy. This is described in more detail in 5.2.1

Figure 3.1: Map of Gouda centrum shaded by the year of construction of the buildings [48].

In order to collect real-world data and make an estimate of how many man hours are needed to collect enough usable real-world crack pattern images, two visits were made to the city of Gouda in the Netherlands. The collected data was also used to see how well the neural network was able to judge the similarity between the crack pattern image pairs.

The city centre of Gouda has many buildings constructed 200 - 400 years ago, as shown in Fig. 3.1. Many of these buildings have subsidence induced cracks, and the reported damages are expected to go up as the buildings continue to age.

The following general guidelines were adhered to in the real-world data collection process:

- Only cracks with a width $\geq$ 1 mm and with a length $\geq$ 100 mm were collected. Cracks smaller than this are not easily detectable with the naked eye or on digital cameras.
- It is necessary to take the picture of the whole cracked façade. The aspect ratio of the façade, the location, and the ratio of crack length to façade dimensions are important information, necessary for the similarity assessment.
- Photos were taken during the daytime. The cracks seemed to be more visible when taken on cloudy days as opposed to bright sunny ones. The surrounding brightness affected the visibility of the cracks on the photographs.
- Care was taken to avoid obstructions in front of the masonry façades wherever possible. Ideally, the angle of the photos must be perpendicular to the façades in order to avoid tilt or distortion of the images.
- A high resolution digital camera was used to take the images of the cracks. Having a resolution of 2-4 pixels per millimetre would be ideal. The resolution of the photograph is important as it impacts the detectability of the cracks. In this thesis, all the images were collected with a pixel resolution of around 6000 x 4000.
- It is highly recommended to use a camera with a tripod to avoid distortion due to shaking. It is also recommended to geo-tag your photographs or have a separate application on your phone to log GPS co-ordinates.

Over two visits that spanned around 10 hours in total, 47 images of masonry façades with settlement induced damages were collected. Most of these images were collected from buildings near the city centre of Gouda. Buildings further away from the city centre showed little to no cracks of sufficient width and length. Of these 47 images, only four are similar to the Pattern IDs and façade geometries considered in the statistics-based approach. These four images are shown in Fig. 3.2.

(a) Real World Image ID 1.

(b) Real World Image ID 2.

(c) Real World Image ID 3.

(d) Real World Image ID 4.

Figure 3.2: Real-world masonry crack patterns obtained from Gouda, The Netherlands.

A new sample of images pairs were generated using these real-world images as described below:

- This sample of image pairs were generated using the images from all the three approaches: the statistics-based approach, the computational physics-based approach and the real-world images. The sample of image pairs were chosen such that a comparison could be made between real-world images and images from the statistics-based approach, as well as between images from the statistics-based approach and the computational physics-based approach. A comparison between the real-world images and the images from the computational physics-based approach was avoided as no real world images could be obtained that were similar to the crack patterns generated using the computational physics-based approach.
- The image pairs were assessed for their similarity using the three labels as mentioned in Section 3.2. These labels are also rated with four rating classes as mentioned in Section 3.3. Each image pair is annotated at least thrice.
- The sample of image pairs generated must contain roughly the same number of image pairs under each rating class. The criteria shown in Table 3.3 was used to select the image pairs shown to the raters. 50 image pairs were generated following these criteria.

Table 3.3: Distribution of image pairs in the sample of images generated from the statistics-based approach, computational physics-based approach and the real-world images.

Image pair selection criteria	Percentage of the sample (%)
Real-world images and images generated from the statistics-based approach	30
Images generated from the statistics-based approach and images generated from the computational physics-based approach	70

- Three industry experts annotated 50 such image pairs in total. The details of these raters are provided in Appendix A.

3.5 THE PROCESSING OF CRACK PATTERN IMAGE DATA

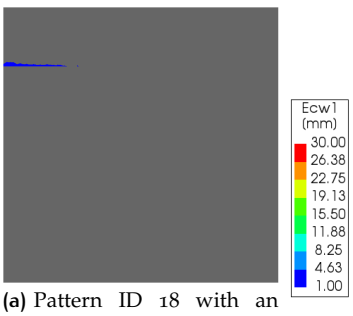
3.5.1 Converting data from the computational physics-based approach to a pixel format

The crack pattern images generated from the statistics-based approach were in a pixel format. They were generated as matrices in Python. Each entry of the matrix represented a pixel. The various entries of these matrices represented a different type of data at the corresponding location on the façade. A detailed description of what each value in the matrix signifies is given in Table 3.4. The results from the computational physics-based approach must be similar and compatible to this pixel format. This ensured that all the annotated crack pattern image pairs could be given as input to the same neural network.

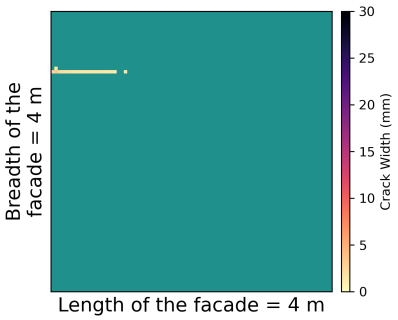
The results from the computational physics-based approach gave the crack width at each point of the cracked façade. To convert this data into a pixel format, a matrix is created such that the number of rows and columns of the arrays correspond to the number of nodes along the length and breadth of the masonry façade. Using the X and Y co-ordinates of each node, the crack width of at that point was fed into to a matrix following the same strategy as explained in Table 3.4. Only crack widths larger than 1 mm were used to complete the matrix. The pixel representation of these arrays were shown in Fig. 3.3.

Table 3.4: Explanation of the numerical values of the matrices used to generated the pixel format of the statistically generated images.

Numerical value	Data it represents at that location
0	Uncracked masonry
-1	Openings in the structure
(0,30]	Crack width of the crack

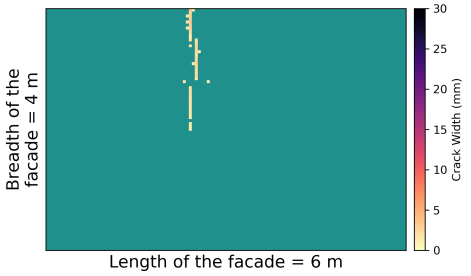


(a) Pattern ID 18 with an aspect ratio 1:1 generated through the computational physics-based approach.

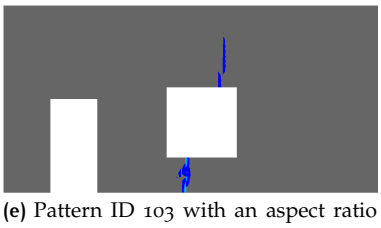


(b) Pixel representation of Pattern ID 18 with an aspect ratio 1:1.

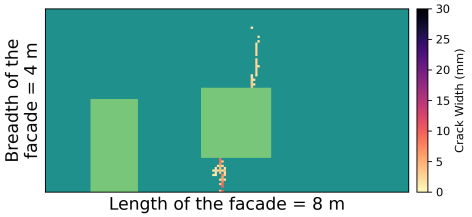
(c) Pattern ID 23 with an aspect ratio 3:2 generated through the computational physics-based approach.



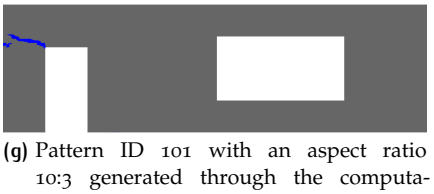
(d) Pixel representation of Pattern ID 23 with an aspect ratio 3:2.



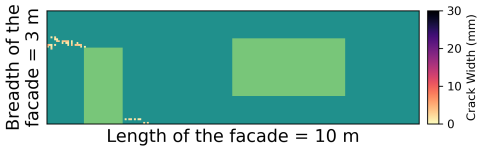
(e) Pattern ID 103 with an aspect ratio 2:1 generated through the computational physics-based approach.



(f) Pixel representation of Pattern ID 103 with an aspect ratio 8:4.



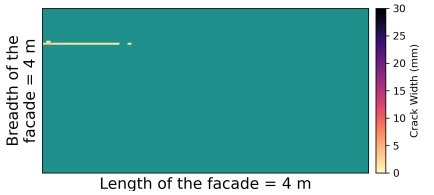
(g) Pattern ID 101 with an aspect ratio 10:3 generated through the computational physics-based approach.



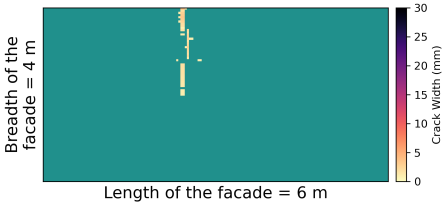
(h) Pixel representation of Pattern ID 101 with an aspect ratio 10:3.

Figure 3.3: Examples of the pixel realisation of the chosen crack types on façades with different dimensions and different orientation of openings using the computational physics-based method.

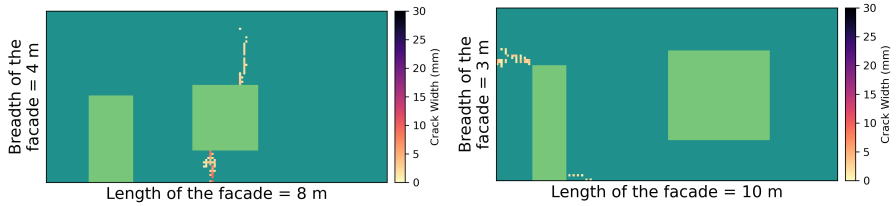
The neural network is able to accommodate only one aspect ratio of the dimensions of these crack pattern images at a time due to the nature of the network architecture and training methods. It is thus only possible to feed the data into the neural network as long as all the façades have the same aspect ratio. As the finite element models are paramterised to give different aspect ratios of the masonry façades, the pixel representations obtained from the finite element models need to be resized to a predefined fixed dimension that is acceptable to the neural network. The images generated from the finite element models contain façades of aspect ratios 1:1, 3:2, 2:1 and 10:3. Therefore, these images were resized to a pixel representation of the façade with an arbitrarily chosen aspect ratio 2:1 (161 pixels by 81 pixels) using the OpenCV library functions with the nearest neighbour interpolation scheme [49]. The examples shown in Fig. 3.3 have been resized to a dimension of 161 pixel by 81 pixels as shown in Fig. 3.4.



(a) Resized pixel representation of Pattern ID 18.



(b) Resized pixel representation of Pattern ID 23.



(c) Resized pixel representation of Pattern ID 103.

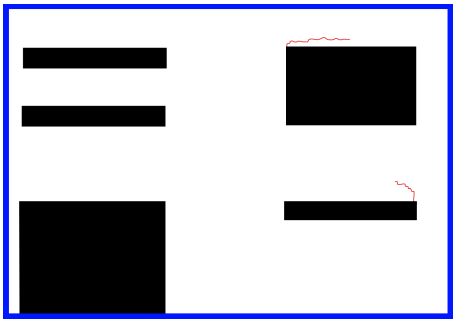
(d) Resized pixel representation of Pattern ID 101.

Figure 3.4: Examples of the pixel representation of the chosen crack types on façades that have been resized to be fed in to the neural network.

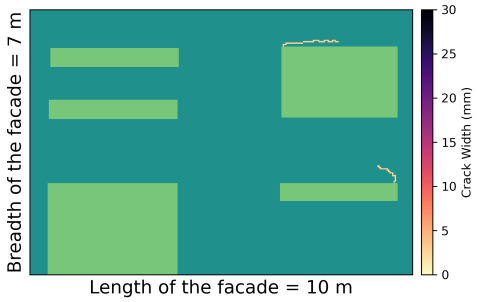
3.5.2 Converting real-world crack pattern to a pixel format

In order to give the real-world images of crack patterns as input to the neural networks, the dimensions of the façades, their openings and their cracks were traced out manually and saved as images, as shown in Fig. 3.5a, Fig. 3.5c, Fig. 3.5e, and Fig. 3.5g. These images and their corresponding RGB values were used to create a pixel representation using the same legend explained in Table 3.4.

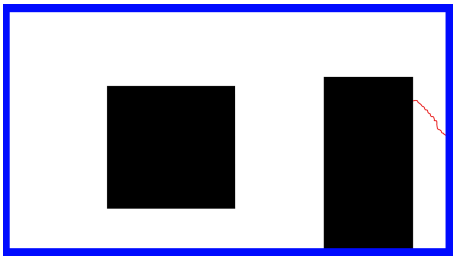
As with the case of the images generated from the computational physics-based approach, the masonry façades with real-world cracks and their pixel representations have varying dimensions. They also needed to be resized to the predefined fixed aspect ratio of 2:1 (161 pixels by 81 pixels). This resizing was achieved with the help of in-built functions from the OpenCV library in python, as explained in Section 3.5.1. The examples shown in Fig. 3.5 were resized to a dimension of 161 pixel by 81 pixels, as shown in Fig. 3.6.



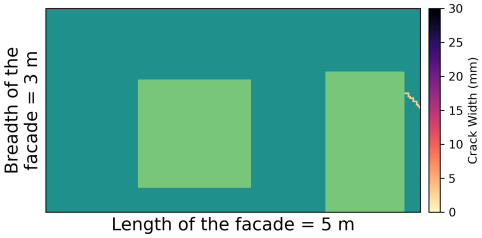
(a) Pictorial representation of Real World Image ID 1.



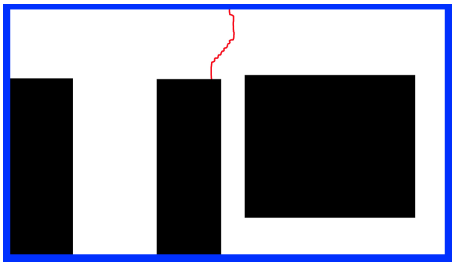
(b) Pixel representation of Real World Image ID 1.



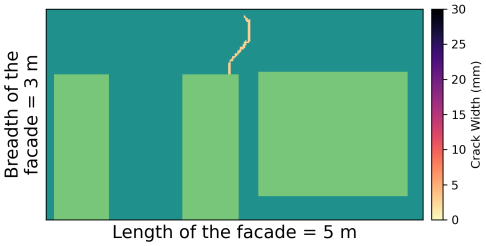
(c) Pictorial representation of Real World Image ID 2.



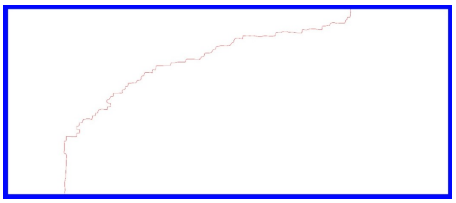
(d) Pixel representation of Real World Image ID 2.



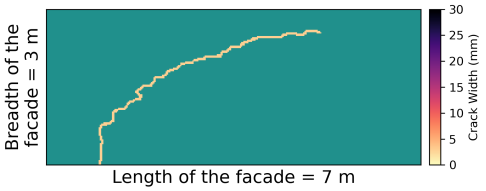
(e) Pictorial representation of Real World Image ID 3.



(f) Pixel representation of Real World Image ID 3.

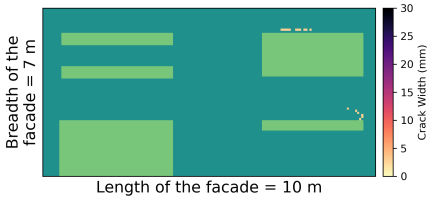


(g) Pictorial representation of Real World Image ID 4.

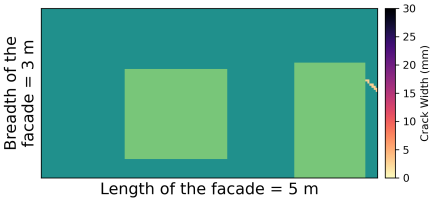


(h) Pixel representation of Real World Image ID 4.

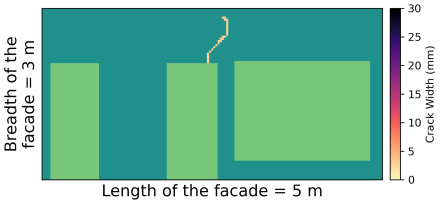
Figure 3.5: Examples of the pixel realisation of the real-world crack pattern images.



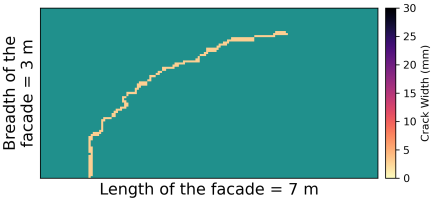
(a) Resized pixel representation of Real World Image ID 1.



(b) Resized pixel representation of Real World Image ID 1.



(c) Resized pixel representation of Real World Image ID 3.



(d) Resized pixel representation of Real World Image ID 4.

Figure 3.6: Pixel representation of chosen real-world images that have been resized to be fed in to the neural network.

3.6 EVALUATING AGREEMENT AMONG THE RATERS

It is useful to study the agreement among the raters before using the crack pattern image pairs to fit the neural networks. This way, it is possible to see if there is a consensus in establishing the ground truth of the neural network. When the raters were given the crack pattern similarity assessment tasks, it is seen that each rater needed some time to decide for themselves how they perceived each label and the corresponding similarity rating. To overcome this, each rater was asked to do a trial assessment on a sample of image pairs before attempting the real task. This helped them to be more confident with the assessment task and to be more consistent with how they themselves perceived the similarity.

Two metrics were used to evaluate the agreement with a standard rater and the overall agreement among all the raters. These are: Lin's Concordance Correlation Coefficient (CCC) and Krippendorff's alpha.

3.6.1 Agreement with a standard rater using Lin's CCC

Lin developed his correlation coefficient to study bio-equivalence [50]. Since then, it has been used in other studies test agreement with a *gold standard* or a *reference standard* [51] [52]. Lin's CCC is calculated by using the distance of a point in a bi-variate sample space from the $y = x$ line as shown in Eq. 3.1.

$$\begin{aligned}\rho_c &= 1 - \frac{\text{Expected orthogonal squared distance from } x = y}{\text{Expected orthogonal squared distance from } x = y \text{ assuming independence}} \\ &= \frac{2\rho\sigma_x\sigma_y}{\sigma_x^2 + \sigma_y^2 + (\mu_x - \mu_y)^2}\end{aligned} \tag{3.1}$$

Where,

ρ = Correlation coefficient between the two variables.
μ_x, μ_y = Mean of the variables
σ_x, σ_y = Variance of the variables.

The Lin's CCC lies in [-1, 1], where a magnitude of -1 indicates perfect disagreement and a magnitude of 1 indicates perfect agreement. If the magnitude of the Lin's CCC is 0, it indicates that there is no correlation between the variables [50]. A perfect agreement with the reference standard would be along the $y = x$ line. Lin noted that the Lin's CCC is robust for sample sizes of 10 and above [50].

Some recommendations for minimum acceptable CCC values are available in the literature. McBirdie recommends a very strict criteria: CCC $\leq$ 0.90 shows poor agreement, 0.95 < CCC $\leq$ 0.99 shows moderate agreement and CCC > 0.99 shows excellent agreement [53]. Another recommendation is to interpret Lin's CCC the same as the Pearson correlation coefficient given as: CCC $\leq$ 0.20 shows poor agreement and CCC > 0.80 shows excellent agreement [54].

3.6.2 Inter-rater agreement

In order to study the degree of agreement between all the raters in a group of raters, the Krippendorff's alpha [55] is a useful metric. The Krippendorff's alpha is applicable to any number of raters, any number of items to be assessed and can account for missing data. This is particularly useful in this thesis, as each rater has annotated a different set of image pairs. While they have assessed image pairs in common with other individual raters, every rater has not assessed every single image pair. This is what is meant by "missing data". Krippendorff's alpha is also useful to look at agreement between categorical and interval data [55]. Here, the ratings are assumed to follow an interval scale, as explained in Table 3.5. Rather than measuring agreement, alpha calculates the disagreement to give a measure of agreement, as shown in Eq. 3.2 [55].

$$\alpha = 1 - \frac{D_o}{D_e} \tag{3.2}$$

Where,

D_o = The observed disagreement.
D_e = The expected disagreement by chance.

Coincidence matrices and difference functions are used to compute D_o and D_e. A coincidence matrix contains the n pairable values from the canonical form of the data assessed by all the raters cross tabulated into a v by v square matrix, where v is the number of items assessed by the raters. The difference function attempts to measure the distance between the individual raters. For an interval scale, it is given by Eq. 3.3 [55].

$$\delta_{\text{interval}}\ (v, v') = (v - v')^2 \tag{3.3}$$

Where,

v, v' = Ratings on an interval scale.

The Krippendorff's Alpha lies in [0, 1], where an α value of 1 signifies perfect agreement and an α value of 0 indicates perfect disagreement [56]. Krippendorff has provided some recommendations for a minimum acceptable alpha value as follows: $\alpha \geq 0.800$ shows that there is good agreement between the raters, $0.800 > \alpha \geq 0.667$ indicates that there is insufficient agreement and the data can only to draw tentative conclusions and $\alpha < 0.667$ shows that the data is unreliable due to significant disagreement [56].

3.7 CONVOLUTIONAL NEURAL NETWORKS FOR CRACK PATTERN SIMILARITY

A convolutional neural network (CNN) was developed by Kleijn [1] such that it can be fitted with crack pattern image pairs with similarity labels. This neural network was expected to be able to learn from this data and predict the similarity of image pairs that it had not seen before. The CNN is developed as a regression neural network, where each image pair's similarity was predicted by categorising them into one of four similarity rating classes: *very dissimilar, dissimilar, similar*, and *very similar*. The similarity ratings were defined on an interval scale, as given in Table 3.5. To each similarity rating class, we assign a numerical value that is the mid-point of the interval class, i.e. *very dissimilar* will have a numerical value of 0.125, *dissimilar* will have a numerical value of 0.375, *similar* will have a numerical value of 0.625 and *very similar* will have a numerical value of 0.875.

The neural network needs to know the ground truth, which is the "real truth" that it needs to learn from and attempt to predict. In this thesis, the ground truth must be established in terms of the similarity label for each image pair. Therefore, for a given image pair, the corresponding ground truth is established by taking the average of all the similarity ratings given by all the raters who evaluated that image

Table 3.5: Converting the categorical ratings to an interval scale for the neural network.

Categorical rating	Interval rating
Very Dissimilar	(0.0, 0.25)
Dissimilar	(0.25, 0.5)
Similar	(0.5, 0.75)
Very Similar	(0.75, 1.0)

pair. The numerical values to establish this ground truth was based on the interval scale provided by Table 3.5. This table also helped to determine what similarity rating class each image pair fell under. The CNN will use the same interval rating system to determine the predicted similarity classes of each image pair as well.

The loss function of this neural network is the mean squared difference between the actual similarity and the estimated similarity. The split for the data to test and fit the the neural network is set to 0.75 (75% for training and 25% for testing). The 25% of the data used to form the testing sample is used to validate the neural network. The image pairs used for fitting the neural network are not used for testing. The results of the similarity assessment are presented using four metrics: accuracy, precision, recall and F-score, which are explained below. These metrics are defined for each individual class using *true positive*, *true negative*, *false positive* and *false negative* of the predicted results. These definitions are explained using the example of an image pair that was annotated as *very similar* by the raters (true value). Thus, the class that represents the similarity rating of *very similar* is taken as the positive class, which makes the classes *very dissimilar*, *dissimilar*, and *similar* the negative classes.

- **True positives**: Number of cases where the similarity rating predicted by the neural network matches that of the test set (true value) that belongs to the positive class. This means that if a rater annotates an image pair with *very similar*, the neural network also predicts the same similarity rating for that image pair.
- **True negative**: For each similarity rating, the true negatives are the number of cases where the predicted value and the true value match (both of which belongs to one of the negative classes). For example, image pairs that were annotated as *very dissimilar*, *dissimilar*, or *similar* by the raters, are predicted as *very dissimilar*, *dissimilar*, or *similar* respectively.
- **False positive**: For each similarity rating, the false positives are the number of cases where the similarity rating is incorrectly predicted as belonging to the positive class when it does not. For example, image pairs that the raters assessed as *very dissimilar*, *dissimilar*, or *similar* are wrongly predicted by the neural network as *very similar*.
- **False negative**: For each similarity rating, the false negatives are the number of cases where the similarity rating is incorrectly predicted as belonging to one of the negative classes when it actually belongs to the positive class. For example, for an image pair that the rater annotates as *very similar*, the neural network predicts a rating of *very dissimilar*, *dissimilar*, or *similar*.

These metrics, typically used to determine the performance of classification neural networks, are chosen to assess the regression neural network as they work well even for regression neural networks. In the previous work done at TNO (Section 2.5.2), the accuracy of the regression neural network was used to measure the network performance. Hence, determining the accuracy of the neural network in this thesis makes possible a comparison between the performance of these two neural networks.

Each set of data fed into the neural network is run five times to determine the mean and standard deviation of each of these metrics. If the obtained standard deviation is greater than 5%, the networks are run 10 times for each label. The accuracy, precision, recall and F-score assesses the performance of the fitted neural networks by class. To evaluate the performance of each fitted neural network by label, these metrics are averaged out among all the classes over all the runs of the fitted neural networks.

Accuracy

Accuracy is the ratio of correctly predicted observations to the total observations for each similarity rating, as given in Eq. 3.4. In supervised learning, the accuracy of the neural network is largely dependent on the data used to fit the neural network. A representative training data set is critical, both in terms of sample size and quality, affect the accuracy of neural network's performance [57].

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \tag{3.4}$$

Precision

Precision is the ratio of correctly predicted positive observations to the total predicted positive observations for each similarity rating, as shown in Eq. 3.5. This metric helps us understand, of all the predicted positive classes, how many are true. A low false positive rate is related to high precision values.

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \tag{3.5}$$

Recall

Recall is the ratio of correctly predicted positive observations to all observations in the actual class computed for each similarity rating. This is shown in Eq. 3.6. Recall helps evaluate, of all true classes in the test set, how many are predicted. Often, recall is referred to as the counterpart of precision.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \tag{3.6}$$

F score

F score is less intuitive than the other metrics. It is the harmonic mean of precision and recall calculated for each similarity rating. Therefore, this score takes both false positives and false negatives into account. This is explained in 3.7. For problems with unbalanced data, the F score may be more appropriate than the Mean Square Error (MSE) or other error metrics [58].

$$\text{F score} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{3.7}$$

Although using accuracy as a defining performance metric for a neural network makes intuitive sense, it is usually helpful to include precision, recall and F score as well. It is possible that even with a high accuracy, the precision or recall is poor. Although it is aimed to achieve a high precision and recall value, it is not always possible achieve both at the same time. For example, if the neural network is modified to give a high recall, it is likely true that we will be able to correctly predict which image pairs fall under each similarity class. However, this will also lead to a large number of image pairs that do not belong to a certain similarity class being put there as well. Similarly, if a high precision is aimed to avoid putting image pairs under the wrong similarity class, likely a large number of image pairs will be put in other classes. Instead of trying to find a balance between a good precision value and good recall, it is helpful to aim for a good F score, which will indicate good precision and recall. It only reaches its ideal 100% if both precision and recall are 100%. And if one of these equals 0%, F score will have the worst value of 0%.

4 FINITE ELEMENT MODELS BASED ON THE CRACK PATTERNS

4.1 INTRODUCTION

A computational physics-based approach was proposed as the next step in the data generation process to fit a neural network that aims to perform similarity assessments on crack pattern image pairs. This approach was expected to give more realistic results than the previously used statistics-based approach. The crack pattern image pairs generated in the computational physics-based approach needed to be converted into a pixel representation in order to be compatible with the neural network, as explained in Section 3.3. The neural network trained on the statistically generated images was fitted to 12 different crack patterns with varying characteristics, as explained in Section 2.5. In this thesis, 8 out of these 12 crack patterns were simulated in the computational physics-based approach, with each crack pattern type requiring unique loading and boundary conditions. The aim of the finite element analysis was to reproduce similar crack patterns from these crack pattern archetypes using appropriate finite element software, paying special attention to the boundary conditions and the loads. This had not been explored in detail previously in the statistics-based approach, or in de Vent's research [7]. Sometimes, unrealistic boundary conditions and loads had to be used to simulate these chosen crack patterns. The general properties of these finite element models and their analyses are described in this chapter. The generation of these crack patterns was achieved using the finite element software DIANA FEA 10.4. Additionally, this chapter also explains how the parametrisation of these models was achieved by varying the geometry, material properties and the crack-inducing loads in each crack pattern type. Finally, an overview of the sensitivity study on the finite element models, and the limitations of the automated crack pattern generation using the finite element method is also given.

4.2 THE FINITE ELEMENT MODELS

4.2.1 Geometry

Two types of geometries of the masonry façades are considered in this project, as explained in Section 2.5. These are façades without openings and façades with openings (e.g., doors and windows). The façades are modelled as rectangular sheets of varying aspect ratios. In the classification neural network fitted to statistically generated images, the aspect ratios of the façades considered were 3:2 for façades without openings and 10:3 for façades with openings. In order to explore the possibility to generalise the neural networks and to increase the range of input façades that can be tested, two more aspect ratios are selected: 1:1 and 2:1. Therefore, the following rectangular façade dimensions were considered for façades with and without openings: 4 x 4 m, 4 x 6 m, 4 x 8 m and 10 x 3 m. The thicknesses of these façades were chosen from a range of values typically seen for masonry structures of this size, i.e. it is varied with 0.2 m [59, 60], 0.3 m [16], 0.4 m [61] and 0.5 m [39].

In the façades with openings, three types of openings were considered: a door of size 1 m x 2 m (denoted as Door 1), a window of size 3 m x 1.5 m (denoted as

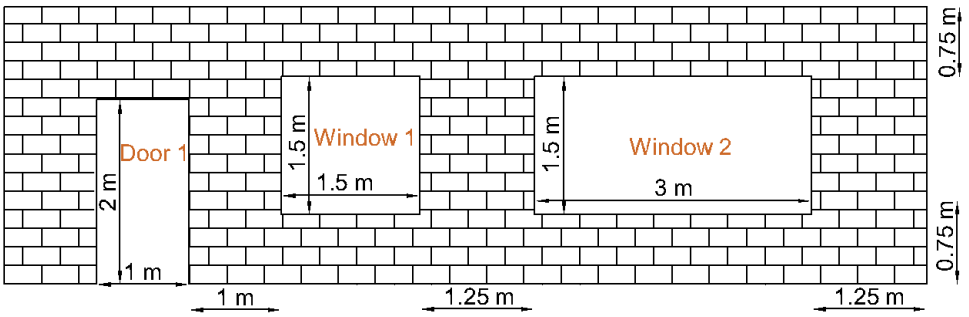


Figure 4.1: Example of the geometry of a façade of dimensions 10 m x 3 m with openings.

Window 1) and another window of size 1.5 m x 1.5 m (denoted as Window 2)[6]. These openings were used in different combinations and at different positions to reflect the various façades seen in real life. An example of these openings on a façade of size 10 m x 3 m is shown in Fig. 4.1. All the openings are at least 1 m away from each other and the edges. The soil-structure behaviour of the masonry façade and the ground was simulated using a boundary interface. The length and thickness of this interface will be the same as that of the façade.

4.2.2 Material properties

Masonry

The masonry was assumed to consist of clay masonry bricks and mortar. It was modelled as a continuum with homogenised properties using the total strain based smeared fixed cracked model in DIANA FEA 10.4. The smeared cracked model spreads the crack over a certain area. The advantage of using a smeared cracked model is that the cracks can initiate anywhere and propagate in any direction. The density of the masonry was assumed to be 1900 kg/m^3 [39] and a constant Poisson's ratio of 0.2 was adopted [39]. A linear softening curve is adopted for the tensile strength with the fracture energy equal to 10 N/m. The behaviour under compression was assumed to be linear-elastic. This is a reasonable assumption as most of the crack-inducing loads are tensile forces, not compressive.

In reality, the masonry is a heterogeneous material with non-uniform values for the various properties along the façade [62]. The structural performance and reliability of unreinforced masonry walls may be greatly influenced by spatial variability in material qualities [63]. Most of the finite element software that deals with concrete, masonry and steel treat their various material parameters as deterministic [64]. There are different mathematical approaches available in the literature to define a random field to model the spacial variability of the material properties of concrete. These random fields make use of different probabilistic distributions for sampling the data points [64]. If the masonry is modelled as a deterministic continuum, there is a potential loss of the uncertainties that exist in real-world structures. The uncertainties can be related to material properties, such as stiffness of the brick and mortar joints, the compressive and tensile strengths of the bricks and the Young's modulus, or geometric properties, such as thickness, texture or layout. Considering the discrete bricks and mortar as a continuum in the numerical model is a process of deterministic homogenisation. This can be extended to a stochastic homogenisation by considering the variability in the geometric variability [65].

For implementing the random field for material properties of masonry structures, no readily available framework could be found in the literature. Hence, the simple methodology explained in Appendix B is adopted. The mean values and coeffi-

Table 4.1: Properties of the random distributions used to generate the random field for the spatial variation of Young's modulus and tensile strength in masonry [39, 65–71].

Material property	Distribution	Mean (N/m^2)	CoV(%)	Standard deviation (N/m^2)
Young's Modulus	Normal	3×10^9	20	0.6×10^9
Tensile Strength	Log-normal	0.1×10^6	35	0.035×10^6

cients of variations (CoVs) of the Young's modulus and tensile strength used to generate these random fields are given in Table 4.1.

Soil-structure interface

The nature of settlement-induced damage on a structure is dependent on the nature and properties of the soil beneath it. Here, the soil-structure interaction was modelled as a nonlinear interface, allowing for the possibility to vary the stiffness along this interface. Generally, structures with a stiffer soil-structure interface tend to suffer more subsidence induced damages than those sitting on a softer soil-structure interface when subjected to the same settlement profile [39]. This is because softer soil-structure interfaces allow for a better redistribution of the settlement stresses. A soil-structure interface also allows for more realistic modelling of the interaction between the soil and the façade when compared to applying a settlement load directly to the bottom edge of the structure. The soil deformations are influenced by the mass and the stiffness of the structure, resulting in a lower vertical displacement and horizontal strain. Thus, this lower subsidence load will also cause reduced subsidence-induced stresses on the façade [72].

The properties of the interface elements were chosen from the literature [39, 72], and by following the general guidelines described by DIANA FEA [73]. The stiffness properties of the interface were also chosen such that the compressive normal stiffness was large enough to avoid overlap of the interface elements and the masonry façade elements. The shear stiffness of the soil structure interface was assumed to be equal in magnitude for both tangential directions of the interface. This is calculated using Eq. 4.1.

$$D_{tt} = D_{ss} = \frac{A^2}{t} \cdot \frac{E_{soil}}{2 \cdot (1 + \nu_{soil})} \tag{4.1}$$

Where,

A = A reduction factor that reflects the flexibility of the soil structure interface in shear when compared to the surrounding soil medium. It may vary from 0.5 to 1.0. In this thesis, the value of A was chosen as 1.
t = Thickness of the interface layer normal to the plane of the interface. Here, it will be the same as the thickness of the façade.
E_{soil} = Young's modulus of the soil.
ν_{soil} = Poisson's ratio of the soil.

The soil below the structure was assumed to be silty sand. The corresponding Young's modulus and Poisson's ratio of the soil are given in Table 4.2.

The shear stiffness values thus obtained are used to compute the normal stiffness of the interface elements as per Eq. 4.2.

Table 4.2: Soil properties used to model the soil-structure interface [72].

Material property	Magnitude
Young's Modulus	$5 \times 10^7\ N/m^2$
Poisson's ratio	0.3

Table 4.3: The normal and shear stiffness values of the soil-structure interface.

Thickness of the façade (m)	D_{nn} (N/m^3)	$D_{tt}(N/m^3)$	D_{ss} (N/m^3)
0.2	8.13×10^{10}	8.13×10^{8}	$8.13e\times10^{8}$
0.3	3.61×10^{10}	3.61×10^{8}	3.61×10^{8}
0.4	2.03×10^{10}	2.03×10^{8}	2.03×10^{8}
0.5	1.30×10^{10}	1.30×10^{8}	1.30×10^{8}

$$D_{\mathrm{nn}} = f \times D_{\mathrm{tt}} \tag{4.2}$$

Where,

D_{nn} = Normal stiffness offered by the soil-structure interface.
f = A multiplication factor that varies from 10 to 100. In this thesis, the value of f was chosen as 100.

Thus, the magnitudes of the normal and shear stiffness used depends on the thickness of the structure, and is shown in Table 4.3. The nonlinearity of the soil-structure interface is realised by incorporating a no-tension interface into the soil-structure interface. The no-tension interface works by reducing the stiffness of the interface elements by a factor when the relative displacement in that direction is greater than a pre-defined critical value. The critical value is given as zero in the normal and shear directions for the soil-structure interface under the façades. As a result, the no-tension interface represents the soil-structure interface's low resistance to tension.

4.2.3 Element types and finite element mesh

The finite elements were chosen such that out-of-plane loading could be applied, if necessary. Quadratic elements with linear interpolation were used to generate the elements of the interface and that of the masonry façade.

Masonry

The masonry was modelled using regular curved shell finite elements in order to facilitate out-of-plane loading wherever necessary, as shown in Fig. 4.2. This resulted in the use of the 8-noded quadrilateral element CQ40S and the 6-noded triangular element CT30S for the masonry façade, as shown in Fig. 4.3.

Soil-structure interface

The soil-structure interface was modelled as a structural plane interface using a curved line. This resulted in the use of the 6-noded curved line interface element CL24I as shown in Fig. 4.4

Element Size

Element sizes of 100 mm x 100 mm were chosen to generate the finite element models of the masonry façade. This size was chosen after conducting a sensitivity

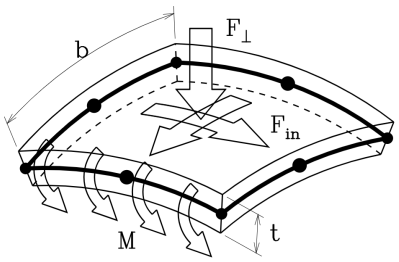
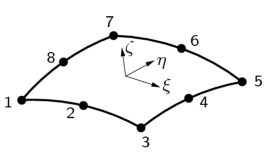
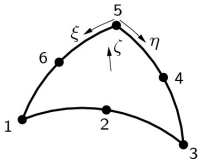


Figure 4.2: Characteristics of regular curved shell elements used in DIANA FEA 10.4 [74].



(a) Finite element CQ40S.



(b) Finite element CT30S.

Figure 4.3: The finite element types used to generate the mesh of the masonry façade in DIANA FEA 10.4 [75].

study by varying the mesh size, as explained in Section 4.4. The quality of the cracks and the computation times were analysed before choosing the optimal mesh size. It was seen that there is no significant difference in the quality of the crack patterns when using a mesh of 50 mm and 100 mm. As hundreds of crack patterns were to be generated, a mesh size of 100 mm was fixed for all the models.

4.2.4 General loading and boundary conditions

There were three types of loads applied onto the masonry façades in this thesis: the self-weight, the pre-compression load and the crack-inducing load. These are explained in detail below.

- **Self-weight:** The self-weight of the masonry wall was applied in the vertical direction. The magnitude of this load was computed from the volume of the façade and the density of the masonry.
- **Pre-compression load:** This load simulated the load that comes on the structures from the dead and live loads on the rest of the building. A pre-compression pressure of 0.5 N/m^2 is assumed to act on the wall before any crack-inducing load takes place [77]. It was applied on the top horizontal edge as a line load whose magnitude was obtained by multiplying 0.5 N/m^2 with the thickness of the façade, as shown in Fig. 4.5.

  The pre-compression load was only applied on the façades with openings. The façades without openings were assumed to be non-load bearing external walls.

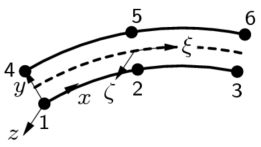


Figure 4.4: Finite element CL24I used in DIANA FEA 10.4 [76].

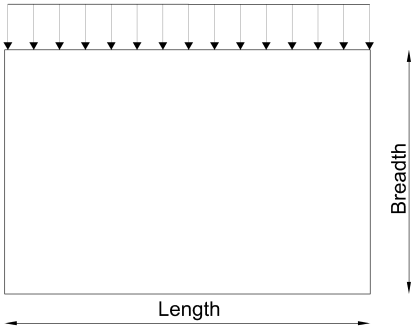


Figure 4.5: Schematic representation of the pre-compression load being applied on a masonry façade.

- **Crack-inducing load:** The aim of applying this load was to obtain the required crack pattern with a sufficiently large crack width. The crack patterns considered in this thesis were assumed to be caused by settlement loads, temperature loads, or external out-of-plane loads. The crack-inducing load was applied either as a force control or as a displacement control, depending on the crack pattern that needs to be simulated. The type of the load, its location, and its magnitude changes with each Pattern ID.

  The magnitude of the load to be applied on each structure was chosen such that it led to significant crack widths. Sometimes, when the magnitude of the load is too large, large stresses develop elsewhere causing cracks in other parts of the façade. These new cracks will in turn cause a redistribution of the stresses and could reduce the crack width of the initial crack. Therefore, the maximum magnitude of the load was applied such that stresses in the façade are not large enough to cause cracking elsewhere.

Additionally, two main types of supports are applied to all of the masonry façades:

- **Soil support**: The soil support was the support offered by the soil-structure interface at the bottom edge of the façade. In general, the soil support offers restraint in the vertical and horizontal directions. Additional translations or rotational supports were provided, if required.
- **Support for the crack-inducing load**: The support for the crack-inducing load was provided in order to initiate the crack pattern. This varies with each Pattern ID.

4.2.5 Analysis

All of the masonry façades were analysed using a phased nonlinear structural analysis with a Newton Raphson/Tangential iteration scheme and a linear search. The arc length method is applied on the load steps wherever possible to account for snap-back behaviour. Both physical and geometrical nonlinearities are considered in the analyses.

The loads are applied in two phases as given below:

- **Phase 1**: The self-weight and the pre-compression load were applied in this phase. They were both applied in one step each. Only the soil support was active in this phase, as shown in Table 4.4.
- **Phase 2**: The support of the crack-inducing load was activated in this phase. It was also assumed that since the construction of the masonry structures,

Table 4.4: Properties of Phase 1 of the nonlinear analysis of the masonry façades.

Load	Iteration Scheme	Number of steps	Maximum number of iterations allowed	Active supports
Self weight	NR/ Tangential	1	10	Soil support
Pre-compression load	NR/Tangential	1	10	Soil support

Table 4.5: Properties of Phase 2 of the nonlinear analysis of the masonry façades.

Load	Iteration Scheme	Number of steps	Maximum number of iterations allowed	Active supports
Crack-inducing load	NR/ Tangential	200	50	Soil support, Support for the crack-inducing load

a significant amount of time has passed before the crack inducing load occurs. Thus, while the stresses due to the self-weight and the pre-compression load still exist in the structures, the displacements can be assumed to be zero. Hence, the rotational and translational displacements were reset to zero before applying the cracking load in this phase. The crack-inducing load was applied in 200 steps.

The final output required from the finite element analysis is the various crack patterns and their crack widths in the principal direction. The principal crack width is calculated as the product of the crack strain in the principal direction and the crack bandwidth h_{cr}. As 2D elements with quadratic interpolation were used in this thesis, the crack bandwidth, h_{cr}, is calculated using the Rot's element-based method [34] as $\sqrt{A}$, where A is the total area of the finite element. The details of the elements, the locations of their nodes and the principal crack-widths at the corresponding locations are exported and saved to generate the pixel representations of these crack patterns.

4.3 CRACK PATTERN RESULTS

4.3.1 Pattern ID 18

The crack pattern shown in Fig. 4.6 represents Pattern ID 18. It shows a crack with a horizontal orientation. These types of cracks tend to initiate near the corner or the column [7] and are usually seen when masonry façades are subjected to dimensional changes induced by moisture or temperature. For example, if a portion of the façade experiences direct sunlight, while the rest of the façade experiences shade, it can lead to a nonuniform expansion along the façade [7].

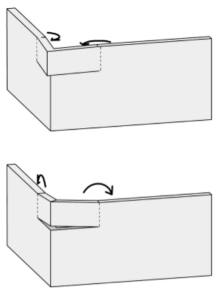


Figure 4.6: Pattern ID 18 [7].

Boundary Conditions and loading condition

As seen in Fig. 4.7, the bottom edge of the façade was fixed in all directions for translation (X, Y, Z). This was the only boundary condition that was active while the façade was being loaded with the self-weight and the pre-compression load. In order to simulate a horizontal crack near the corner or the column of the façade, a line was imprinted onto the surface of the façade. The deformation load was applied along this imprinted line. To take this load and generate the required crack pattern, the imprinted horizontal line was fixed in the vertical (Y) direction as shown in Fig. 4.7. Additionally, the top edge was restrained in the horizontal (X) direction. As the crack widens when the applied load is increased, the stresses in the part of the façade directly above the crack start increasing. This can cause the façade to *tear* away as shown in Fig. 4.9 similar to Fig. 4.6. The horizontal restraint on the top of façade helped the façade achieve a wider horizontal crack. This was necessary to obtain a larger variety of the crack widths that can be fed into the neural networks.

The general loading conditions were explained in Section 4.2.4. The crack inducing-load used to produce this type of crack pattern is shown in Fig. 4.8. The crack-inducing load is applied as a constant load using a displacement control applied along the horizontally imprinted line. This prescribed deformation is supported with a boundary support as explained above. In order to parametrise this crack type, the crack initiation point and the crack length is varied in each model.

The crack initiation point was randomly chosen such that it occurs in the top 70 - 80 % of the height of the façade. Any cracks that initiate closer to the top edge (i.e cracks that initiate at 80 - 100% of the height of the façade) tore away before significant crack width could be achieved. Attempts to initiate the crack width at less that

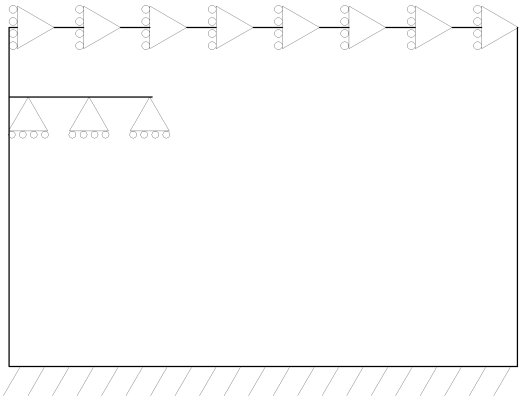


Figure 4.7: Schematic representation of the boundary conditions of Pattern ID 18.

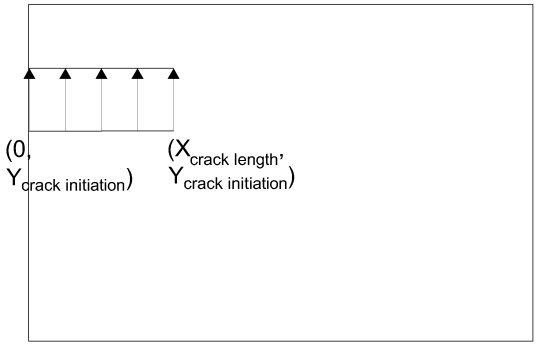


Figure 4.8: Schematic representation of the crack-inducing load given to Pattern ID 18.

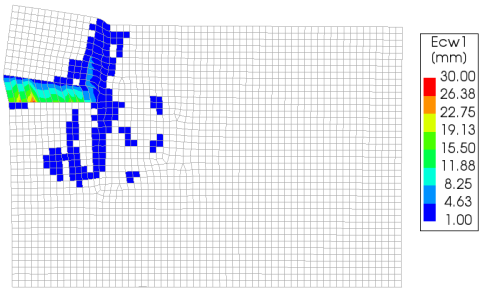


Figure 4.9: Tearing away of the corner of the façade in some cases of Pattern ID 18.

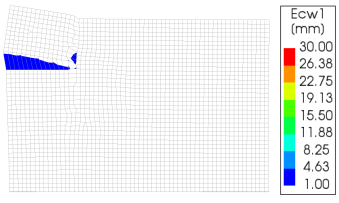
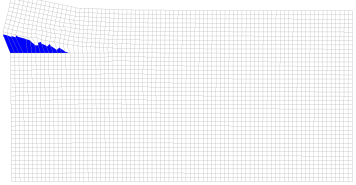
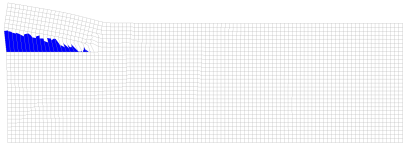
70% of the height of the façade failed as extremely large forces were required to produce even small cracks. The length of the imprinted line that produces this crack is randomly chosen between 1 and 2 m. Smaller crack lengths are not long enough to be comparable to the crack lengths obtained from the statistically generated images. As the crack widens and the portion of the façade above the crack pattern bends upwards, stresses develop causing the crack to tear away. This happens much more easily when the crack lengths are greater than 2 m, causing cracks that are different from the expected crack patterns.

The magnitude of the prescribed deformation load applied on the horizontally imprinted line is randomly chosen between 0.1 mm and 0.5 mm. This magnitude was enough to produce cracks with widths between 1 and 5 mm. Larger loads led to tearing away of the façade corner. Sometimes, larger loads also caused numerical issues with the modelling and led to difficulties in achieving equilibrium in the finite element models.

Crack patterns

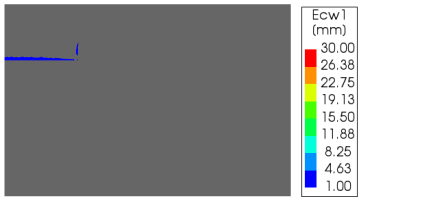
A nonlinear phased analysis of each of the models was completed using the procedure explained in Section 4.2.5. Examples of the normalised (or scaled) deformed façades with the crack pattern at the last step of the finite element analysis are given in Table 4.6.

Table 4.6: The deformed shape of the finite element model of Pattern ID 18 at the last step (deformation factor = 0.05).

Pattern ID 18		
	Aspect ratio 1:1.	 Aspect ratio 3:2.
	 Aspect ratio 2:1.	 Aspect ratio 10:3.

The absolute deformation of the cracked façades (where the deformation is shown relative to the façade dimensions), as shown to the masonry experts for the similarity assessment, are shown below in Table 4.7.

Table 4.7: The contour plot of the crack pattern of Pattern ID 18 in a masonry façade of various aspect ratios.

Pattern ID 18	 Aspect ratio 1:1.	 Aspect ratio 3:2.
	 Aspect ratio 2:1.	 Aspect ratio 10:3.

4.3.2 Pattern IDs 20, 21, 23, 24, 101, 102 and 103

The boundary and loading conditions used to simulate Pattern IDs 20, 21, 23, 24, 101, 102 and 103 are explained in table 4.8. A detailed description of these and the finite element models are provided in Appendix C.

Table 4.8: Summary of the boundary and loading conditions used to simulate the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103.

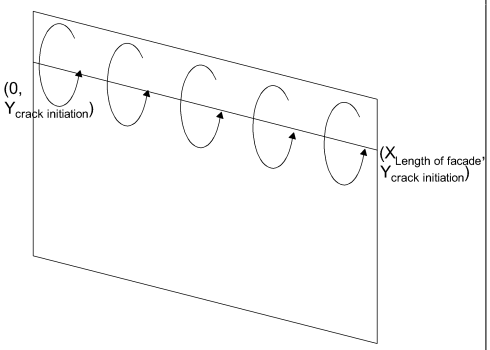
Pattern ID 20	 Boundary condition.	 Loading condition.

Table 4.8: Summary of the boundary and loading conditions used to simulate the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103.

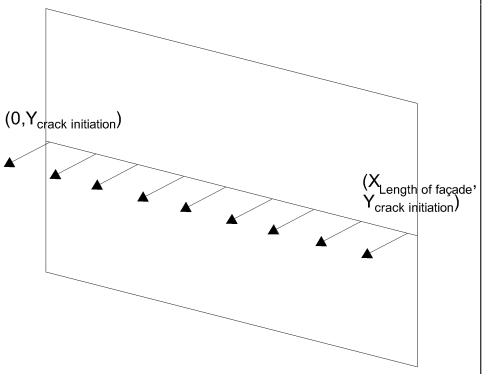
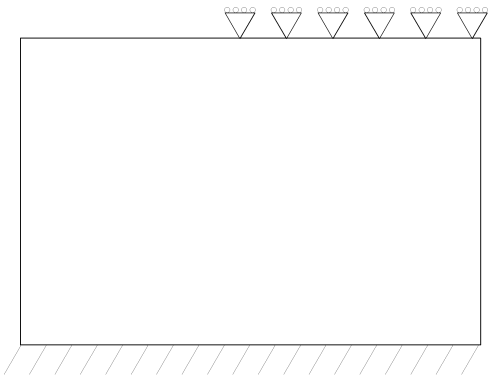
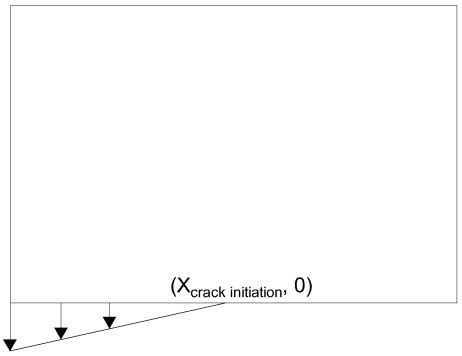
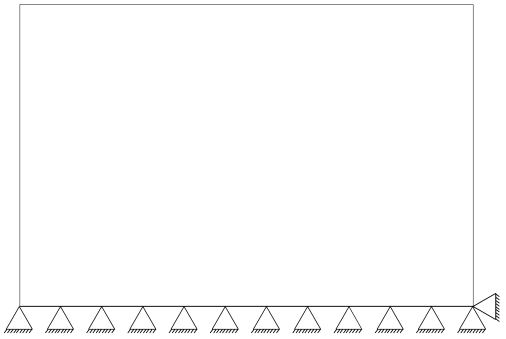
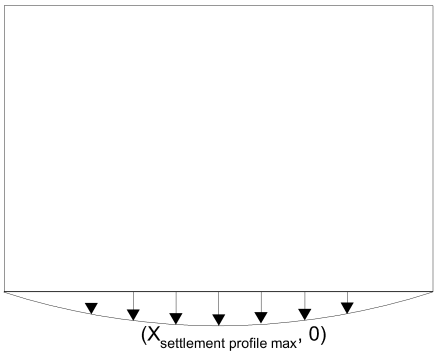
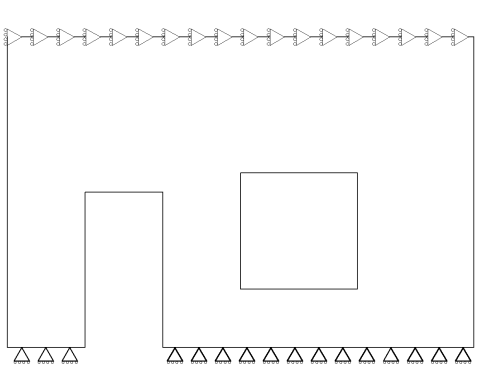
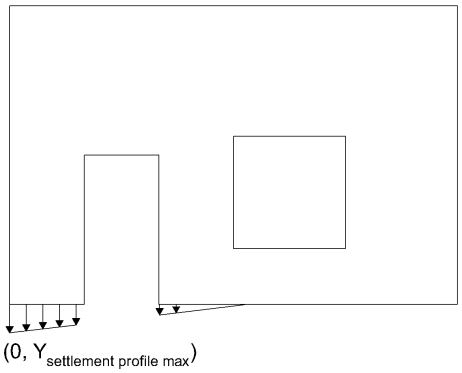
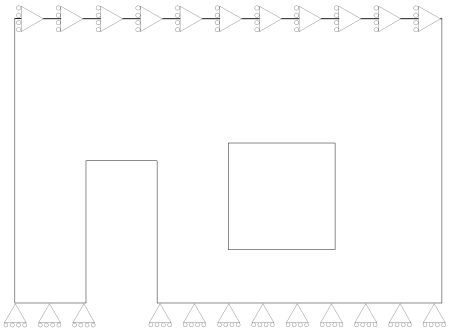
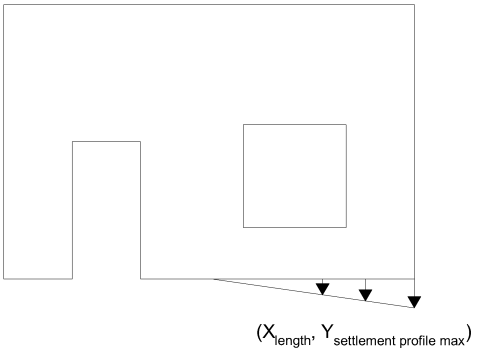
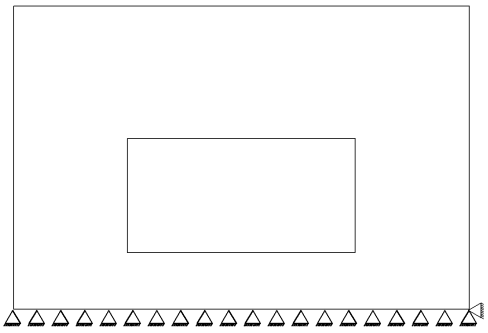
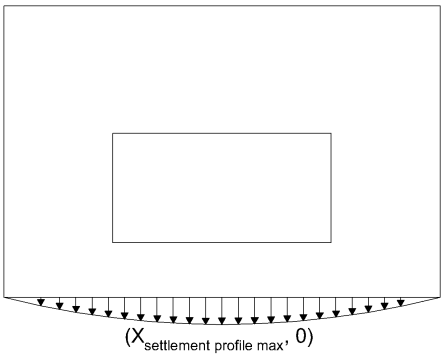
Pattern ID 21	 Boundary condition.	 Loading condition.
Pattern ID 23	 Boundary condition.	 Loading condition.
Pattern ID 24	 Boundary condition.	 Loading condition.

Table 4.8: Summary of the boundary and loading conditions used to simulate the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103.

Pattern ID 101	 Boundary condition.	 Loading condition.
Pattern ID 102	 Boundary condition.	 Loading condition.
Pattern ID 103	 Boundary condition.	 Loading condition.

A summary of the crack patterns obtained by simulating these remaining Pattern IDs is provided in Table 4.9. These are the crack patterns used to generate the image pairs which were shown to the raters for the similarity assessments.

Table 4.9: Summary of the crack patterns simulated for the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103 shown as contour plot on masonry façade of various aspect ratios.

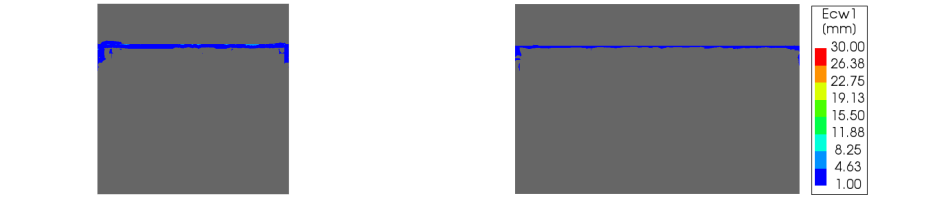
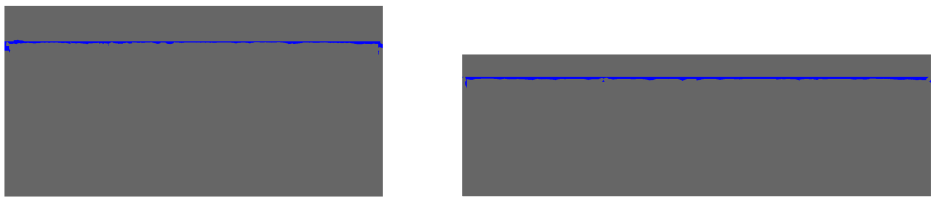
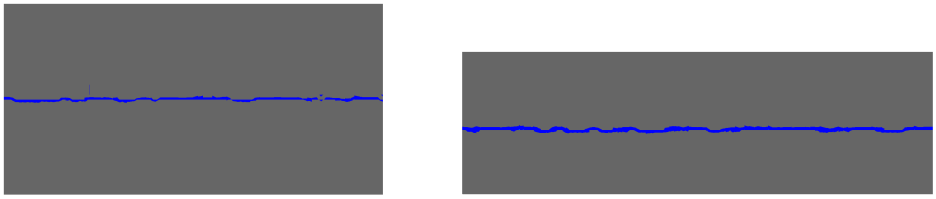
Pattern ID 20	 Aspect ratio 1:1. Aspect ratio 3:2.  Aspect ratio 2:1. Aspect ratio 10:3.
Pattern ID 21	 Aspect ratio 1:1. Aspect ratio 3:2.  Aspect ratio 2:1. Aspect ratio 10:3.
Pattern ID 23	 Aspect ratio 1:1. Aspect ratio 3:2.  Aspect ratio 2:1. Aspect ratio 10:3.

Table 4.9: Summary of the crack patterns simulated for the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103 shown as contour plot on masonry façade of various aspect ratios.

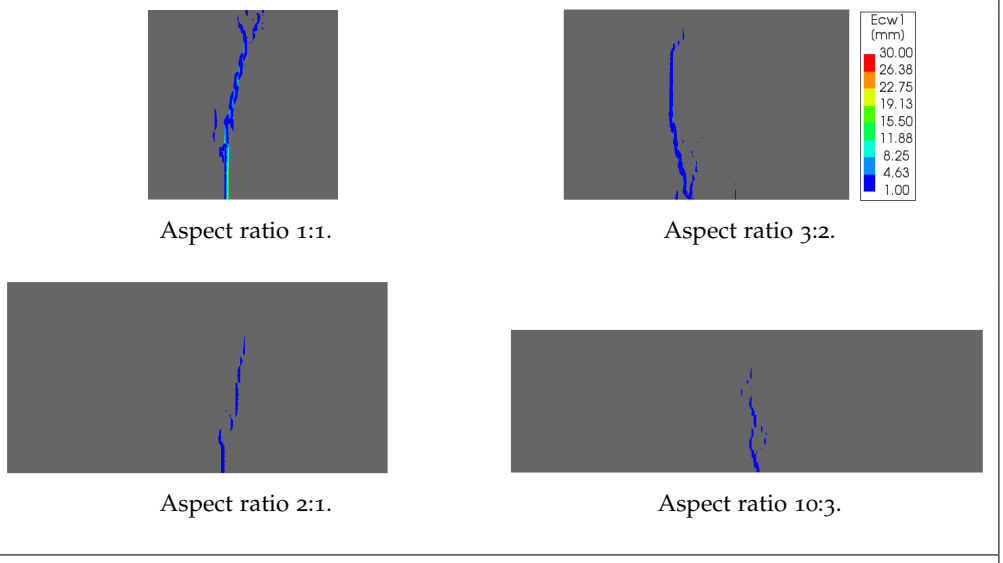
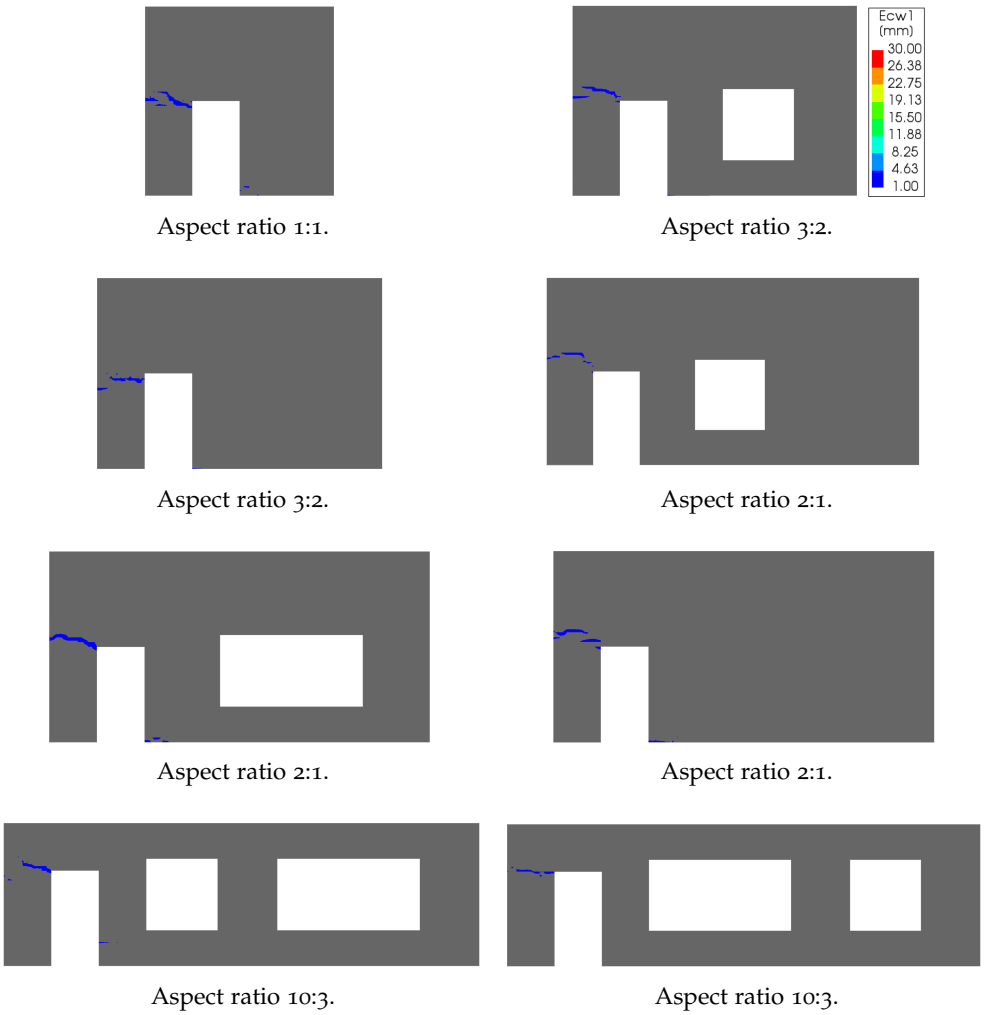
Pattern ID 24	 Aspect ratio 1:1. Aspect ratio 3:2. Aspect ratio 2:1. Aspect ratio 10:3.
Pattern ID 101	 Aspect ratio 1:1. Aspect ratio 3:2. Aspect ratio 3:2. Aspect ratio 2:1. Aspect ratio 2:1. Aspect ratio 2:1. Aspect ratio 10:3. Aspect ratio 10:3.

Table 4.9: Summary of the crack patterns simulated for the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103.

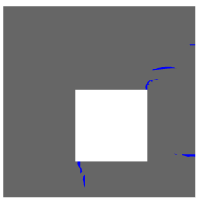
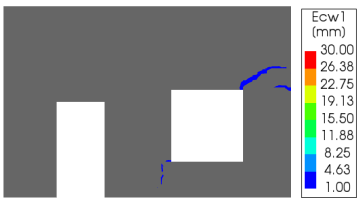
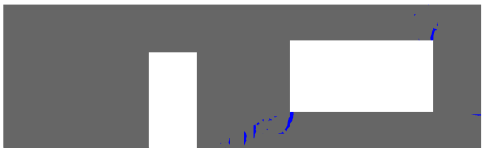
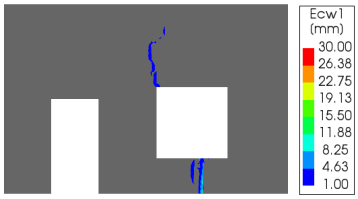
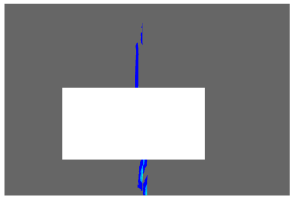
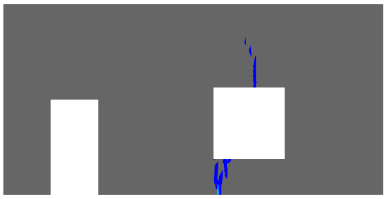
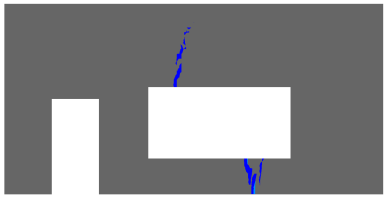
Pattern ID 101	 Aspect ratio 10:3.  Aspect ratio 10:3.
Pattern ID 102	 Aspect ratio 1:1.  Aspect ratio 3:2.  Aspect ratio 3:2.  Aspect ratio 2:1.  Aspect ratio 2:1.  Aspect ratio 2:1.  Aspect ratio 10:3.  Aspect ratio 10:3.  Aspect ratio 10:3.  Aspect ratio 10:3.

Table 4.9: Summary of the crack patterns simulated for the Pattern IDs ID 20, 21, 23, 24, 101, 102 and 103.

Pattern ID 103	
 Aspect ratio 1:1.	 Aspect ratio 3:2.
 Aspect ratio 3:2.	 Aspect ratio 2:1.
 Aspect ratio 2:1.	 Aspect ratio 2:1.
 Aspect ratio 10:3.	 Aspect ratio 10:3.
 Aspect ratio 10:3.	 Aspect ratio 10:3.

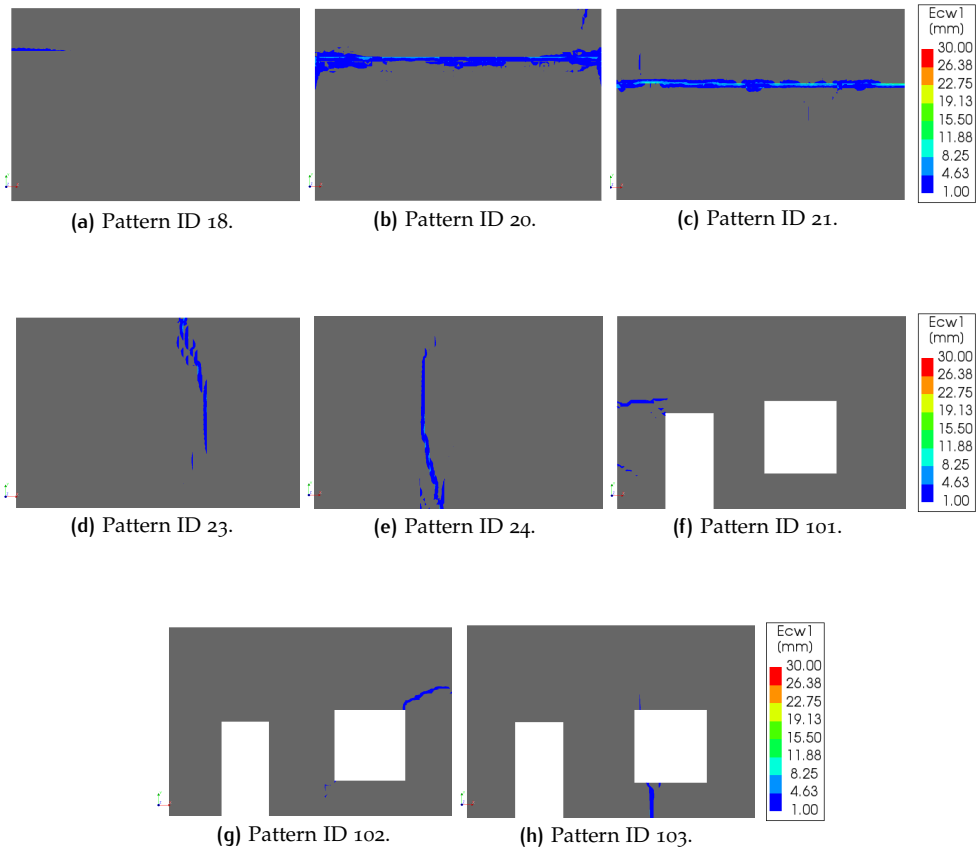
4.4 SENSITIVITY STUDY

A sensitivity study was conducted on chosen finite element models to explore the effect of the mesh size, soil properties and initial loading conditions on the behaviour of the masonry façades and crack development.

4.4.1 Reference cases

The reference case taken for each Pattern ID which shows the crack width at the 200^{th} step are shown in Fig.4.10. These reference cases were generated for masonry

façades with a thickness of 0.2 m and and aspect ratio of 3:2. The mesh size of the each finite element model is 100 mm. The material properties of the masonry and the soil were as explained in Section 4.2.2. The loading condition of each model was explained in Section 4.3.1 and Appendix C.



(a) Pattern ID 18. (b) Pattern ID 20. (c) Pattern ID 21.

(d) Pattern ID 23. (e) Pattern ID 24. (f) Pattern ID 101.

(g) Pattern ID 102. (h) Pattern ID 103.

Figure 4.10: Reference model of each Pattern ID

4.4.2 Element size variation

This subsection explores the mesh sensitivity of the models. A mesh sensitivity analysis is useful to verify that the structural response of the masonry façades are independent of mesh size. The ideal mesh size is a trade-off between accuracy and computing time. The optimum mesh size is chosen such that the use of smaller mesh sizes have no discernible effect on the performance of the finite element models. For this purpose, a finer and coarser mesh than the reference case was used in this section. The mesh sizes chosen were 200 mm and 50 mm. The results were compared with the reference cases generated with a mesh size of 100 mm, as seen in Fig. 4.10.

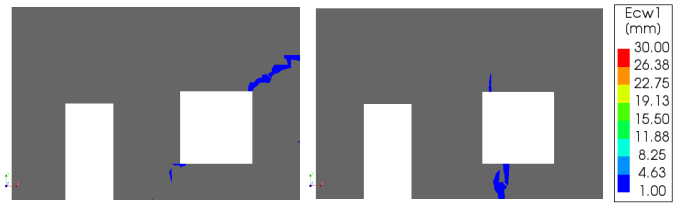
Mesh size of 200 mm

The crack width was computed at each node in the finite element model. With a mesh size of 200 mm, the nodes were farther apart from each other, giving the impression that the cracks are spread over a larger area. The crack widths at the 200^{th} step for all the Pattern IDs except Patter ID 20 can be seen in Fig. 4.11. It was seen that Pattern ID 20 was not able to attain convergence, and it diverged at the 97^{th} step after attaining a crack width of 11.28 mm. The crack width at the 96^{th} step was shown in Fig. 4.11b. Facade ID 20 had 11 non-convergence points after 50 iterations when using a mesh size of 100 mm. The divergence of Pattern ID 20 with

a mesh size of 200 mm is is likely caused due to the instability of the model due to a large horizontal crack along the length of the façade. This could also be exacerbated by a large mesh size and consequently, an insufficient number of integration points.

(a) Pattern ID 18. **(b)** Pattern ID 20 (96^{th} step). **(c)** Pattern ID 21.

(d) Pattern ID 23. **(e)** Pattern ID 24. **(f)** Pattern ID 101.



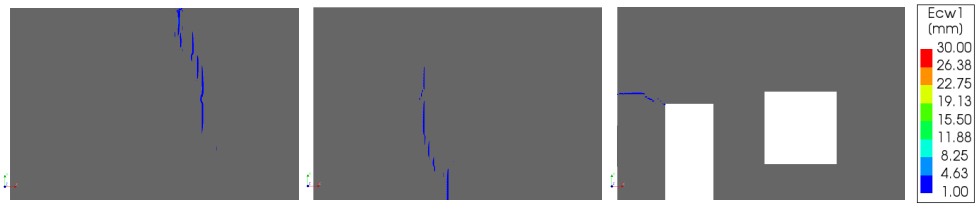
(g) Pattern ID 102. **(h)** Pattern ID 103.

Figure 4.11: Each Pattern ID generated for a mesh size of 200 mm.

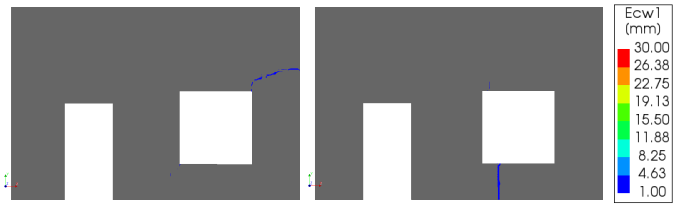
Mesh size of 50 mm

With a mesh size of 50 mm, the crack widths were computed at nodes located closer to each other, creating an impression of finer cracks, as seen in Fig. 4.12. The crack widths at the 200^{th} step for all the Pattern IDs can be seen in Fig. 4.12. The crack pattern images are seen to be very similar to the reference case. The pixel representation of the crack patterns generated with a mesh size of 50 mm is most similar to the images generated by statistics-based approach in terms of pixels used along the crack width. However, the small mesh size led to a large computation time for generating the random fields for the material properties, as well as for the analysis, as shown in Table 4.10. For this reason, it was decided to keep a mesh size of 100 mm for the computational physics-based approach.

(a) Pattern ID 18. **(b)** Pattern ID 20. **(c)** Pattern ID 21.



(d) Pattern ID 23. (e) Pattern ID 24. (f) Pattern ID 101.



(g) Pattern ID 102. (h) Pattern ID 103.

Figure 4.12: Each Pattern ID generated for a mesh size of 50 mm.

Table 4.10: Time required to run the analysis for each Pattern ID shown as a function of mesh size.

Pattern ID	Run time of the analysis[1] as a function of mesh size (minutes)		
	50 mm	100 mm	200 mm
18	26.72	18.00	1.45
20	75.72	18.42	4.07[2]
21	50.60	8.38	2.48
23	21.63	4.70	1.38
24	25.23	5.90	1.38
101	17.27	4.10	0.98
102	13.78	3.40	1.10
103	14.68	3.45	0.97

4.4.3 Soil-structure interaction

The behaviour of the masonry and the development of the crack patterns are greatly influenced by the stiffness of the soil-structure interface. As per Eq. 4.1, the stiffness is dependent on the thickness of the structure. For all the reference cases, the normal stiffness is assumed to be $8.125 \times 10^{10}\ N/m^3$, and the tangential stiffness along the length of the masonry and perpendicular it are both $8.125 \times 10^{8}\ N/m^3$. Here, we investigate the structural response of the masonry façade with an order of normal stiffness that is both greater and less than the reference stiffness.

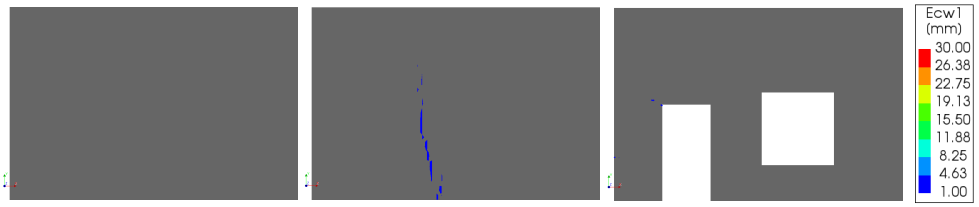
Normal stiffness of the soil-structure interaction $= 8.125 \times 10^{9}\ N/m^3$

With a normal stiffness of the soil-structure interaction as $8.125 \times 10^{9}\ N/m^3$, it is seen from Fig. 4.13, that many of the finite element models fail to produce crack widths of sufficient magnitude, i.e. $\geq$ 1 mm. Pattern ID 23, 101 and 102 had a very small crack development when compared to the reference cases. These models were subjected to subsidence induced loads that were applied to the bottom edges of the façade. This is due to the softer normal stiffness of the interface, which caused a

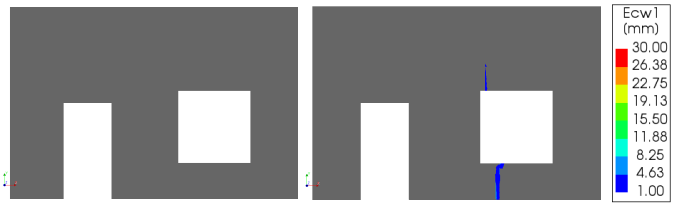
1 On a workstation with 6 cores and 16 GB RAM
2 Divergence occurs at the 96^{th} step.

stronger redistribution of the stresses and base settlements in the façade, causing smaller damage.

(a) Pattern ID 18. **(b)** Pattern ID 20. **(c)** Pattern ID 20.



(d) Pattern ID 23. **(e)** Pattern ID 24. **(f)** Pattern ID 101.



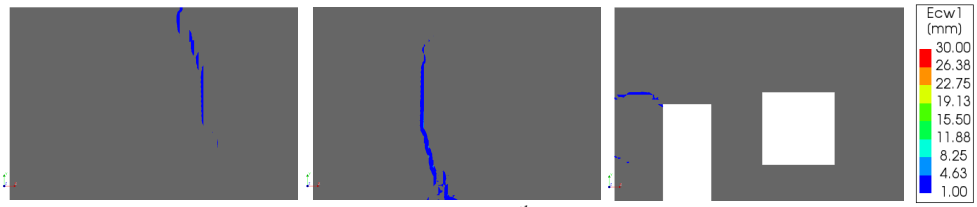
(g) Pattern ID 102. **(h)** Pattern ID 103.

Figure 4.13: Each Pattern ID with the normal stiffness of the soil-structure interaction as $8.125 \times 10^9\ N/m^3$.

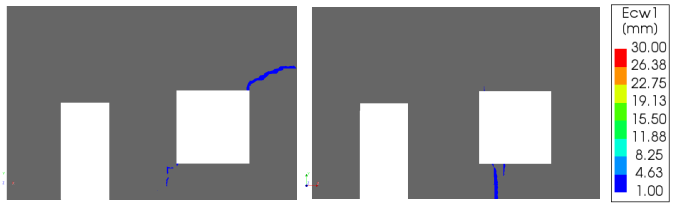
Normal stiffness of the soil-structure interaction $= 8.125 \times 10^{11}\ N/m^3$

The crack patterns produced with a normal stiffness of the soil-structure interaction as $8.125 \times 10^{11}\ N/m^3$, is seen from Fig. 4.14. Pattern ID 24 underwent divergence at the 115^{th} step after achieving a maximum magnitude of crack width of 6.13 mm, and the 114^{th} step is shown in Fig. 4.14e. It is seen that all the façades have a slightly more developed crack pattern than the reference case. This is due to a concentration of the stress at the bottom edge due to the higher normal stress, inducing a more rigid behaviour of the façade. This could also explain why Patter ID 24 failed to achieve convergence.

(a) Pattern ID 18. **(b)** Pattern ID 20. **(c)** Pattern ID 21.



(d) Pattern ID 23. **(e)** Pattern ID 24 (114^{th} step). **(f)** Pattern ID 101.



(g) Pattern ID 102. **(h)** Pattern ID 103.

Figure 4.14: Each Pattern ID with the normal stiffness of the soil-structure interaction as $8.125 \times 10^{11}\ N/m^3$.

4.4.4 Initial loading conditions

The dependence of the structural response of the masonry façades to the initial loading conditions is studied here. This is done by looking at crack patterns in two cases, the first is that of masonry façades subjected to no pre-compression loads, and the second is that of masonry façades subjected to twice the pre-compression load as that of the reference case. As Pattern IDs 101, 102 and 103 are not subjected to any pre-compression loads, these Pattern IDs are not included in this section.

Self-weight only

Fig. 4.15 shows the final damages of the different pattern IDs that were subjected to the self-weight alone as the initial condition. The crack patterns are quite similar to the reference case, with some Patter IDs achieving maximum magnitudes larger than those seen in the reference case. This can be seen prominently in the case of Pattern IDs 20 and 21 especially. This could be because the presence of a vertical pre-compression load tends to oppose the opening up of horizontal cracks. In the absence of this load, the crack widths were able to achieve higher magnitudes.

(a) Pattern ID 18. **(b)** Pattern ID 20. **(c)** Pattern ID 21.

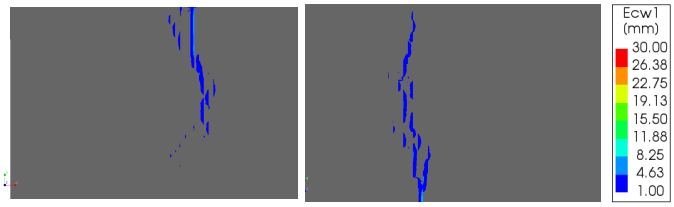
(d) Pattern ID 23. **(e)** Pattern ID 24.

Figure 4.15: Each Pattern ID loaded with only self-weight as the initial condition.

Self-weight and a pre-compression load of 1 MPa

The resultant damages of the various pattern IDs subjected to self-weight and a pre-compression load of 1 MPa as the initial condition are shown in Fig 4.16. The magnitude of the pre-compression load was chosen such that it was twice that of the reference case. The crack patterns are quite similar to those observed in the reference instance, except for Pattern ID 20 and 24. Both these crack pattern types underwent divergence at intermediate steps after achieving a maximum crack width of 6.72 mm and 8.94 mm. This could be due to the instability in the structure after achieving these large crack widths compounded by the additional stresses due to the large pre-compression load.

(a) Pattern ID 18. **(b)** Pattern ID 20 (63^{rd} step). **(c)** Pattern ID 21.



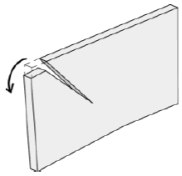
(d) Pattern ID 23. **(e)** Pattern ID 24 (119^{th} step).

Figure 4.16: Each Pattern ID loaded with only self-weight as the initial condition.

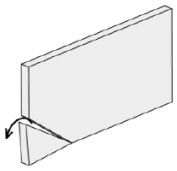
4.5 LIMITATIONS OF THE FINITE ELEMENT MODELLING

4.5.1 Pattern IDs excluded from this thesis

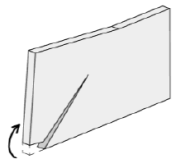
Pattern IDs 30, 31, 32 and 201 were included in the data generation process when the classification of crack pattern images by crack pattern type was attempted at TNO in 2019, as explained in Section 2.5.1. Pattern IDs 30, 31 and 32 were also included in the data generation process to fit a CNN that could perform the similarity assessment between crack pattern image pairs in 2020. This is explained in Section 2.5.2.



(a) Pattern ID 30 [7].



(b) Pattern ID 31 [7].



(c) Pattern ID 32 [7].

(a) Pattern ID 201 [6].

After many attempts, sufficiently robust finite element models that can simulate these crack patterns could not be generated. A few attempts generated very small cracks ($\leq$ 1 mm), which is too small for the purpose of this thesis. For this reason, these Pattern IDs are excluded from this thesis.

4.5.2 Automation of the generation of finite element models

All operations in DIANA FEA 10.4 can be recorded as python commands. The paramterisation of the masonry models was achieved by taking advantage of the Python scripting language in combination with the script commands of the Diana FEA 10.4. This was also useful to incorporate randomness into the model. The spatial variation of the material properties realised through the generation of random fields (Appendix B), the choice of location of the crack initiation, and the magnitude of the load on each structure was realised through the inbuilt python functions. The reproduciblity of these models is achieved with the use of a random seed. With this implementation, the automated generation of each crack pattern can be realised on masonry façades with varying material and geometric properties.

However, due to the randomness incorporated in each model, not every model produced in the automated computational physics-based approach is usable in this project. For example, sometimes the combination of the spatially varying material properties and magnitude of the load fail to produce cracks with sufficiently large widths ($\geq$ 1 mm). Other times, the same combination of parameters produced localised points of weakness that generated crack patterns different from those expected. Sometimes, this also caused stability issues in the models that led to divergence. This was especially true regarding the localisation of damage around openings and the development of additional small flexural cracks. Even when the model ran successfully to completion, some of the crack patterns generated were not useful for this reason. These crack patterns were generated again with different random seeds to obtain expected crack patterns.

5 RESULTS

5.1 INTRODUCTION

The results of crack pattern similarity assessments performed by the raters are reported in this chapter. An inter-rater and intra-rater agreement is also provided. Additionally, the results of the similarity assessment performed by deep neural networks fitted to crack pattern images generated via the statistics-based approach are also shown here.

5.2 RESULTS OF ANALYSIS OF THE SIMILARITY LABELS AND INTER-RATER VARIABILITY

5.2.1 Data generated from the statistics-based approach

The neural network, which is used to predict the masonry crack pattern similarity, was fitted to crack pattern image pairs assessed by raters. Twenty eight raters volunteered to assess each crack pattern image pair generated using the statistics-based approach. Three similarity labels were used to perform this similarity assessment: crack pattern similarity label, damage severity label, and overall similarity label. The similarity labels are rated using these five rating classes: *very dissimilar, dissimilar, similar, very similar*, and *I cannot say*. The raters come from different academic backgrounds and have varying levels of experience working with masonry structures. Each rater was provided with a unique Rater ID. Three broad expertise levels were defined: industry experts (masonry expert and structural engineering expert), PhD students, and MSc students. A summary of the expertise level and number of annotations of each rater was provided in Table 5.1.

Of all the image pairs, only the image pairs assessed by at least three separate raters were used to fit and test the neural network. A total of 2587 such image pairs were obtained. A summary of the number of annotations performed by each rater is provided in Fig. 5.1.

Table 5.1: Raters who performed assessment of masonry crack pattern image pairs generated using the statistics-based approach.

Rater ID	Expertise level	Rater ID	Expertise level	Rater ID	Expertise level
1	Masonry expert	10	PhD Student	20	MSc Student
2	Structural engineering expert	11	PhD Student	21	MSc Student
		12	PhD Student	22	MSc Student
3	Masonry expert	13	PhD Student	23	MSc Student
4	Masonry expert	14	PhD Student	24	MSc Student
5	Masonry expert	15	MSc Student	25	MSc Student
6	Masonry expert	16	MSc Student	26	MSc Student
7	Masonry expert	17	MSc Student	27	MSc Student
8	PhD Student	18	MSc Student	28	MSc Student
9	PhD Student	19	MSc Student		

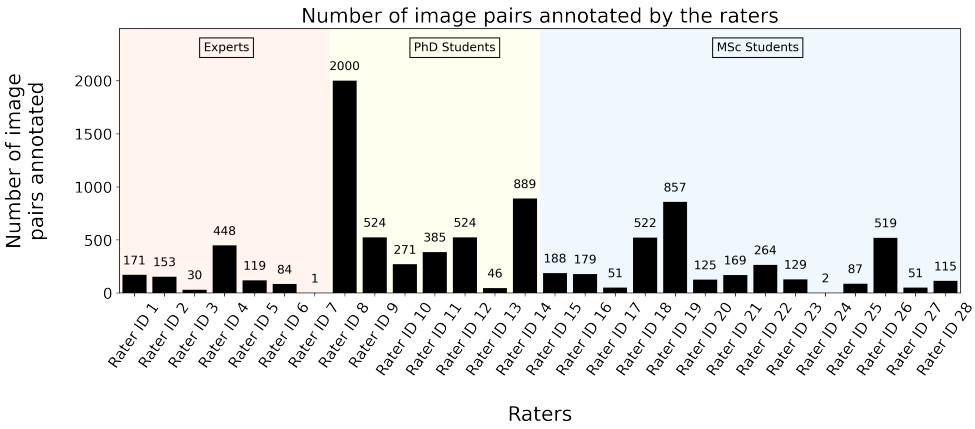


Figure 5.1: Summary of the number of image pairs generated using the statistics-based approach assessed by each rater.

Reliability of the similarity assessment performed by the raters

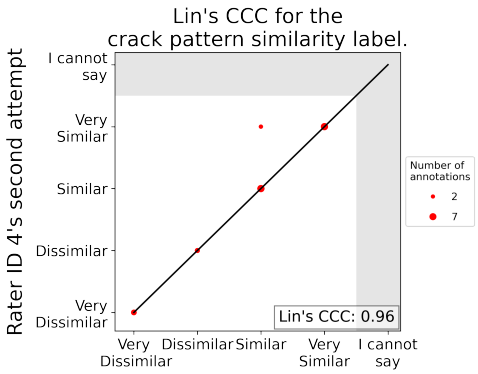
INTRA-RATER AGREEMENT To study how consistent a rater was with themselves, 23 image pairs were annotated twice by the same expert, Rater ID 4. These 23 image pairs were provided along with the larger sample of image pairs, therefore the expert was unaware that these were images they had annotated before. The intra-rater agreement was then studied using Lin's Concordance Correlation Coefficient (CCC), as explained in Section 3.6.1. The rater performed the similarity assessment by choosing one from the five available ratings for each label, i.e. *very dissimilar, dissimilar, similar, very similar*, and *I cannot say*. The labels were initially given to the raters to be evaluated on a categorical scale consisting of named ratings as mentioned above. However, it was seen that a more useful evaluation of the the similarity assessment was on an interval scale. This is because, a rating of *similar* and *very similar* would be closer qualitatively than a rating of *dissimilar* and *very similar*. An interval scale was able to capture this difference better than a categorical scale. Thus, the ratings were considered to be part of an interval scale for the assessment of agreement and for training the neural network. Table 5.2 shows how the categorical ratings are converted to an interval scale.

The Lin's CCC was calculated for all the three labels for 23 image pairs assessed by Rater ID 4 twice and is shown in Fig. 5.2. The assessment done by the rater the first time was kept as the reference rating. The crack pattern similarity label had the highest intra-rater agreement among the labels with a Lin's CCC of 0.96. This was followed by the overall similarity label with a Lin's CCC of 0.87. The damage severity label had the lowest intra-rater agreement with a Lin's CCC value of 0.72.

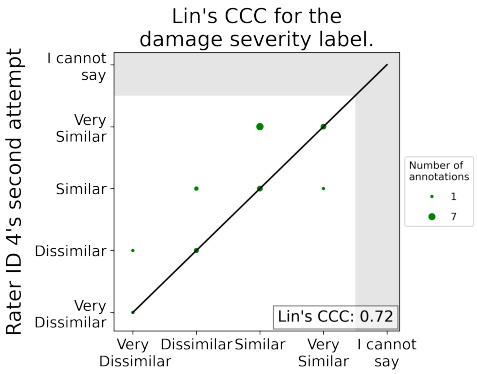
What value of the Lin's CCC is needed to prove sufficient agreement between raters depends on the nature of the task at hand. Raters assessing tasks which require a high agreement between raters has more stringent requirements for minimum value of the Lin's CCC to prove the agreement. The recommended lowest

Table 5.2: Converting the categorical rating to an interval scale.

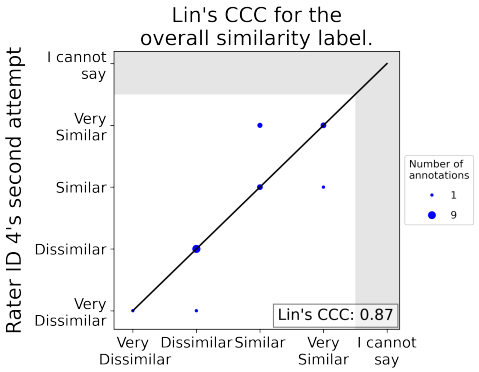
Categorical rating	Interval rating
Very Dissimilar	1
Dissimilar	2
Similar	3
Very Similar	4
I cannot say	5



(a) Lin's CCC for the crack pattern similarity label.



(b) Lin's CCC for the damage severity label.



(c) Lin's CCC for the overall similarity label.

Figure 5.2: Lin's CCC to determine the intra-rater agreement for Rater ID 4 for all three similarity assessment labels.

value for a Lin's CCC as per literature is given in Section 3.6.1. However, as the lowest intra-rater agreement seen is 0.72, in this project, it was assumed that a Lin's CCC of 0.70 was enough to show sufficient agreement to the reference ratings in the masonry crack pattern similarity assessment. This has been summarised in Fig. 5.3.

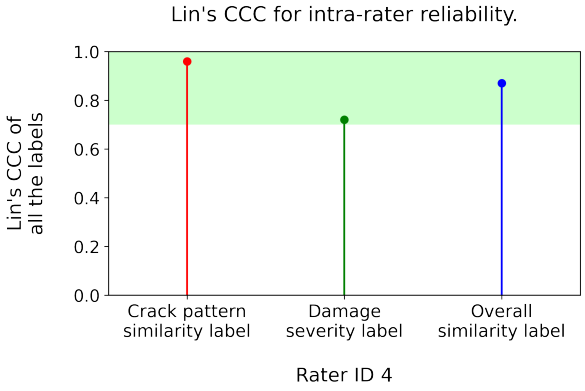


Figure 5.3: Summary of the intra-rater agreement assessment.

AGREEMENT WITH THE STANDARD RATER The raters who took part in this similarity assessment come from varying academic backgrounds and varying degrees of experience on working with masonry structures. It is useful to compare each

rater's performance to a standard rater. For this purpose, Rater ID 4 was chosen as the standard rater due to his/her level of expertise and previous experience with assessing masonry crack pattern similarities. The agreement of each rater with the standard rater was assessed for all three labels. It is necessary for a rater to have assessed at least two image pairs in common with the standard rater to be able to compute a Lin's CCC. Two raters, Rater ID 7 and Rater ID 24, had not assessed sufficient image pairs in common with the standard rater (Rater ID 4), hence they were not included in the analysis for determining the agreement with the standard rater. Additionally, each rater must have assessed 10 image pairs in common with the standard rater to compute a reliable Lin's CCC, as explained in Section 3.6.1.

Lin's CCC was computed for the remaining 26 raters with Rater ID 4 as the standard rater for the crack pattern similarity label as shown in Fig. 5.4 and Fig. 5.5. It was seen from Fig. 5.5 that 10 raters had a CCC value $\geq$ 0.70 and at least 10 annotations in common with Rater ID 4: Rater ID 2, Rater ID 6, Rater ID 8, Rater ID 9, Rater ID 14, Rater ID 15, Rater ID 20, Rater ID 21, Rater ID 23, and Rater ID 26. Rater ID 21 was seen to have the highest agreement with Rater ID 4, with a Lin's CCC value of 0.96.

The Lin's CCC was evaluated for the same 26 raters against Rater ID 4 for the damage severity label as shown in Fig. 5.6 and Fig. 5.7. It was seen that there were no raters that had a sufficient agreement with the standard rater. With a Lin's CCC score of 0.67, Rater ID 21 had the most agreement with Rater ID 4.

For the overall similarity label, the Lin's CCC was calculated for the same 26 raters against Rater ID 4 as shown in Fig. 5.8 and Fig. 5.9. Rater ID 21 seemed to have the maximum agreement with Rater ID 4 with a Lin's CCC score of 0.88. From Fig. 5.9, it was seen that six raters had a CCC $\geq$ 0.70 and at least 10 annotations in common with Rater ID 4: Rater ID 2, Rater ID 5, Rater ID 6, Rater ID 8, Rater ID 20, and Rater ID 21. It was also seen that Rater ID 27 had a Lin's CCC value of -0.13 for this label. Since this value is very low and suggests disagreement with the standard rater, it was not plotted in Fig. 5.9.

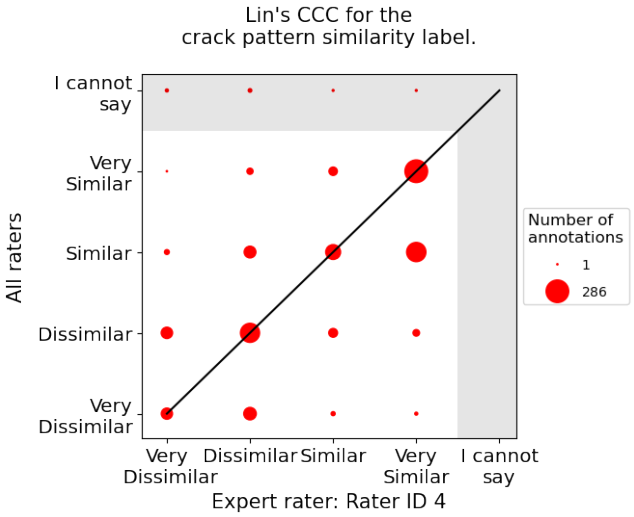


Figure 5.4: Each rater's attempt to assess the crack pattern similarity label plotted against Rater ID 4's attempt.

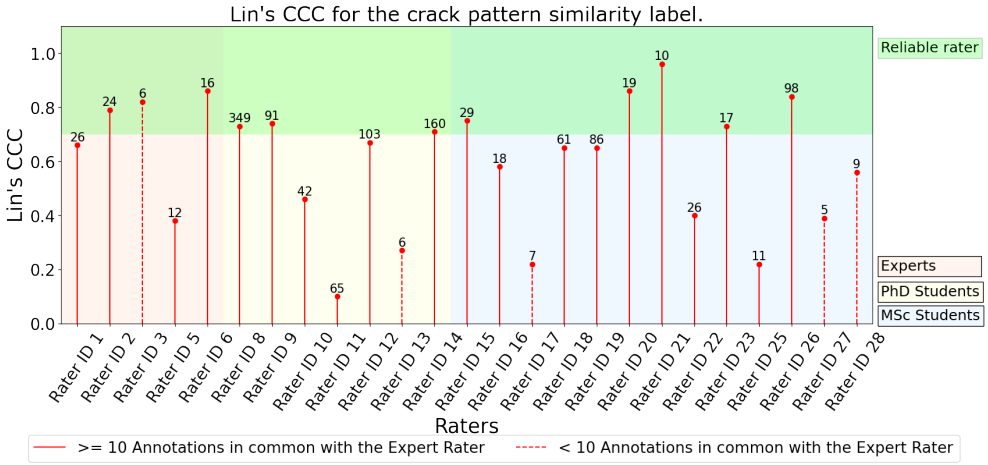


Figure 5.5: Each rater's agreement with Rater is ID 4 shown using Lin's CCC for the crack pattern similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

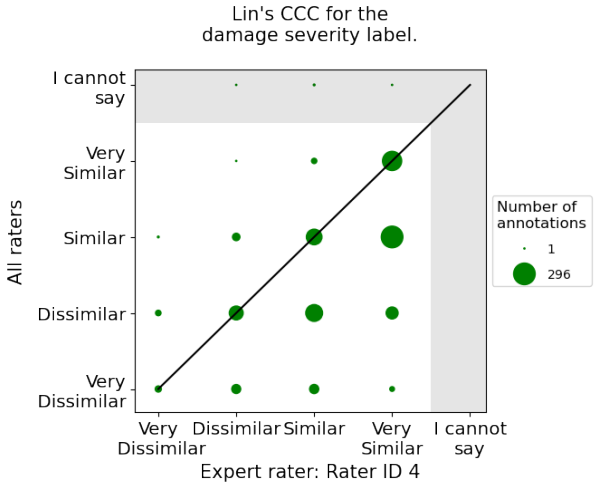


Figure 5.6: Each rater's attempt to assess the damage severity label plotted against Rater ID 4's attempt.

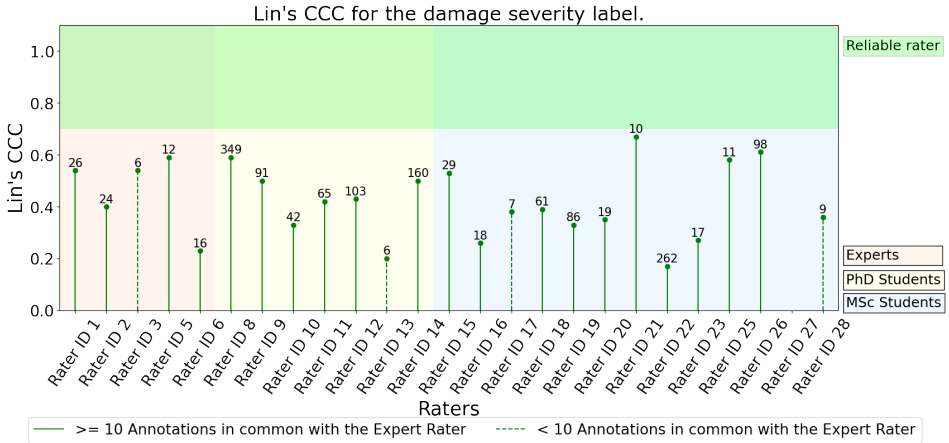


Figure 5.7: Each rater's agreement with Rater ID 4 is shown using Lin's CCC for the damage severity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

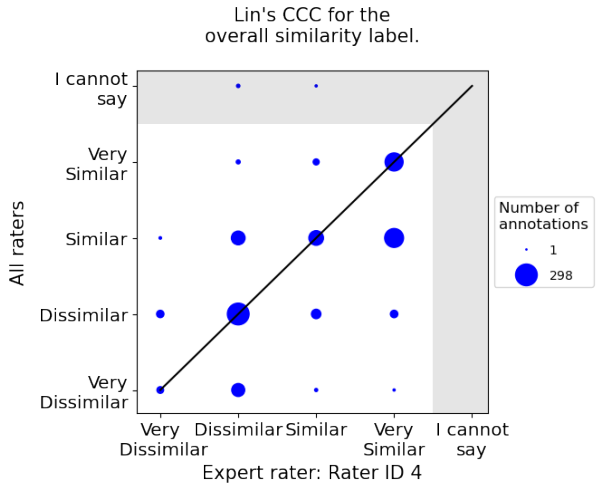


Figure 5.8: Each rater's attempt to assess the overall similarity label plotted against Rater ID 4's attempt.

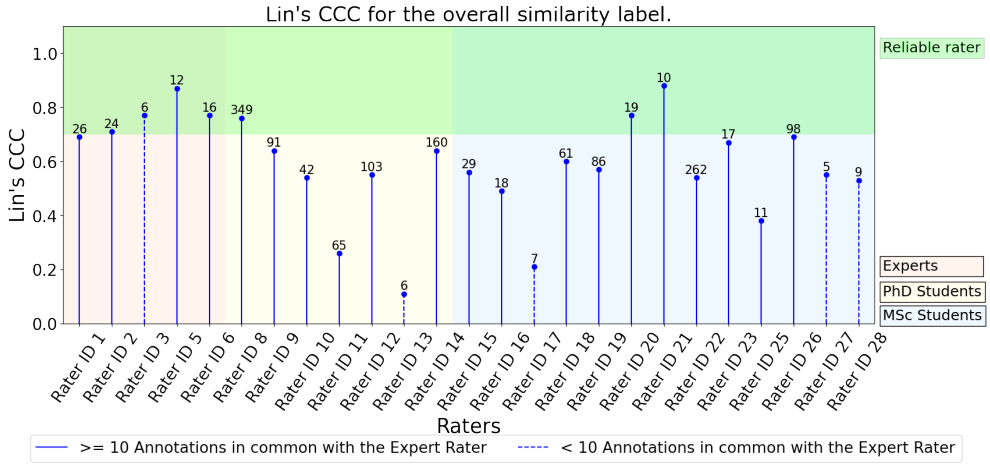


Figure 5.9: Each rater's agreement with Rater ID 4 is shown using Lin's CCC for the overall similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

From Fig. 5.4, Fig. 5.6, and Fig. 5.8, it was seen that there were very few image pairs that were annotated with the *I cannot say* label. Only about 0.7% of the total sample of annotations contained the *I cannot say* label. This shows that, the rating classes were not present uniformly in the annotated image pairs, and currently the labelled data formed an unbalanced data set. This could interfere with the performance of the neural network and decrease its prediction accuracy. Additionally, it was also seen that the standard rater, Rater ID 4, did not use the *I cannot say* label at all. Considering these factors, it was decided to avoid the *I cannot say* label in future annotation tasks.

OVERALL INTER-RATER AGREEMENT The assessment of all 28 raters had been used to calculate the inter-rater reliability for all three labels using the Krippendorff's Alpha, as explained in Section 3.6.2. Additionally, the Krippendorff's Alpha was also calculated by dividing the raters based on their expertise level as Experts, PhD students, and MSc students. This is shown in Fig. 5.10.

As with the Lin's CCC, what minimum value of the Krippendorff's Alpha is needed to establish a good agreement between the raters depends on the nature

of the task at hand. In this thesis, the minimum *alpha* value needed to prove sufficient inter-rater agreement was decided based on Krippendorff's recommendations, as explained in Section 3.6.2. Thus, based on these recommendations, it was seen that no category of raters had sufficient agreement with each other to provide a reliable assessment of masonry crack pattern similarity. This is shown in Fig. 5.10. It was also seen that the group of expert raters had the highest agreement amongst themselves for all three labels. It was also seen that the crack pattern similarity label had the highest rate of agreement among the three labels for all categories of raters, while the damage severity label consistently had the least agreement among all categories of raters.

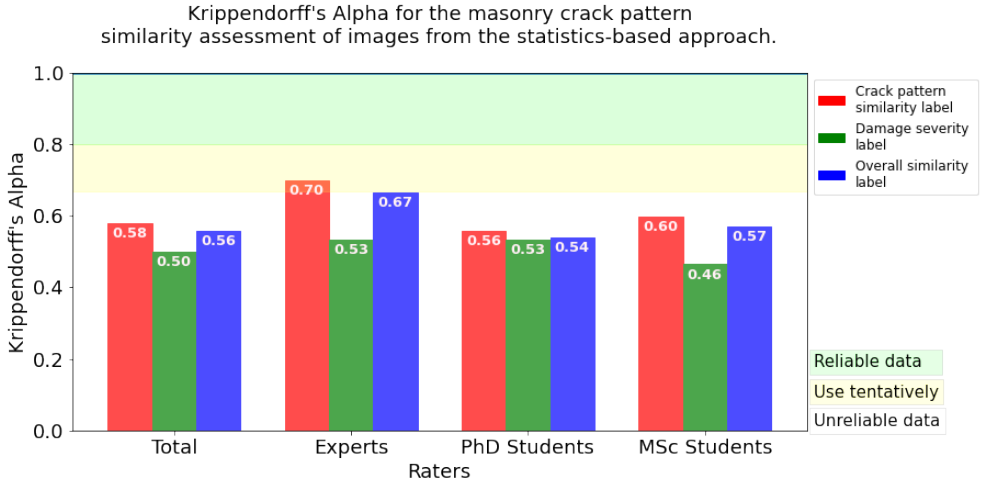


Figure 5.10: Krippendorff's Alpha for the inter-rater agreement of the masonry crack pattern similarity assessment of images generated in the statistics-based approach.

5.2.2 Data generated from the computational physics-based approach

To assess the similarity of the crack pattern image pairs generated in the computational physics-based approach, seven raters volunteered to annotate a total of 500 image pairs with similarity labels. Each image pair was assessed at least three times by different raters. The three similarity labels used in the similarity assessment of the images generated in the statistics-based approach were used here as well: crack pattern similarity label, damage severity label and overall similarity label. The similarity labels were rated using the modified rating scale with four classes: *very dissimilar, dissimilar, similar*, and *very similar*. Each rater was provided with a unique Rater ID as provided earlier in Table 5.1. A summary of the number of annotations performed by each rater is provided in Fig. 5.11.

Reliability of the similarity assessment performed by the raters

As with the data generated from the statistics-based approach, it is useful to study the agreement among the raters regarding the similarity of the crack pattern image pairs generated using the computational physics-based approach. This is discussed in the following sections.

AGREEMENT WITH THE STANDARD RATER To evaluate the agreement among the raters regarding the similarity assessment, the modified similarity classes presented to the raters as categorical ratings had to be converted to an interval scale. The numerical values used to make this conversion was the same as that in the statistics-based approach, as shown in Table 5.2. Rater ID 4 was selected as the standard rater.

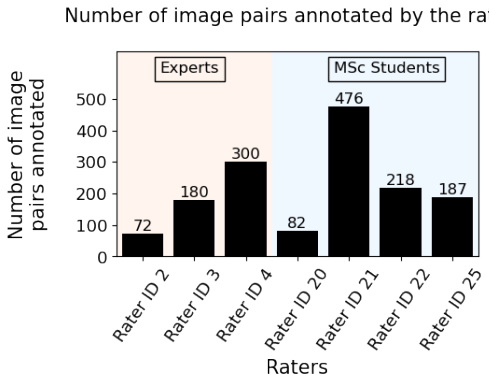


Figure 5.11: Summary of the number of image pairs generated using the computational physics-based approach assessed by the raters.

The similarity assessments by all the other raters are compared against those done by Rater ID 4 for all three labels. All remaining six raters had assessed more than two image pairs in common with the Rater ID 4. The Lin's CCC was computed for all these six raters, as shown in Fig. 5.13, Fig. 5.15 and Fig.5.15.

It is seen from Fig. 5.13 that four raters had a CCC value $\geq$ 0.70 and at least 10 annotations in common with Rater ID 4: Rater ID 2, Rater ID 3, Rater ID 20, and Rater ID 21. With a Lin's CCC score of 0.86, Rater ID 3 and Rater ID 21 appeared to have the most agreement with Rater ID 4.

On evaluating the Lin's CCC is for the damage severity label, as shown in Fig. 5.15, it was seen there was only one rater that had a sufficient agreement with standard rater. This is Rater ID 2, with Lin's CCC value of 0.76.

For the overall similarity label, the Lin's CCC was calculated for the same six raters against Rater ID 4 as shown in Fig. 5.17. Four raters had sufficient agreement with Rater ID 4: Rater ID 2, Rater ID 3, Rater ID 20, and Rater ID 21. With a Lin's CCC score of 0.89, Rater ID 21 had the highest level of agreement with Rater ID 4.

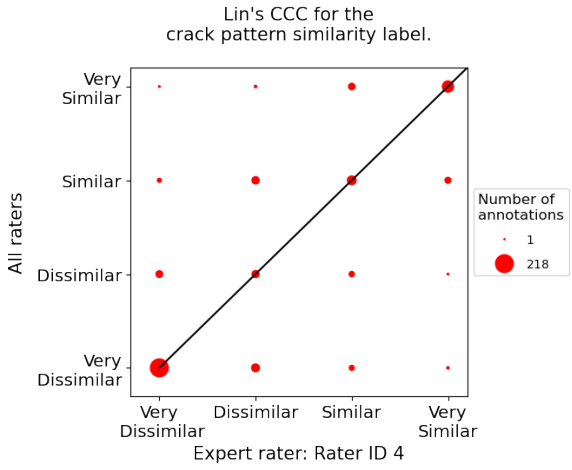


Figure 5.12: Each rater's attempt to assess the crack pattern similarity label plotted against Rater ID 4's attempt.

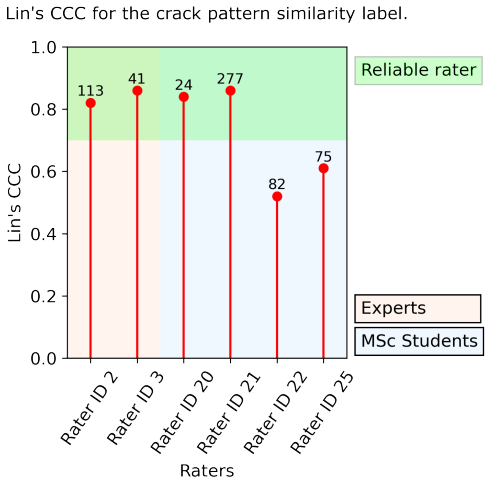


Figure 5.13: Each rater's agreement with Rater ID 4 is shown using Lin's CCC for the crack pattern similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

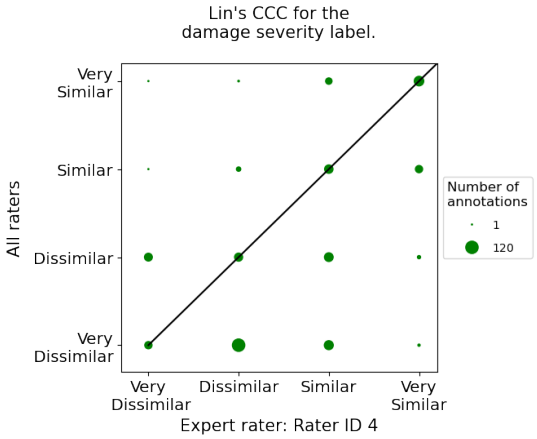


Figure 5.14: Each rater's attempt to assess the damage severity label plotted against Rater ID 4's attempt.

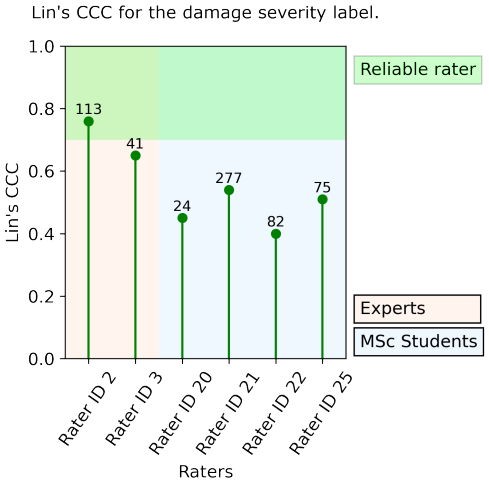


Figure 5.15: Each rater's agreement with Rater ID 4 shown using Lin's CCC for the damage severity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4's attempt.

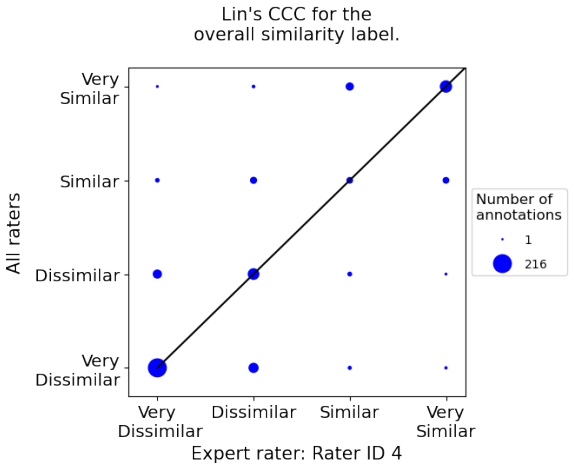


Figure 5.16: Each rater's attempt to assess the overall similarity label plotted against Rater ID 4's attempt.

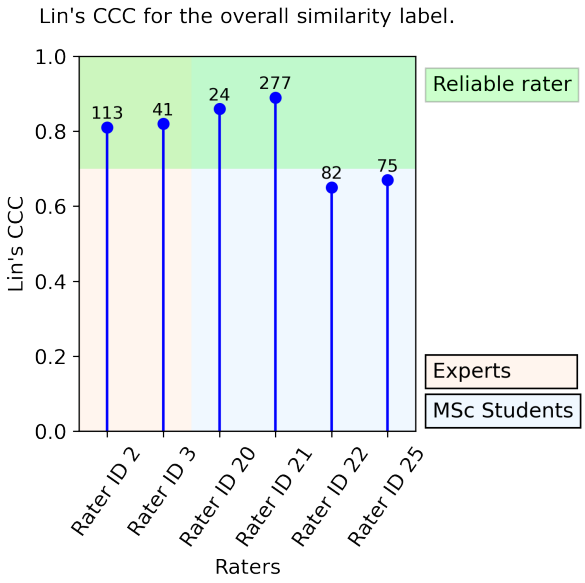


Figure 5.17: Each rater's agreement with Rater ID 4 shown using Lin's CCC for the overall similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

OVERALL INTER-RATER AGREEMENT The similarity assessments performed by all seven raters were used to calculate the Krippendorff's Alpha to investigate the inter-rater reliability for all three labels, as explained in Section 5.2.1. Additionally, the Krippendorff's Alpha was also calculated by dividing the experts based on their expertise level as Experts and MSc students. This is summarised in Fig. 5.18.

Based on the recommendations for minimum α value to establish sufficient agreement (Section 5.2.1), it was seen that only the crack pattern similarity label and the overall similarity label of the expert raters had sufficient agreement to provide a reliable assessment of masonry crack pattern similarity. The crack pattern similarity label, assessed by the MSc students, showed the lowest degree of agreement. This could be the reason why the inter-rater agreement of the crack pattern similarity label among all the raters showed a lower degree of agreement than anticipated.

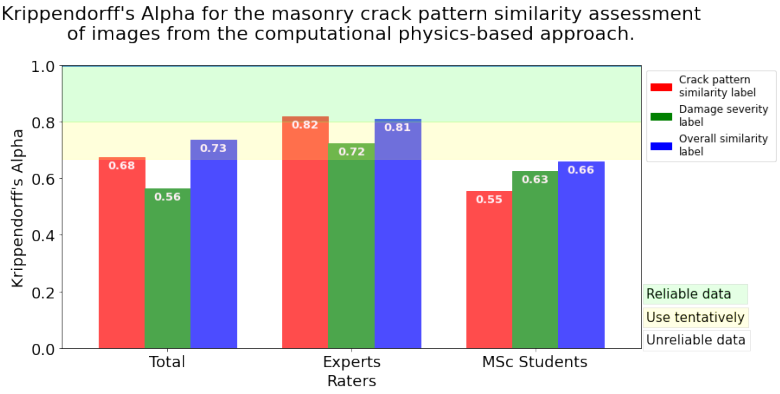


Figure 5.18: Krippendorff's Alpha for the inter-rater agreement of the masonry crack pattern similarity assessment of images generated in the computational physics-based approach.

5.2.3 Data generated from the statistics-based approach, computational physics-based approach and the real-world images

A third batch of data was generated, which consisted of image pairs from all three approaches: the statistics-based approach, the computational physics-based approach, and the real-world images. Three raters consented to annotate a total of 50 image pairs to assess the similarity of these crack pattern images. Each image pair was assessed by each rater. The three similarity labels used in the previous tasks were used here as well: crack pattern similarity label, damage severity label, and overall similarity label. The similarity labels were rated using the modified rating scale with four classes: *very dissimilar, dissimilar, similar*, and *very similar*. A summary of the number of annotations performed by each rater is provided in Fig. 5.19.

Reliability of the similarity assessment performed by the raters

The agreement among the raters regarding the similarity assessment of the crack pattern image pairs are discussed below.

AGREEMENT WITH THE STANDARD RATER Rater ID 4 was selected as the standard rater. Rater ID 2's and Rater ID 3's similarity assessments were compared against those done by Rater ID 4 for all three labels. The Lin's CCC (Section 3.6.1) was computed for both these raters, as shown in Fig. 5.21, Fig. 5.23, and 5.25.

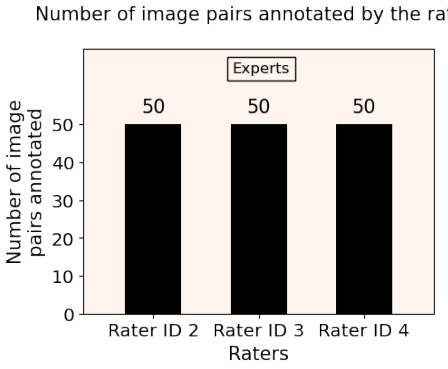


Figure 5.19: Summary of the number of image pairs generated using all three approaches assessed by the raters.

It is seen from Fig. 5.21 that both raters had a Lin's CCC value ≥ 0.70 with at least 10 annotations in common with Rater ID 4. Rater ID 2 and 3 had Lin's CCC values of 0.77 and 0.71 respectively.

On evaluating the Lin's CCC for the damage severity label, against Rater ID 4, as shown in Fig. 5.22 and Fig. 5.23, it is seen that neither rater had a sufficient agreement with standard rater. Both raters exhibited the lowest Lin's CCC values for any label in this data set.

For the overall similarity label, the Lin's CCC was calculated as shown in Fig. 5.25. Only Rater ID 2 had a Lin's CCC value that was just enough to establish sufficient agreement with Rater ID 4, with a magnitude of 0.70.

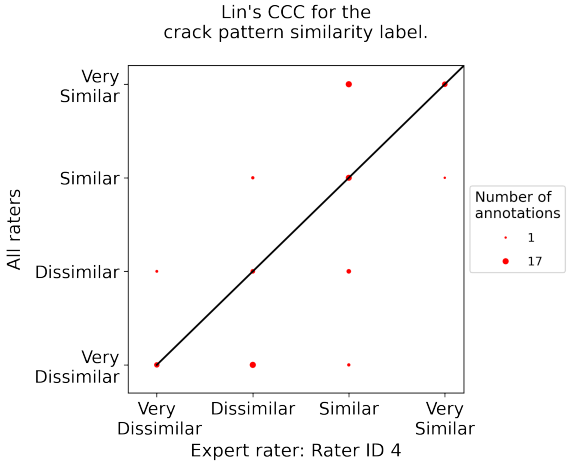


Figure 5.20: Each rater's attempt to assess the crack pattern similarity label plotted against Rater ID 4.

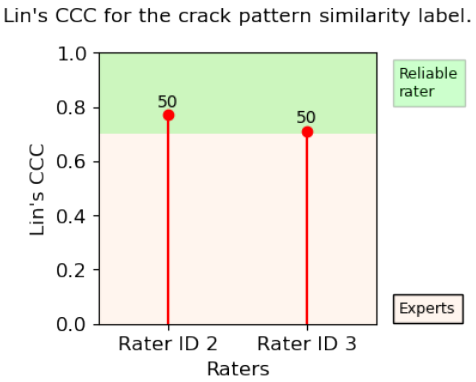


Figure 5.21: Each rater's agreement with Rater ID 4 shown using Lin's CCC for the crack pattern similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

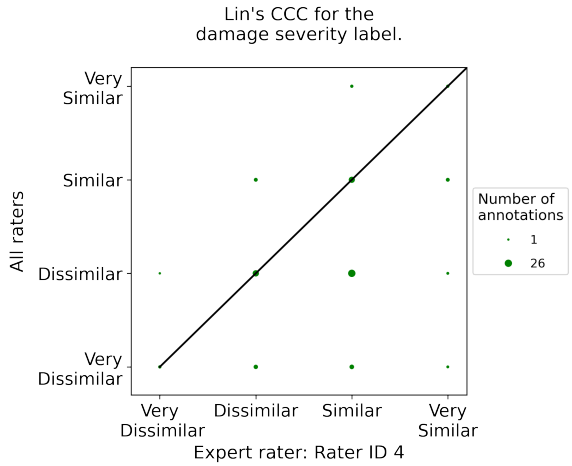


Figure 5.22: Each rater's attempt to assess the damage severity label plotted against Rater ID 4.

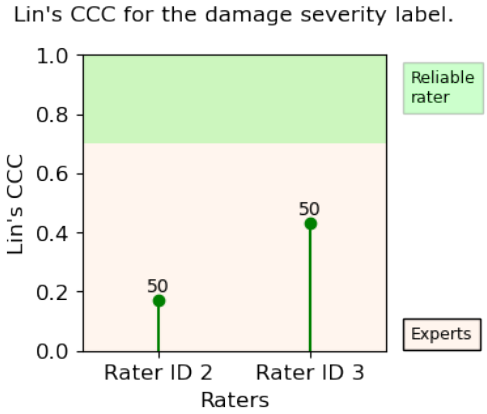


Figure 5.23: Each rater's agreement with Rater ID 4 shown using Lin's CCC for the damage severity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

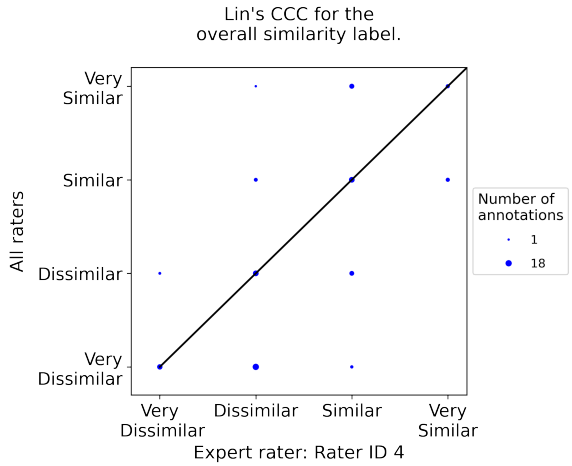


Figure 5.24: Each rater's attempt to assess the overall similarity label plotted against Rater ID 4.

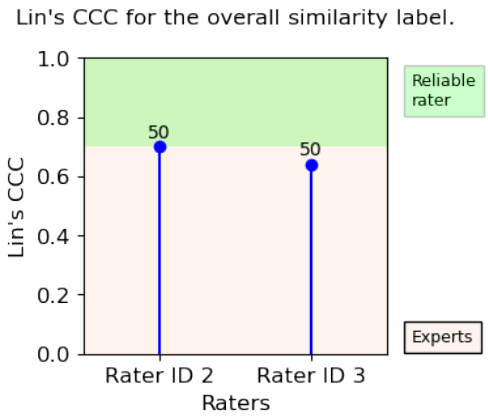


Figure 5.25: Each rater's agreement with Rater ID 4 shown using Lin's CCC for the overall similarity label. The number above each rater's CCC value represents the annotations they have in common with Rater ID 4.

OVERALL INTER-RATER AGREEMENT The Krippendorff's Alpha, discussed in Section 3.6.2, was calculated using the similarity assessments completed by all 3 raters to evaluate the inter-rater reliability for the 3 labels, as shown in Fig. 5.26. All the three raters who performed this assessment were experts. Considering the minimum α value suggestions for establishing sufficient agreement, as explained in Section 3.6.2, it was seen that only the crack pattern similarity label had sufficient agreement among the raters to provide a reliable assessment of masonry crack pattern similarity. The overall similarity label had enough agreement to be used with caution, while there was significantly lower agreement among the raters regarding the damage severity label.

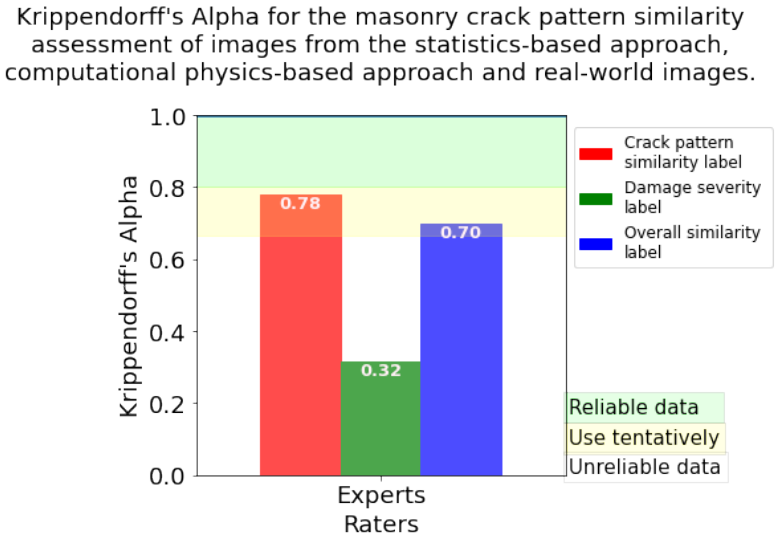


Figure 5.26: Krippendorff's Alpha for the inter-rater agreement of the masonry crack pattern similarity assessment of images generated using all three approaches.

.

5.3 RESULTS OF THE SIMILARITY ASSESSMENT PERFORMED BY THE CNN

A convolutional neural network (CNN) was developed to perform the similarity assessment of crack pattern image pairs and its performance was evaluated against the performance of the raters who took part in this project. The details of this neural network are explained in Section 3.7 and in the research done by Kleijn [1]. Four different subsets of data were defined in order to test the accuracy of the neural network in predicting the similarity of crack pattern image pairs generated from the computational physics-based approach. These four subsets are defined as follows:

- Subset 1: The performance of the neural network was evaluated as a function of the different raters' assessments used to fit the neural network. The generalisability of the network was not explored here.
- Subset 2: The generalisability of the fitted neural networks were explored with respect to the crack pattern type (Pattern ID).
- Subset 3: The generalisability of the fitted neural networks were explored with respect to the geometry of the façades. This included variation in the aspect ratio and the presence of openings in the façade.
- Subset 4: The generalisability of the fitted neural networks were tested to see how well it can evaluate the similarity in crack patterns generated from the statistics-based approach and the real-world images.

These subsets were further divided into smaller subsections, as explained in the following sections. It is to be noted that in this thesis, the neural network had only been fitted to image pairs generated from the computational physics-based approach. It is to be noted that, as the ground truth was established by considering the average of the similarity labels assigned by the rater, information was lost in this process.

The performance of the neural network was measured using four metrics: accuracy, precision, recall, and the F-score, which are explained in Section 3.7. For each subset of data, the accuracy, recall, precision, and F-score was calculated. Each subset was run 5 times in the neural network, which gave a mean and a standard deviation per subset for each of these metrics. The accuracy of each subset is presented below. The precision, recall, and F-score of each subset is presented in Appendix E.

5.3.1 Subset 1: No generalisation

This subset of data tested the performance of the neural network, where the ground truth of the fitting and testing data was established by different groups of raters. Five such subsections were established based on the agreement of the raters with the standard rater and the expertise level of each rater. This can be seen in Table 5.3. All the Pattern IDs, aspect ratios, and geometries generated in the computational physics-based approach were used in each subsection formed from Subset 1.

CRACK PATTERN SIMILARITY LABEL It is seen in Fig. 5.27 that the crack pattern similarity label had a higher accuracy than the other labels in all subsets except for Subset 1.5. The highest accuracy, 94%, for the crack pattern label is obtained by Subset 1.2, where the fitting and testing was done solely on the assessments performed by the standard rater, Rater ID 4. The lowest accuracy for this label was 79%, obtained for Subset 1.5 when the neural network was fitted and tested to the

Table 5.3: Details of Subset 1, formed from the labelled data generated from the computational physics-based approach used to fit and test the neural network.

Subset No	Criteria used to choose raters		Raters included in the subset		Number of image pairs used	
	Fitting	Testing	Fitting	Testing	Fitting	Testing
1.1	All raters	All raters	2, 3, 4, 20, 21, 22, 25	2, 3, 4, 20, 21, 22, 25	375	125
1.2	Standard rater	Standard rater	4	4	225	75
1.3	Raters with good agreement with the standard rater[1]	Raters with poor agreement with the standard rater[2]	2, 3, 4, 20, 21	22, 25	375	125
1.4	Expert raters	Expert raters	2, 3, 4	2, 3, 4	293	98
1.5	MSc student raters	MSc student raters	20, 21, 22, 25	20, 21, 22, 25	362	121

similarity assessment by the MSc students. This could be related to the low inter-rater agreement for this label among the MSc student raters, as seen in Fig. 5.18. The accuracy of Subset 1.1 for the crack pattern similarity label, which contains the similarity assessment of all seven raters, was 87%. This was a 5% decrease from the accuracy of 92% obtained from the neural network fitted to only statistically-generated crack pattern images with labels given by the masonry expert at TNO in 2020 (Section 2.5.2). The statistically generated data used in 2020, contained 395 crack pattern images on masonry façades without openings that were assessed by Rater ID 4. Subset 1.2 was most similar to this data set, with 300 crack pattern image pairs generated through the computational physics-based approach on façades with and without openings, assessed by Rater ID 4. It is also interesting to note that the neural network fitted to Subset 1.2 predicted the crack pattern similarity label with a 2% higher accuracy than the neural network fitted to the statistically-generated image pairs.

DAMAGE SEVERITY LABEL The performance of similarity prediction of the damage severity label as a function of accuracy is seen in Fig. 5.27. The highest accuracy, 85%, was obtained when the neural network was fitted to Subset 1.2, containing only the annotations performed by standard rater, Rater ID 4. This was about 6% lower than the 92% accuracy obtained by the neural network fitted to the statistically generated images in 2020 (Section 2.5.2). The accuracy obtained by the neural network fitted to Subset 1.4 (containing the assessment of only the expert raters) and Subset 1.5 (containing the assessment of only the MSc student raters) were only slightly lower at 85% and 81%, respectively. The lowest accuracy, 72%, was obtained when the neural network was fitted to Subset 1.3.

OVERALL SIMILARITY LABEL The accuracy of the overall similarity label, was seen to follow the general trend of the accuracy of the crack pattern similarity label, as seen in Fig. 5.27. The highest accuracy of this label was seen when the neural network was fitted to and tested with Subset 1.2 containing only the similarity assessments performed by Rater ID 4, with a magnitude of 85%. This was around 6% less accurate than the neural network fitted to statistically produced images in 2020 (Section 2.5.2), which achieved 92% accuracy. When the neural network was

1 These raters were chosen such that they have a Lin's CCC value of at least 0.75 for at least one of the similarity labels, as seen in Fig. 5.13, Fig. 5.15, and Fig. 5.17.

2 These raters were chosen such that they have a Lin's CCC value less than 0.75 for all three similarity labels, as seen in Fig. 5.13, Fig. 5.15 and Fig. 5.17.

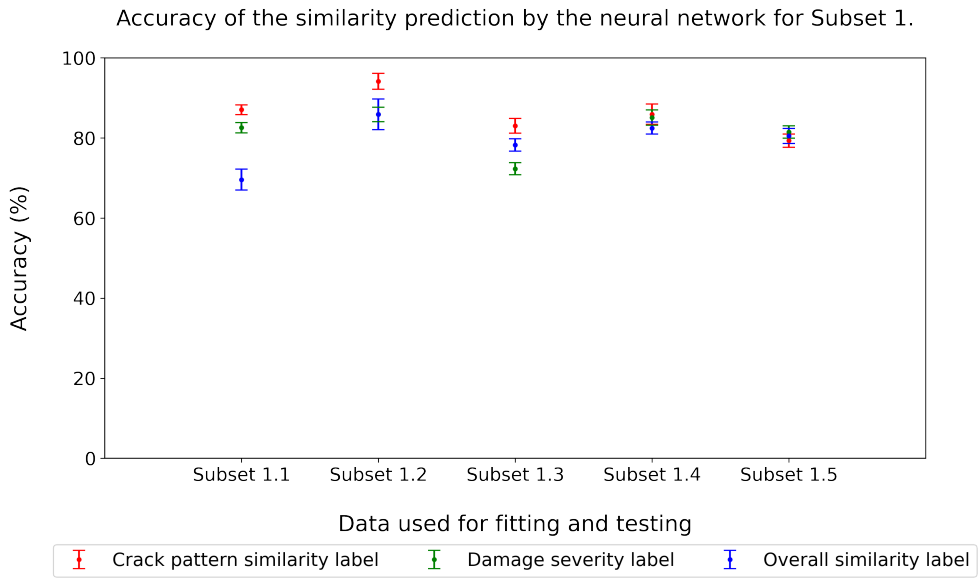


Figure 5.27: Summary of accuracy of the labels with the neural network fitted and tested to Subset 1.

fitted to Subset 1.1, the accuracy dropped to 69%, the lowest accuracy of any label for any subsection from Subset 1 used to fit the neural network.

5.3.2 Subset 2: Generalisability with respect to Pattern ID

In order to explore the generalisability of the neural network, it was tested how well the network was able to predict the similarity of a crack pattern ID not used for fitting. With this aim, eight subsections of data were defined under Subset 2, as seen in Table 5.4. In each of these subsections of data, one Pattern ID was left out from the fitting of the neural network. This neural network was then tested with only the image pairs that contained at least one image of the crack pattern type that was left out from the fitting. In each such subsection of data, the annotations given by all the raters (Rater IDs 2, 3, 4, 20, 21, 22, and 25) for all the images generated in the computational physics-based approach were used. Additionally, all the aspect ratios and geometries were included in all the fitting and test sets.

CRACK PATTERN SIMILARITY LABEL It is seen in Fig. 5.28 that the neural network generally had a poor degree of accuracy in its ability to predict the crack pattern similarity label. The highest accuracy was only, 56%, obtained for Subset 2.1 when

Table 5.4: Details of the Subset 2 of the labelled data that was annotated by all the raters from the computational physics-based approach used to fit and test the neural network.

Subset No	Pattern ID used		Number of image pairs used	
	Fitting	Testing	Fitting	Testing
2.1	20, 21, 23, 24, 101, 102, 103	18	369	131
2.2	18, 21, 23, 24, 101, 102, 103	20	370	130
2.3	18, 20, 23, 24, 101, 102, 103	21	389	111
2.4	18, 20, 21, 24, 101, 102, 103	23	392	108
2.5	18, 20, 21, 23, 101, 102, 103	24	383	117
2.6	18, 20, 21, 23, 24, 102, 103	101	399	101
2.7	18, 20, 21, 23, 24, 101, 103	102	407	93
2.8	18, 20, 21, 23, 24, 101, 102	103	398	102

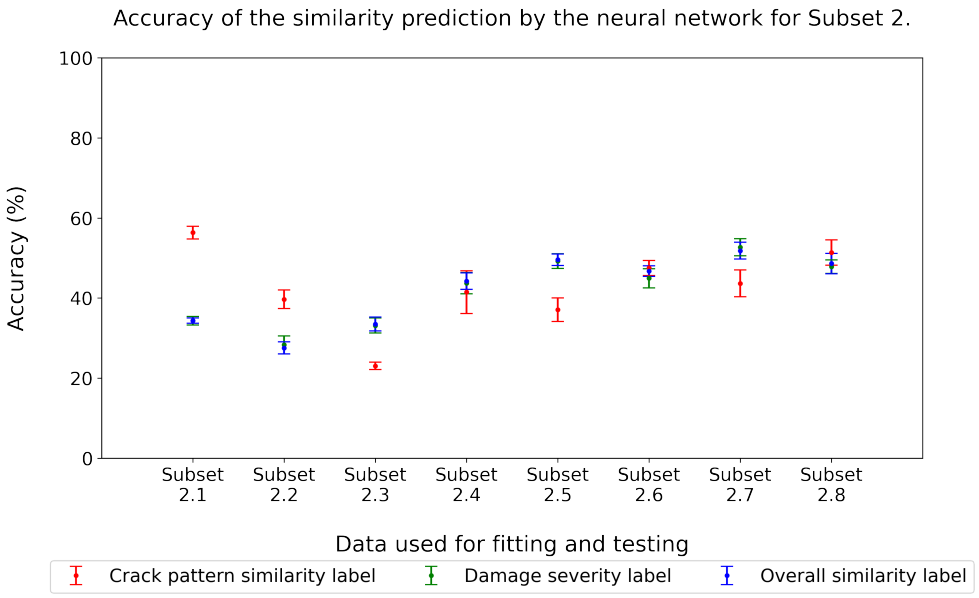


Figure 5.28: Summary of accuracy of the labels with the neural network fitted and tested to Subset 2.

Pattern ID 18 was left out of the fitting sample but included in the testing sample. The lowest accuracy for this label was 23%, obtained for Subset 2.3. This was obtained when the neural network attempted to predict the crack pattern similarity label for Pattern ID 21 without using it in the fitting step. The accuracy of the neural network to similarly predict the crack pattern similarity label for all the other IDs was about ∼40-50%. Pattern IDs 18, 20, and 21 (Subsets 2.1, 2.2, and 2.3) have horizontal crack orientations, while Pattern IDs 23 and 24 (Subsets 2.4 and 2.5) have vertical crack orientations. There seemed to be no recognisable trend in the prediction accuracy of the neural network when crack patterns of similar orientation (or shape) were left out of the fitting procedure.

DAMAGE SEVERITY LABEL The damage severity label was also seen to perform very poorly in Fig. 5.28. Subset 2.7 had the highest accuracy for this label in this subsection at 53%, achieved when Pattern ID 102 was dropped out of the fitting sample but included in the testing sample. It was also seen that Subset 2.2 had the lowest label accuracy of 28% when the neural network tried to predict the damage severity label for Pattern ID 20 without using it in the fitting phase.

OVERALL SIMILARITY LABEL The accuracy of the overall similarity label showed similar performance to that of the damage severity label, as seen in Fig. 5.28. The highest prediction accuracy of 52% for this label was seen again for Subset 2.7, as with the damage severity label. Similarly, the lowest prediction accuracy was only 27%, seen in Subset 2.2.

5.3.3 Subset 3: Generalisability with respect to façade geometry

Another question regarding the generalisability of the neural network was if it can handle different geometries of the masonry façades. Five subsections of data were defined under Subset 3 in order to investigate this, as seen in Table 5.5. In subsections 3.1, 3.2, 3.3, and 3.4, one out of the four aspect ratios of the masonry façades generated in this project was left out in the fitting procedure of the neural network. This neural network was then tested with image pairs of which at least one was of the aspect ratio that was left out from the fitting. Additionally, one more subsection was defined where the fitting of the neural network was done with masonry

façades without openings, and it was tested with image pairs that contain at least one façade with an opening. This was subset 3.5. In each such subsection of data, the annotations given by all the raters (Rater IDs 2, 3, 4, 20, 21, 22 and 25) for all the images generated in the computational physics-based approach were used. Additionally, all the Pattern IDs were included in all the fitting and test sets.

CRACK PATTERN SIMILARITY LABEL It is seen in Fig. 5.29 that while the crack pattern similarity labels perform relatively poorly with respect to accuracy, this label had a higher accuracy than the other labels in almost all subsections, except for Subset 3.4. The highest accuracy, 61%, for the crack pattern label was obtained by Subset 3.1, where fitting was done by leaving out the aspect ratio of 1:1 and tested image pairs that included it. The lowest accuracy for this label was 36%, obtained for Subset 3.5, when the neural network was fitted to façades with no openings and tested with façades with openings.

DAMAGE SEVERITY LABEL The similarity prediction accuracy of the damage severity label appeared to be performing poorly as well in Fig. 5.29. Similar to the crack

Table 5.5: Details of the Subset 3 of the labelled data that was annotated by all the raters from the computational physics-based approach used to fit and test the neural network.

Subset No	Aspect ratio of the façades used		Type of geometry o fthe façades used (with/without opening)		Number of image pairs used	
	Fitting	Testing	Fitting	Testing	Fitting	Testing
3.1	3:2, 2:1, 10:3	1:1	Façades with and without openings	Façades with and without openings	313	187
3.2	1:1, 2:1, 10:3	3:2	Façades with and without openings	Façades with and without openings	308	192
3.3	1:1, 3:2, 10:3	2:1	Façades with and without openings	Façades with and without openings	308	192
3.4	1:1, 3:2, 2:1	10:3	Façades with and without openings	Façades with and without openings	316	184
3.5	1:1, 3:2, 2:1, 10:3	1:1, 3:2, 2:1, 10:3	Façades without openings	Façades with openings	234	267

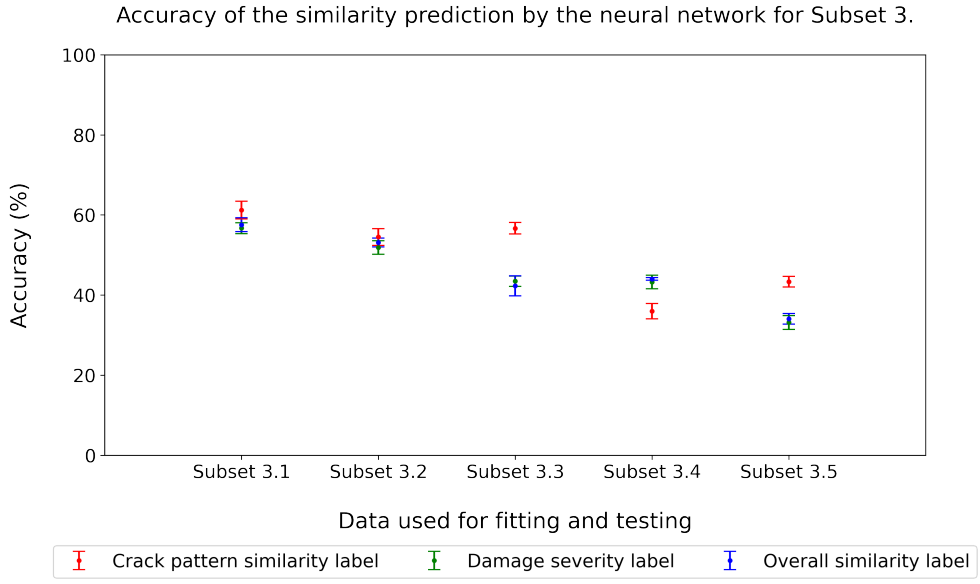


Figure 5.29: Summary of the accuracy of the labels with the neural network fitted and tested to Subset 3.

pattern similarity label, Subset 3.1 had the highest accuracy for this label in this subsection at 57% and Subset 3.5 had the lowest label accuracy of 33%. In Subset 3.4, when the label is fitted to all aspect ratios except 10:3 and fitted to it, this label was seen to perform better than the crack pattern similarity label.

OVERALL SIMILARITY LABEL The accuracy of the overall similarity label while poor, almost exactly mimicked the behaviour of the damage severity label. This is seen in Fig. 5.29. The highest prediction accuracy for this label was seen again for Subset 3.1, as with the damage severity label, with an accuracy of 58%. Similarly, the lowest prediction accuracy was only 34%, seen in Subset 3.5.

5.3.4 Subset 4: Generalisability with respect to statistically generated images and real-world images

Finally, to see how well the neural network was able to predict the similarity of real-world images when fitted to images generated in the computational physics-based approach, Subset 4 was defined as shown in Table 5.6. The data used for fitting was chosen based on the expertise level of the raters and their agreement with the standard rater. These were then tested with images generated from all three approaches in subsections 4.1, 4.2, and 4.3. Additionally, Subsets 4.4, 4.5, and 4.6 were defined such that the neural network was tested with image pairs that contain at least one real-world crack pattern. All the Pattern IDs, aspect ratios and geometries in the computational physics-based approach were used in each subset used to fit the neural network. Subsets 4.4, 4.5, and 4.6 were run 10 times per label in the neural network due to the large standard deviations seen.

CRACK PATTERN SIMILARITY LABEL A very poor performance was seen for the neural network when tested with images generated from all three approaches for this label, as seen in Fig. 5.30. The highest accuracy, 33%, for the crack pattern label was obtained by Subset 4.4, where the neural network was fitted to finite element images annotated by all the raters and tested with image pairs that contain at least one real-world image. The lowest accuracy for this label was 23% when the same data was tested on a neural network fitted to annotations from the standard rater alone. Both these subsets also had very large standard deviations, making these accuracies less reliable.

DAMAGE SEVERITY LABEL In Fig. 5.30, the damage severity label's similarity prediction accuracy looked to be weak as well. Subset 4.2 had the highest accuracy for this label in this subsection at 50% and Subset 4.5 had the lowest label accuracy of 30%. This label was also seen to have a very large standard deviation for the prediction accuracy, hinting that these values were not very reliable.

OVERALL SIMILARITY LABEL While the overall similarity label's accuracy was weak, it closely resembled the behaviour of the damage severity label. This is seen in Fig. 5.30. The highest prediction accuracy for this label was seen again for Subset 4.2, as with the damage severity label, with an accuracy of 48%. The lowest prediction accuracy was only 25%, seen in Subset 4.6. This label also had a very significant standard deviation for prediction accuracy, implying that these numbers were questionable.

Table 5.6: Details of the Subset 4 of the labelled data that was used to fit and test the neural network.

Subset No	Approach used to generate the image pairs		Criteria used to choose raters		Raters included in the subset		Number of image pairs used	
	Fitting	Testing	Fitting	Testing	Fitting	Testing	Fitting	Testing
4.1	Computational physics-based approach	All three approaches[3]	All raters	All raters[4]	2, 3, 4, 20, 21, 22, 25	2, 3, 4	500	50
4.2	Computational physics-based approach	All three approaches[3]	Standard rater	All raters[4]	4	2, 3, 4	300	50
4.3	Computational physics-based approach	All three approaches[3]	Raters with good agreement with the standard rater	All raters[4]	2, 3, 4, 20, 21	2, 3, 4	500	50
4.4	Computational physics-based approach	Subset from all three approaches[5]	All raters	All raters[4]	2, 3, 4, 20, 21, 22, 25	2, 3, 4	500	15
4.5	Computational physics-based approach	Subset from all three approaches[5]	Standard rater	All raters[4]	4	2, 3, 4	300	15
4.6	Computational physics-based approach	Subset from all three approaches[5]	Raters with good agreement with the standard rater	All raters[4]	2, 3, 4, 20, 21	2, 3, 4	500	15

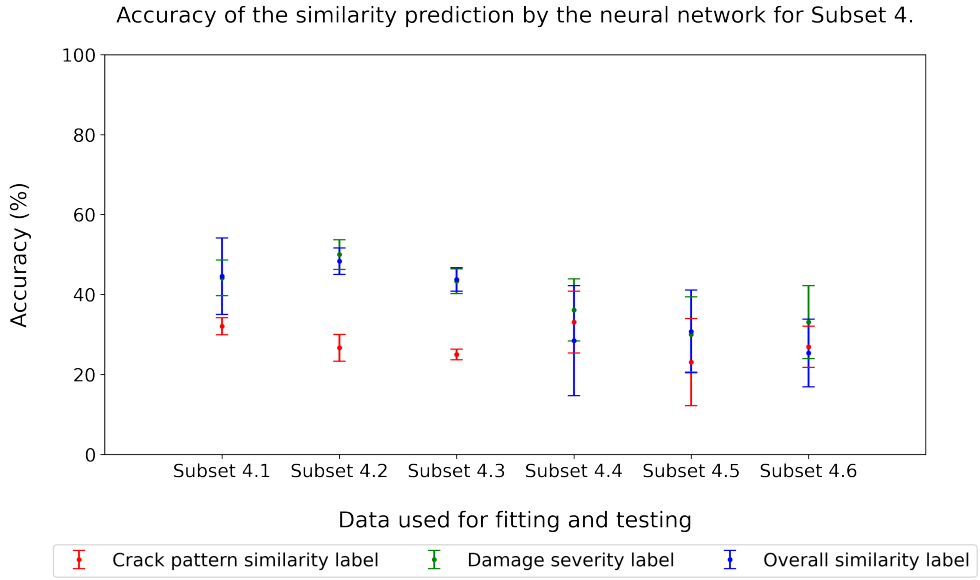


Figure 5.30: Summary of the accuracy of the labels with the neural network fitted to and tested to Subset 4.

3 The subset contains image pairs generated from the statistics-based approach and real-world images, from the computational physics-based approach and real-world images.

4 Only Rater IDs 2, 3, and 4 assessed the image pairs generated from all three approaches. Thus, *all raters* here refers to only these three raters.

5 The subset contains image pairs generated from the statistics-based approach and real-world images, as well as image pairs and from the computational physics-based approach and real-world images.

5.4 DISCUSSION

The similarity of masonry crack pattern images is currently assessed by masonry experts and structural engineers. The findings of this thesis corroborate the previous research done at TNO, showing that it is indeed possible to automate the similarity assessment of crack pattern images using convolutional neural networks. However, the neural network developed in this thesis shows very poor generalisability when fitted and tested on the current data set. There is potential to improve this by making changes to the architecture of the neural network [1]. Additionally, using larger quantities of more balanced data to train the neural network may improve the performance. This data can be expanded to image pairs that include more crack pattern types and a more uniform distribution of similarity ratings. Generating tens of thousands of image pairs using the computational physics-based approach to fit and test this neural network will take a few weeks worth of effort on a moderately powered workstation [6]. Collecting hundreds, or even thousands of real-world images will be a few weeks of effort as well; however in contrast with the previous point it would involve almost solely human labour. Furthermore, getting raters to annotate this large quantity of image pairs can take up to weeks or months. The judgements of masonry experts were instrumental in the training of the neural network to a great extent. The low prediction accuracy values of the fitted neural networks in the generalisabilty experiments suggest that the neural network may need to be fitted to a much larger sample of crack patterns that cover all the relevant cases to obtain a higher prediction accuracy. Additionally, the high inter-rater variability seen in assigning the labels to crack pattern image pairs indicated that even a *perfect* neural network need not work well, as it would be fitted to very inconsistent data. Hence, the neural networks cannot completely replace the masonry experts. At best, it can facilitate easier access to the experts' opinion on the similarity assessment of crack pattern images. If neural networks fitted to expert opinions are sufficiently generalisable and able to achieve accuracy greater than 90% in predicting the similarity labels, they can be used as a less expensive and faster substitute to masonry experts in crack pattern similarity assessment tasks.

6 Any work station with a CPU of at least 8 cores and a RAM of at least 32 GB.

6 CONCLUSIONS

The main research goal of this thesis was to explore how accurately a neural network fitted to data from a computational physics-based approach (finite element analysis) can predict the similarity of crack pattern image pairs. Four sub-research questions were provided to give an insight into the research aim and methodology. Based on the results presented in the previous chapter these questions are answered here. Moreover, recommendations for future research are also provided.

6.1 RETURNING TO THE RESEARCH QUESTIONS

6.1.1 Main Research Question

How accurately can a neural network – fitted to data generated from finite element models – predict masonry crack pattern similarities?

The crack pattern image pairs generated from the computational physics-based approach were assessed for their similarity by seven masonry experts. Three similarity labels were used to perform this assessment: the crack pattern similarity label, the damage severity label, and the overall similarity label. The neural network fitted to 375 image pairs with expert-given labels was able to predict the similarity with reasonably good accuracy. The neural network assigned the crack pattern similarity label with an accuracy of 87%, the damage severity label with an accuracy of 82% and the overall similarity label with an accuracy of 69%. The degree of accuracy that can be achieved by the neural network depends greatly on the type of the data used to fit the neural networks, i.e. which raters were used to establish the ground truth, which crack pattern archetypes and geometries were used in the fitting set and the testing set, as well as the sources used to generate the crack pattern image pairs. When the neural network was only fitted to the labelled data from the standard rater, it showed an average increase of 12% in the accuracy of all three labels. The generalisability experiments indicated a significant drop in accuracy for all three label groups, by an average of 47%.

6.1.2 Sub Research Questions

What criteria are used by masonry experts to assess the similarity in crack pattern image pairs? What is the agreement between the masonry experts in using these criteria?

From conversations with masonry experts, it was deduced that the way an expert interprets the similarity of a crack pattern image pair greatly depends on the nature of the task at hand. Following this, the three similarity labels that attempted to include the characteristics that masonry experts use for visual similarity assessment were defined: crack pattern similarity label, damage severity label, and overall similarity label. The crack pattern similarity label takes into account the similarity in the cracking mechanism and the similarity of the geometrical shape of the crack pattern. The damage severity label captures the similarity in the severity by accounting for the crack length and width. A third overall similarity label was used to see how ma-

sonry experts combined the previous two labels. An inter-rater agreement analysis was conducted by computing Krippendorff's alpha on the labelled data generated from the computational physics-based approach. This revealed that there was insufficient agreement among all the raters regarding all the three labels. It was also seen that the highest agreement among all the raters was for the overall similarity label. This was followed by the crack pattern similarity label, and finally, the damage severity label. The corresponding α values were 0.73, 0.68, and 0.56, respectively. It was also observed that, in general, there was higher agreement among raters when they were grouped by their experience with masonry structure, i.e. there was more agreement among expert raters than among MSc student raters. The consistency of a chosen standard rater, Rater ID 4, with themself, was also tested using Lin's Concordance Correlation Coefficient (CCC) by having them assess the same set of image pairs twice. It was seen that the standard rater was most consistent for the crack pattern similarity label, followed by the overall similarity label and finally the damage severity label with Lin's CCC values of 0.96, 0.87 and 0.72, respectively. Following this, it was decided that if any rater was able to have a Lin's CCC value of at least 0.70 with Rater ID 4, for at least one out of the three similarity labels, they could be considered to be as reliable as Rater ID 4. It is also to be noted that as the similarity ratings of each label for each image pair was averaged out before fitting the neural network, this high inter-rater variability was not transferred to the neural networks. Thus, the performance of the neural network was not correlated to the Krippendorff's alpha values for any label.

What are the parameters that should be varied in the finite element models to generate different crack patterns (e.g.: geometry, loading, material properties)?

In order to make the generation of the finite element models as realistic and generalisable as possible, the geometry, material properties, and loading conditions were varied with each model. The variation of the geometry included variation in height, width, thickness, and shape of the façade, as well as a variation in the number, size, shape, and location of openings in the façade. Four aspect ratios of the masonry façades were considered, along with different thicknesses for each aspect ratio. Additionally, three types of openings were considered for the façades with openings. As an infinite number of possibilities exist in choosing the location and combination of these openings, ten different orientations of the openings were chosen for each crack pattern type. The variation in material properties of the masonry was realised through the generation of random fields for the Young's modulus and tensile strength. All the other properties of masonry and the soil-structure interaction were kept constant. The location and magnitude of the crack-inducing loads were also varied for each crack pattern in this thesis.

How generalisable is the fitted neural network in dealing with different masonry façades?

The fitted neural networks were tested for their generalisability by evaluating its capability to predict the similarity of a crack pattern image pair that was not used to fit the neural network. Three types of generalisability experiments were used to (*i*) predict the crack pattern similarity involving a crack archetype that was excluded from the fitting procedure; (*ii*) predict the crack pattern similarity involving a façade geometry that was excluded from the fitting procedure; and (*iii*) predict the crack pattern similarity of images generated from the statistics-based approach, computational physics-based approach and the real world images when the network was fitted to images from only the computational physics-based approach. When the neural network predicted the similarity of a Pattern ID that was not used in the fitting, on an average, the accuracy was 50% for the crack pattern similarity label, and 43% for the damage severity and overall similarity label. When the neural network predicted the crack pattern on a façade with a different geometry than the ones

used in the fitting process, the average accuracy of prediction for the crack pattern similarity, damage severity, and overall similarity labels were 50%, 46%, and 46%, respectively. Finally, when the neural network was tested with image pairs generated from the statistics-based approach, computational physics-based approach, and the real world images when the network was fitted to images from only the computational physics-based approach, the average accuracy of the crack pattern similarity label was 28%, the damage severity label was 39% and the overall similarity label was 36%. Thus, it can be concluded that, in general, the fitted neural networks shows very poor generalisability. This could be due to insufficiency of the data set used to fit the neural networks.

How do the fitted neural networks perform on real-world crack patterns?

Over 50 images of settlement induced crack pattern images were collected from the Dutch city of Gouda. However, only four of them have similar crack pattern type and façade geometry as the crack patterns generated in this thesis through the computational physics-based approach. Fifteen image pairs were created that compare these four real-world images to images generated from the statistics-based approach. These images were then tested on a neural network fitted to data from the computational physics-based approach. The neural network predicts the crack pattern similarity label, the damage severity label, and the overall similarity label with an accuracy of 28%, 33%. and 28%. Thus, the neural network performed very poorly on real-world crack patterns with known real world damage causes. It should be noted that the sample of real-world images used for testing in this thesis was very small. Hence, these results and conclusions are to be taken at best as indicative.

6.2 FURTHER RECOMMENDATIONS

This thesis focused on masonry damage caused by soil subsidence, as well as a few cases of damage caused by temperature changes and out-of-plane stresses. The finite element models created were limited to clay brick masonry material qualities and a soil-structure interface that replicates the behaviour produced by a structure sitting on clayey soil. Using the smeared crack model, the masonry was described as a non-linear continuum, and the soil-structure interface was modelled using linear properties. The images of the simulated crack patterns were composed of only rectangular 2D shapes and four aspect ratios. In the façades with openings, only one type of door and two types of windows were considered as openings. Thus, the sample of crack pattern image pairs generated through the computational physics-based approach was not entirely reflective of the façades seen in real life. This affects the generalisability of the neural networks. Keeping in mind the main research question and the goal of the research work at TNO, the following development can be considered in further investigations:

- Pattern IDs 30, 31, 32 and 201 can be developed with the computational physics-based approach. Additional crack pattern types can also be included in future studies, such as, crack patterns generated on different façade geometries due to different damage causes.
- More focus can be given to cracks induced by in-plane loads. These are seen to be more recurrent in buildings that are not subjected to earthquake-induced damage.
- The finite element models can be expanded to include more variety of masonry types and soil properties in order to make the finite element models more generalisable.

- It would also be interesting to study the crack patterns generated from multiple or different combinations of the crack-inducing loads in the finite element models.
- Currently, the real-world images are converted to a pixel format manually. It is expected that the technology to achieve this through automation will be available in the future. A larger variety of real-world image pairs can be collected and used for the similarity assessment in this scenario.
- Future iterations of the similarity assessment performed by raters can include a larger number of experts with a more versatile background.
- To help improve the inter-rater variability, more clear instructions can be given to the raters. This could even include some background theory into the cracking mechanisms of common crack pattern types. A mandatory trial attempt of the crack pattern similarity assessment could also be done to help the raters achieve consistency.
- The neural network can be fitted to the images generated from the statistics-based approach. A comparison of the similarity assessment can be performed to see if the additional computational effort of the computational physics-based approach is rewarded with a higher accuracy in similarity, especially with regard to similarity assessments including real-world images.
- A comparison of the similarity between image pairs generated from the computational physics-based approach and the real-world images can be explored in the future studies.
- The ground truth of the neural network can be established by using a different means other than by averaging the similarity ratings given by the masonry expert as this leads to a loss of information. For example, a weighted average or a median can be employed.

BIBLIOGRAPHY

[1] Kelvin Kleijn. *Automation of Crack Assessment in Masonry Facades using Deep Learning*. M.Sc. thesis, Leiden Institute of Advanced Computer Science, 2022.

[2] Wilson Ricardo Leal da Silva and Diogo Schwerz de Lucena. Concrete cracks detection based on deep learning image classification. *Proceedings*, 2(8):489, 2018.

[3] Krisada Chaiyasarn, Wasif Khan, Luqman Ali, Mayank Sharma, Daniel Brackenbury, and Matthew Dejong. Crack detection in masonry structures using convolutional neural networks and support vector machines. *Proceedings of the 35th International Symposium on Automation and Robotics in Construction (ISARC)*, 2018.

[4] Cao Vu Dung and Le Duc Anh. Autonomous concrete crack detection using deep fully convolutional neural network. *Automation in Construction*, 99:52–58, 2019.

[5] Rebecca Napolitano and Branko Glisic. Methodology for diagnosing crack patterns in masonry structures using photogrammetry and distinct element modeling. *Engineering Structures*, 181:519–528, 2019.

[6] Arpád Rózsás, Arthur Slobbe, Wyke Huizinga, Maarten Kruithof, and Giorgia Giardina. Development of a neural network embedding for quantifying crack pattern similarity in masonry structures. *12th International Conference on Structural Analysis of Historical Constructions*, 2020.

[7] Ilse de Vent. Prototype of a diagnostic decision support tool for structural damage in masonry. *Supplement to PhD Thesis of Ilse de Vent*, 2011.

[8] Antonella Saisi and Carmelo Gentile. Investigation strategy for structural assessment of historic towers. *Infrastructures*, 5(12):106, 2020.

[9] Mitchell J. Hallee, Rebecca K. Napolitano, Wesley F. Reinhart, and Branko Glisic. Crack detection in images of masonry using cnns. *Sensors*, 21(14):4929, 2021.

[10] Kiarash M. Dolatshahi and Katrin Beyer. Stiffness and strength estimation of damaged unreinforced masonry walls using crack pattern. *Journal of Earthquake Engineering*, pages 1–20, 2019.

[11] Amna Rekik and Frédéric Lebon. Identification of the representative crack length evolution in a multi-level interface model for quasi-brittle masonry. *International Journal of Solids and Structures*, 47(22-23):3011–3021, 2010.

[12] Reza Akbari. Crack survey in unreinforced concrete or masonry abutments in short- and medium-span bridges. *Journal of Performance of Constructed Facilities*, 27(2):203–208, 2013.

[13] J B Burland and C P Wroth. *Settlement of Buildings and Associated Damage*. British Geotechnical Society, 1974.

[14] Marco D. Boscardin and Edward J. Cording. Building response to excavation-induced settlement. *Journal of Geotechnical Engineering*, 115(1):1–21, 1989.

[15] Gorgia Giardina. *Modelling of settlement induced building damage*. Delft University of Technology, 2013.

[16] Paul A. Korswagen, Michele Longo, Edwin Meulman, and Jan G. Rots. Crack initiation and propagation in unreinforced masonry specimens subjected to repeated in-plane loading during light damage. *Bulletin of Earthquake Engineering*, 17(8):4651–4687, 2019.

[17] A. Ellenberg, A. Kontsos, I. Bartoli, and A. Pradhan. Masonry crack detection application of an unmanned aerial vehicle. *Computing in Civil and Building Engineering (2014)*, 2014.

[18] Srinidhi Lakshmish Kumar, H.B. Aravind, and Nabil Hossiney. Digital image correlation (dic) for measuring strain in brick masonry specimen using ncorr open source 2d matlab program. *Results in Engineering*, 4:100061, 2019.

[19] Tiago Ramos, André Furtado, Shayan Eslami, Sofia Alves, Hugo Rodrigues, António Arêde, Paulo J. Tavares, and P.M.G.P. Moreira. 2d and 3d digital image correlation in civil engineering – measurements in a masonry wall. *Procedia Engineering*, 114:215–222, 2015.

[20] Benjamín Torres, Francisco B. Varona, F. Javier Baeza, David Bru, and Salvador Ivorra. Study on retrofitted masonry elements under shear using digital image correlation. *Sensors*, 20(7):2122, 2020.

[21] M. Didier, G. Abbiati, F. Hefti, M. Broccardo, and B. Stojadinovic. *Damage Quantification in Plastered Unreinforced Masonry Walls Using Digital Image Correlation*. 10th Australasian Masonry Conference, 2018.

[22] Á. Rózsás. *A literature overview on quantifying the similarity of masonry crack patterns*. TNO report. 2021.

[23] Zhenheng Kong. *Probabilistic identification of pile foundation settlements from cracked masonry facades*. Internship Report. 2018.

[24] L Wasserman. Optimal transport and wasserstein distance. *Statistical Methods for Machine Learning*, pages 36–708, 2019.

[25] Mustafizur Rahman, Md. Rafiul Hassan, and Rajkumar Buyya. Jaccard index based availability prediction in enterprise grids. *Procedia Computer Science*, 1(1):2707–2716, 2010.

[26] R.C. Veltkamp. Shape matching: similarity measures and algorithms. *Proceedings International Conference on Shape Modeling and Applications*.

[27] Dushan Chakmakov and Emilija Celakoska. Estimation of curve similarity using turning functions. *International Journal of Applied Mathematics*, 15:403–416, 01 2004.

[28] S.K Parui and D Dutta Majumder. Shape similarity measures for open curves. *Pattern Recognition Letters*, 1(3):129–134, 1983.

[29] Remco C. Veltkamp and Michiel Hagedoorn. Shape similarity measures, properties and constructions. *Lecture Notes in Computer Science*, pages 467–476, 2000.

[30] A Zucchini and P.B Lourenco. A micro-mechanical model for the homogenisation of masonry. *International Journal of Solids and Structures*, 39(12):3233–3255, 2002.

[31] V. Giamundo, V. Sarhosis, G.P. Lignola, Y. Sheng, and G. Manfredi. Evaluation of different computational modelling strategies for the analysis of low strength masonry structures. *Engineering Structures*, 73:160–169, 2014.

[32] A.D. Tzamtzis and P.G. Asteris. *FINITE ELEMENT ANALYSIS OF MASONRY STRUCTURES: PART I- REVIEW OF PREVIOUS WORK*. 2003.

[33] P.B. Lourenco. *Computational strategies for masonry structures*. Delft University of Technology, 1996.

[34] Jan G Rots. *Computational modeling of concrete fracture*. Delft University of Technology, 1988.

[35] Shenghan Zhang, Seyedeh Mohadeseh Taheri Mousavi, Nicolas Richart, Jean-François Molinari, and Katrin Beyer. Micro-mechanical finite element modeling of diagonal compression test for historical stone masonry structure. *International Journal of Solids and Structures*, 112:122–132, 2017.

[36] Antonio Maria D'Altri, Vasilis Sarhosis, Gabriele Milani, Jan Rots, Serena Cattari, Sergio Lagomarsino, Elio Sacco, Antonio Tralli, Giovanni Castellazzi, and Stefano de Miranda. Modeling strategies for the computational analysis of unreinforced masonry structures: Review and classification. *Archives of Computational Methods in Engineering*, 27(4):1153–1185, 2019.

[37] Paul J Fanning and Thomas E Boothby. Three-dimensional modelling and full-scale testing of stone arch bridges. *Computers & Structures*, 79(29-30):2645–2662, 2001.

[38] Leidy Bejarano-Urrego, Els Verstrynge, Giorgia Giardina, and Koen Van Balen. Crack growth in masonry: Numerical analysis and sensitivity study for discrete and smeared crack modelling. *Engineering Structures*, 165:471–485, 2018.

[39] Giorgia Giardina, Anne V. van de Graaf, Max A.N. Hendriks, Jan G. Rots, and Alessandra Marini. Numerical analysis of a masonry façade subject to tunnelling-induced settlements. *Engineering Structures*, 54:234–247, 2013.

[40] Mark J. Masia, Peter W. Kleeman, and Robert E. Melchers. Modeling soil/structure interaction for masonry structures. *Journal of Structural Engineering*, 130(4):641–649, 2004.

[41] Yanming Guo, Yu Liu, Ard Oerlemans, Songyang Lao, Song Wu, and Michael S. Lew. Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48, 2016.

[42] Weibo Liu, Zidong Wang, Xiaohui Liu, Nianyin Zeng, Yurong Liu, and Fuad E. Alsaadi. A survey of deep neural network architectures and their applications. *Neurocomputing*, 234:11–26, 2017.

[43] Weidong Wang, Wenbo Hu, Wenjuan Wang, Xinyue Xu, Mengdi Wang, Youyin Shi, Shi Qiu, and Erol Tutumluer. Automated crack severity level detection and classification for ballastless track slab using deep convolutional neural network. *Automation in Construction*, 124:103484, 2021.

[44] Yu Zhang, G. C. Zhou, Yi Xiong, and M. Y. Rafiq. Techniques for predicting cracking pattern of masonry wallet using artificial neural networks and cellular automata. *Journal of Computing in Civil Engineering*, 24(2):161–172, 2010.

[45] Guangchun Zhou, Deng Pan, Xun Xu, and Yaqub M. Rafiq. Innovative ann technique for predicting failure/cracking load of masonry wall panel under lateral load. *Journal of Computing in Civil Engineering*, 24(4):377–387, 2010.

[46] Yiqing Liu and Justin K.W. Yeoh. Automated crack pattern recognition from images for condition assessment of concrete structures. *Automation in Construction*, 128:103765, 2021.

[47] Eissa Fathalla, Yasushi Tanaka, and Koichi Maekawa. Remaining fatigue life assessment of in-service road bridge decks based upon artificial neural networks. *Engineering Structures*, 171:602–616, 2018.

[48] Code.waag.org. Smart citysdk. https://code.waag.org/buildings/#52.0101,4.712,15. Last accessed on 08/11/2021.

[49] Alan C. Bovik. Basic gray level image processing. *The Essential Guide to Image Processing*, pages 43–68, 2009.

[50] Lawrence I-Kuei Lin. A concordance correlation coefficient to evaluate reproducibility. *Biometrics*, 45(1):255, 1989.

[51] Jennifer A. Zellers, Paul K. Commean, Ling Chen, Michael J. Mueller, and Mary K. Hastings. A limited number of slices yields comparable results to all slices in foot intrinsic muscle deterioration ratio on computed tomography and magnetic resonance imaging. *Journal of Biomechanics*, 129:110750, 2021.

[52] Wei Zhao, Liyue Shen, Bin Han, Yong Yang, Kai Cheng, Diego A.S. Toesca, Albert C. Koong, Daniel T. Chang, and Lei Xing. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology*Biology*Physics*, 105(2):432–439, 2019.

[53] G. B. McBirdie. *A Proposal for Strength of Agreement Criteria for Lin's Concordance Correlation Coefficient*. NIWA Client Report: HAM2005-062. 2005.

[54] Haldun Akoglu. User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18(3):91–93, 2018.

[55] Klaus H Krippendorff. *Content Analysis - 3rd Edition*. SAGE Publications, Inc, 2013.

[56] Klaus krippendorff. Measuring the reliability of qualitative text analysis data. *Quality & Quantity*, 38(6):787–800, 2004.

[57] Taskin Kavzoglu. Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software*, 24(7):850–858, 2009.

[58] Joan Pastor-Pellicer, Francisco Zamora-Martínez, Salvador España-Boquera, and María José Castro-Bleda. F-measure as the error function to train neural networks. *Advances in Computational Intelligence*, pages 376–384, 2013.

[59] M. Boonpichetvong and J. G. Rots. *SETTLEMENT DAMAGE OF MASONRY BUILDINGS IN SOFT-GROUND TUNNELLING*. Taylor & Francis, 2002.

[60] Waleed A. Al-Awsh, Naef A.A. Qasem, Omar S. Baghabra Al-Amoudi, and Mohammed A. Al-Osta. Experimental and numerical investigation on innovative masonry walls for industrial and residential buildings. *Applied Energy*, 276:115496, 2020.

[61] Natalino Gattesco and Ingrid Boem. Out-of-plane behavior of reinforced masonry walls: Experimental and numerical study. *Composites Part B: Engineering*, 128:39–52, 2017.

[62] R Pukl, M Jansta, and J Červenka. Spatial variability of material properties in nonlinear computer simulation. *Computational Modelling of Concrete Structures - Proceedings of EURO-C 2006*, pages 891–896, 2006.

[63] Jie Li, Mark G. Stewart, Mark J. Masia, and Stephen J. Lawrence. Spatial correlation of material properties and structural strength of masonry in horizontal bending. *Journal of Structural Engineering*, 142(11):04016112, 2016.

[64] Robin van der Have. Random fields for non-linear finite element analysis of reinforced concrete. 2015.

[65] L. Schueremans, D. Van Gemert, M.A. Maes, and L.J. Huyse. *Stochastic Homogenization Technique for Unreinforced masonry*. Numerical modelling of uncertainties. 1999.

[66] L Schueremans and D Van Gemert. *Predicting masonry properties from component properties using probabilistic techniques*. 2001.

[67] Rob Van Del' Pluijm. Non-Linear Behaviour of Masonry under Tension. *Nonlinear Behaviour of Masonry under Tension*, 42(1):25–54, 1997.

[68] Abdelmounaim Zanaz, Sylvie Yotte, Fazia Fouchal, and Alaa Chateauneuf. Young's modulus variability: Influence on masonry vault behavior. http://waset.org/publication/young-s-modulus-variability-influence-on-masonry-vault-behavior/10002228. *World Academy of Science, Engineering and Technology, International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering*, 9:1098–1104, 01 2015.

[69] Joanna Zięba, Izabela Skrzypczak, and Lidia Buda-Ożóg. Calibration of partial safety factors of sample masonry structures. *Materials*, 14(17):5003, 2021.

[70] H P Backes. *Tensile strength of masonry*. 1985.

[71] Nebojša Mojsilović. Tensile strength of clay blocks: An experimental study. *Construction and Building Materials*, 25(11):4156–4164, 2011.

[72] Jan G. Rots, Paul Korswagen Eguren, and Michele Longo. *Computational modelling checks of masonry building damage due to deep subsidence*. 2021.

[73] DIANA FEA. Guidelines to set-up material parameters of interface elements. https://dianafea.com/manuals/d101/GeoTech/node22.html. Last accessed on 26/09/2021.

[74] DIANA FEA. Curved shell elements. https://dianafea.com/manuals/d102/Theory/Theorych23.html. Last accessed on 09/09/2021.

[75] DIANA FEA. Finite element types. https://dianafea.com/manuals/d96/Theory/node46.html#SECTION0523417000000000000000. Last accessed on 09/09/2021.

[76] DIANA FEA. Cl24i. https://dianafea.com/manuals/d101/ElmLib/node352.html. Last accessed on 09/09/2021.

[77] F. Messali, R. Esposito, G. J. P. Ravenshorst, and J. G. Rots. Experimental investigation of the in-plane cyclic behaviour of calcium silicate brick masonry walls. *Bulletin of Earthquake Engineering*, 18(8):3963–3994, 2020.

[78] Vishal Sharma, Mamta Sharma, Shevita Pandita, Jasmeet Kour, and Namrata Sharma. Application of geographic information system and remote sensing in heavy metal assessment. *Heavy Metals in the Environment*, pages 191–204, 2021.

[79] P Schubertl. Tensile and Flexural Strength of Masonry-Influences, Test Methods, Test Results. *10th IB2 MaC*, 7 1994.

[80] Shenghan Zhang, Nicolas Richart, and Katrin Beyer. Numerical evaluation of test setups for determining the shear strength of masonry. *Materials and Structures 2019 51:4*, 51(4):1–12, 8 2018.

[81] L M Heffler, M G Stewart, M J Masia, and M R S Correa. Spatial Correlation of Flexural Bond Strength for Masonry Walls: An Experimental and Statistical Study. *Masonry International*, 21(2):57–70, 2008.

[82] Robert Kissell and Jim Poserina. Advanced math and statistics. *Optimal Sports Math, Statistics, and Fantasy*, pages 103–135, 2017.

[83] MIT OpenCourseWare. Relationships between mean and variance of normal and lognormal distributions. https://ocw.mit.edu/courses/civil-and-environmental-engineering/1-151-probability-and-statistics-in-engineering-spring-2005/lecture-notes/briefnts8_relnl.pdf. Last accessed on 09/09/2021.

A LIST OF RATERS WHO ATTEMPTED THE SIMILARITY ASSESSMENT OF MASONRY CRACK PATTERNS

Ajithkumar Pillai, Krishna
Faculty of Civil Engineering and Geosciences, TU Delft

Bianchini, Nicoletta
Department of Civil Engineering, University of Minho

Brouwers, Raymond
Faculty of Civil Engineering and Geosciences, TU Delft

Foroughnia, Mahtab
Faculty of Civil Engineering and Geosciences, TU Delft

Giardina, Giorgia
Faculty of Civil Engineering and Geosciences, TU Delft

Heinemann, Herdis
Expertise Group for Structural Reliability, TNO

Jacob, Joseph
Faculty of Civil Engineering and Geosciences, TU Delft

Lee, Ka Ho
Faculty of Civil Engineering and Geosciences, TU Delft

Macchiarulo, Valentina
Faculty of Civil Engineering and Geosciences, TU Delft

Malinowska, Dominika
Department of Architecture & Civil Engineering, University of Bath

Mehrotra, Anjali
Faculty of Civil Engineering and Geosciences, TU Delft

Mithapelli, Nitesh
Faculty of Civil Engineering and Geosciences, TU Delft

Murano, Antonio
Department of Civil Engineering, University of Minho

Naldini, Silvia
Faculty of Architecture and the Built Environment, TU Delft

Palla, Pallavi
Faculty of Civil Engineering and Geosciences, TU Delft

Parisse, Francesco
Department of Civil Engineering, University of Minho

Paxinou, Katerina
Faculty of Civil Engineering and Geosciences, TU Delft

Priyadarshi, Richa
Faculty of Civil Engineering and Geosciences, TU Delft

Raghavan, Arjun
Faculty of Civil Engineering and Geosciences, TU Delft

Rózsás, Árpád
Expertise Group for Structural Reliability, TNO

Saravanakumar, Karthikeyan
Faculty of Civil Engineering and Geosciences, TU Delft

Shah, Anjali
Faculty of Civil Engineering and Geosciences, TU Delft

Shah, Manank
Faculty of Civil Engineering and Geosciences, TU Delft

Slobbe, Arthur
Expertise Group for Structural Reliability, TNO

Sushin, Harikrishnan
Faculty of Civil Engineering and Geosciences, TU Delft

van Hees, Rob
Faculty of Architecture and the Built Environment, TU Delft

van Hout, Ylana
Faculty of Civil Engineering and Geosciences, TU Delft

Warnaar, Martijn
Faculty of Civil Engineering and Geosciences, TU Delft

B IMPLEMENTATION OF SPATIAL VARIABILITY OF MASONRY PROPERTIES AS A RANDOM FIELD

The spatial variability of the Young's modulus and the tensile strength was implemented in the finite element models by generating them as random fields, as described in Section 4.2.2. The methodology used to generate these random fields is explained in this appendix.

- **Define an autocorrelation function:** An autocorrelation function defines the degree of similarity or correlation between points in the sample space [78]. Exponential and squared exponential functions are commonly used to define autocorrelation functions in engineering applications [64]. For the definition of the parametric masonry models in this project, a squared auto-correlation function, R_{xx}, was assumed as given in Eq. B.1 [64].

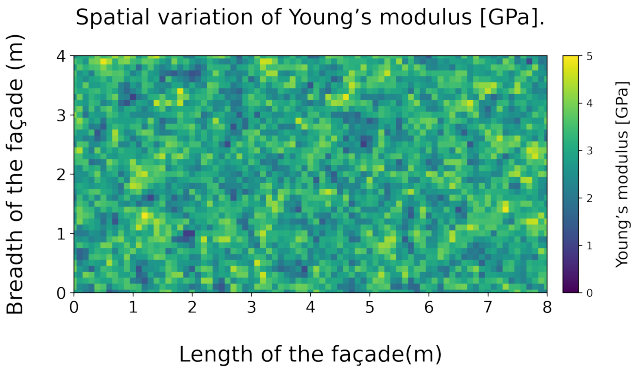
$$R_{xx} = \exp\left(-\frac{\Delta x^2}{L_c^2}\right) \tag{B.1}$$

Where,

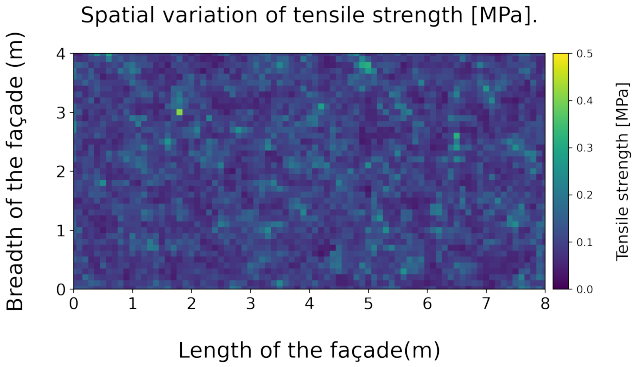
Δx = The distance between the nodes in the mesh.
L_c = The correlation length.

The correlation length indicates the distance beyond which the autocorrelation between two points in the field vanishes. The correlation length for material properties in concrete was seen to vary from 0.5 m to 5 m [64]. The correlation length of masonry properties has been studied far less. The tensile strength and the Young's modulus vary spatially along the volume of a masonry wall depending on the type of the masonry unit, the type of mortar used for bonding, shrinkage stresses and other age-related issues such as carbonation [79].

The correlation length of masonry could be taken as $1/16^{th}$ of the brick length or 12.5 mm [80]. A correlation length of comparative magnitude, such as 10 mm, would be representative of situations where the spatial correlation does not extend beyond the mortar joints. The correlation length had also been studied as a function of brick length which assumed that the correlation length of the material properties of masonry was affected by the adjacent bricks of each course [81]. Thus, a correlation of 20 cm, as shown in Fig. B.1a and Fig. B.1b, was representative of the situations where there was strong correlation only between individual masonry units. The correlation length for the spatial variation of Young's modulus and tensile strength in concrete structures usually are of larger magnitudes, such as 2-3 m. Such values are also sometimes used for masonry structures as the repetitive use of a single type of brick and mortar can lead to a strong correlation over larger lengths [81]. In this study, the correlation length of the masonry was assumed to be 0.2 m. The other material properties are as stated earlier in Table 4.1



(a) Spatial variation of Young's modulus [GPa].



(b) Spatial variation of tensile strength [MPa].

Figure B.1: Spatial variation of material properties are shown for a random field with a correlation length of 20 cm.

- **Define a marginal distribution for the material properties**: The Young's modulus was assumed to follow a normal distribution [66] and the tensile strength was assumed to follow a lognormal distribution [66, 71]. To generate a random field for the Young's modulus and the tensile strength of masonry, the mean and variance of the random distribution were required. The mean values of the Young's modulus and the tensile strength were taken from previous studies of using a continuum model for modelling the properties of masonry structures [39]. The coefficient of variation (COV) of the Young's modulus [66–69] and the tensile strength [65, 67, 70, 71] of the masonry were approximated from experimental research of masonry units of similar properties. This was used to calculate the standard deviation of these properties as well. These values are given in Table 4.1.

- **Generate a pair-wise distance matrix using the centre points of the finite element:** The centre points of the finite elements were taken as (x_i, y_i). Using these points, a pair-wise distance matrix was generated. A distance matrix is a square matrix that contains the pair-wise distance between any two points in the field. The size of this matrix is dependent on the mesh size.

- **Compute the correlation matrix:** A correlation matrix is a square matrix that contains the correlation between any two points. This correlation was obtained using the distance matrix and the autocorrelation function. As each element always correlates with itself perfectly, the diagonal elements of the correlation matrix is equal to 1.

- **Define a multivariate probability distribution:**

  Young's modulus

  A one-dimensional normal distribution generalised to a higher dimension gives a multivariate normal distribution. The assumed normal distribution and the computed correlation matrix was used to define the multivariate normal distribution. In this thesis, this multivariate distribution was generated using inbuilt python functions.

  Tensile strength

  A one-dimensional lognormal distribution generalised to a higher dimension gives a multivariate log-normal distribution. The assumed uni-variate log-normal distribution and the computed correlation matrix was used to define the multivariate log-normal distribution.

  In this project, the multivariate lognormal distribution was generated in python. However, there were no inbuilt python functions to generate the multivariate log-normal distribution. So the mean and variance of the log-normal distribution were converted to the mean and variance of a corresponding normal distribution so that the inbuilt python function for generating the multivariate normal distribution could be used. This conversion was done using Eq. B.2 and Eq. B.3[82, 83]:

$$m_X = 2\ln(m_Y) - \frac{1}{2}\ln\left(\sigma_Y^2 + m_Y^2\right) \tag{B.2}$$

$$\sigma_X^2 = -2\ln(m_Y) + \ln\left(\sigma_Y^2 + m_Y^2\right) \tag{B.3}$$

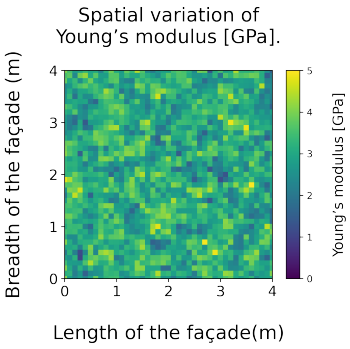
  Where,

  m_X, σ_X^2 = The mean and variance of the normal distribution.
  m_Y, σ_y^2 = The mean and variance of the log-normal distribution.

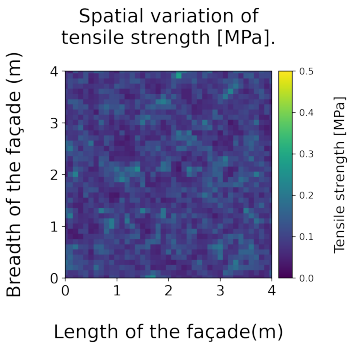
  The exponent of the samples of the multivariate normal distribution (generated using the inbuilt python function) was then used to generate the samples of the multivariate log-normal distribution.

- **Generate the random field:** A random sample chosen from these multivariate probability distributions is a particular realisation of the random field. These random variables were then assigned to the corresponding finite element as a material property.

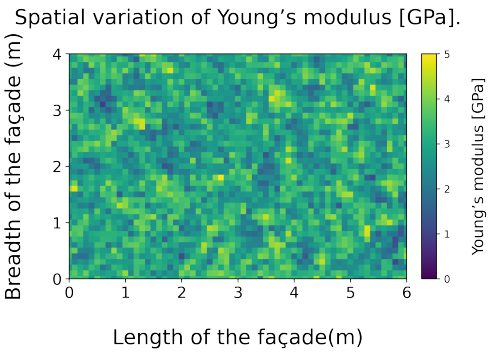
In Fig. B.2, the spatial variation of Young's modulus and tensile strength are shown for masonry walls of the different aspect ratios considered in this project for a mesh size of 100 mm.



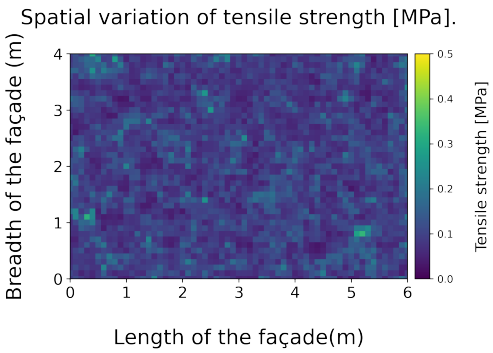
(a) Spatial variation of Young's modulus [GPa] of a masonry wall with aspect ratio 1:1.



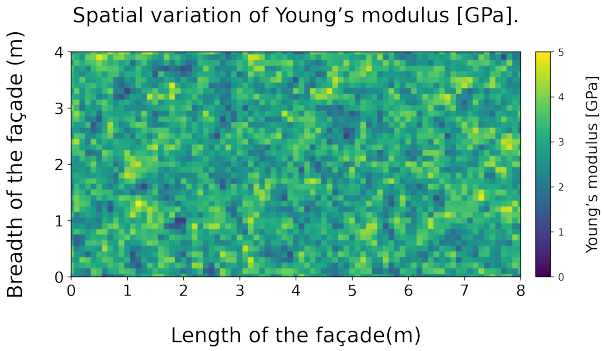
(b) Spatial variation of tensile strength [MPa] of a masonry wall with aspect ratio 1:1.



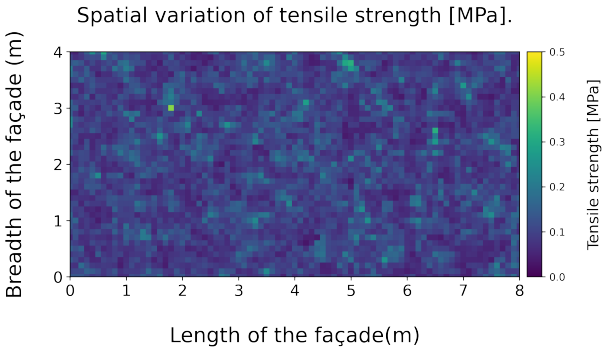
(c) Spatial variation of Young's modulus [GPa] of a masonry wall with aspect ratio 3:2.



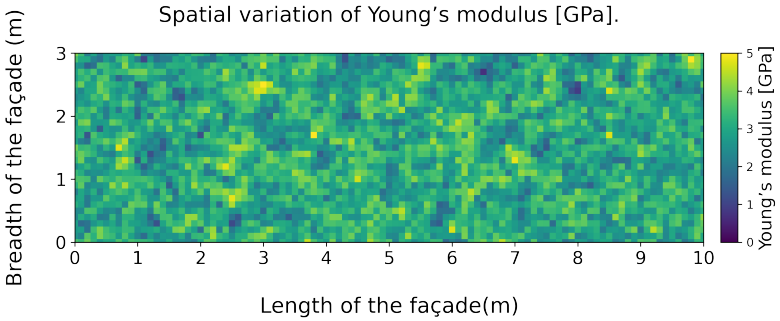
(d) Spatial variation of tensile strength [MPa] of a masonry wall with aspect ratio 3:2.



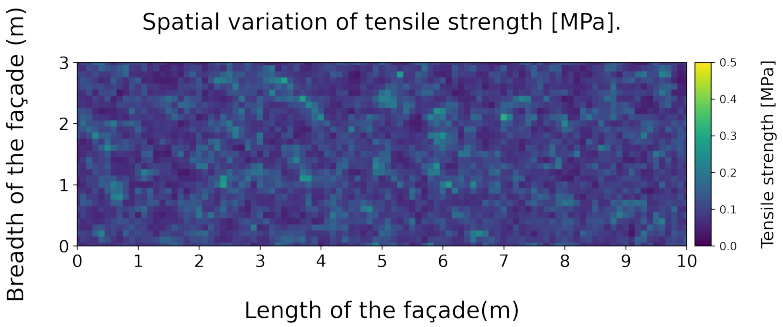
(e) Spatial variation of Young's modulus [GPa] of a masonry wall with aspect ratio 2:1.



(f) Spatial variation of tensile strength [MPa] of a masonry wall with aspect ratio 2:1.



(g) Spatial variation of Young's modulus [GPa] of a masonry wall with aspect ratio 10:3.



(h) Spatial variation of tensile strength [MPa] of a masonry wall with aspect ratio 10:3.

Figure B.2: The random field of the Young's modulus and the tensile strength for various aspect ratios of the masonry façades.

C DETAILED DESCRIPTION OF THE FINITE ELEMENT MODELS

The boundary and loading conditions used to simulate Pattern IDs 20, 21, 23, 24, 101, 102 and 103 are discussed in detail below.

C.1 PATTERN ID 20

Fig. C.1 shows Pattern ID 20, which consists of a horizontal crack along almost the entire length of the wall. These types of cracks tend to occur close to the roof or floor level [7]. They may be caused by overloading caused by a change in the load or the load path. Masonry façades subjected to dimensional changes induced by moisture or temperature may also cause these types of cracks [7].

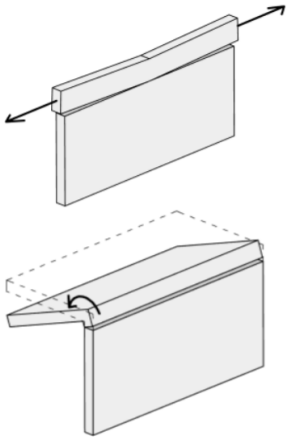
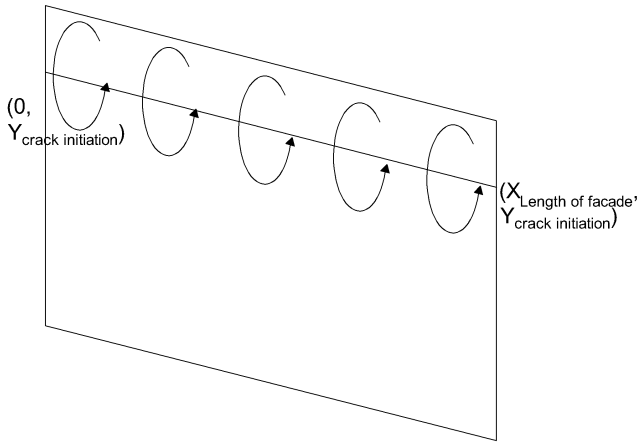


Figure C.1: Pattern ID 20 [7].

Boundary and loading conditions

In order to simulate this type of crack pattern, a horizontal line was imprinted along the length of the façade close to the roof. A moment, which was applied about this horizontally imprinted line, caused out of plane bending. The part of the masonry below this imprinted line was assumed to be fixed against all translational motion (X, Y, Z) along all four edges. The remaining upper part of the façade was only restrained by the shared edge it has with the lower portion of the façade, as shown in Fig. C.2a. The self-weight and the pre-compression load were applied to the structure before the application of the crack inducing load, as explained in Section 4.2.4. The crack inducing load applied to produce this type of crack pattern is shown in Fig. C.2b. The crack inducing load was applied using the force control as a moment along the horizontally imprinted line. In order to parametrise this crack type, the location of this horizontal line was randomly chosen in the top 70 - 85% of the façade height. The magnitude of the moment applied on the horizontally imprinted line was randomly chosen between 100 kNm/m and 200 kNm/m for

(a) Schematic representation of the boundary conditions of Pattern ID 20.



(b) Schematic representation of the loading conditions of Pattern ID 20.

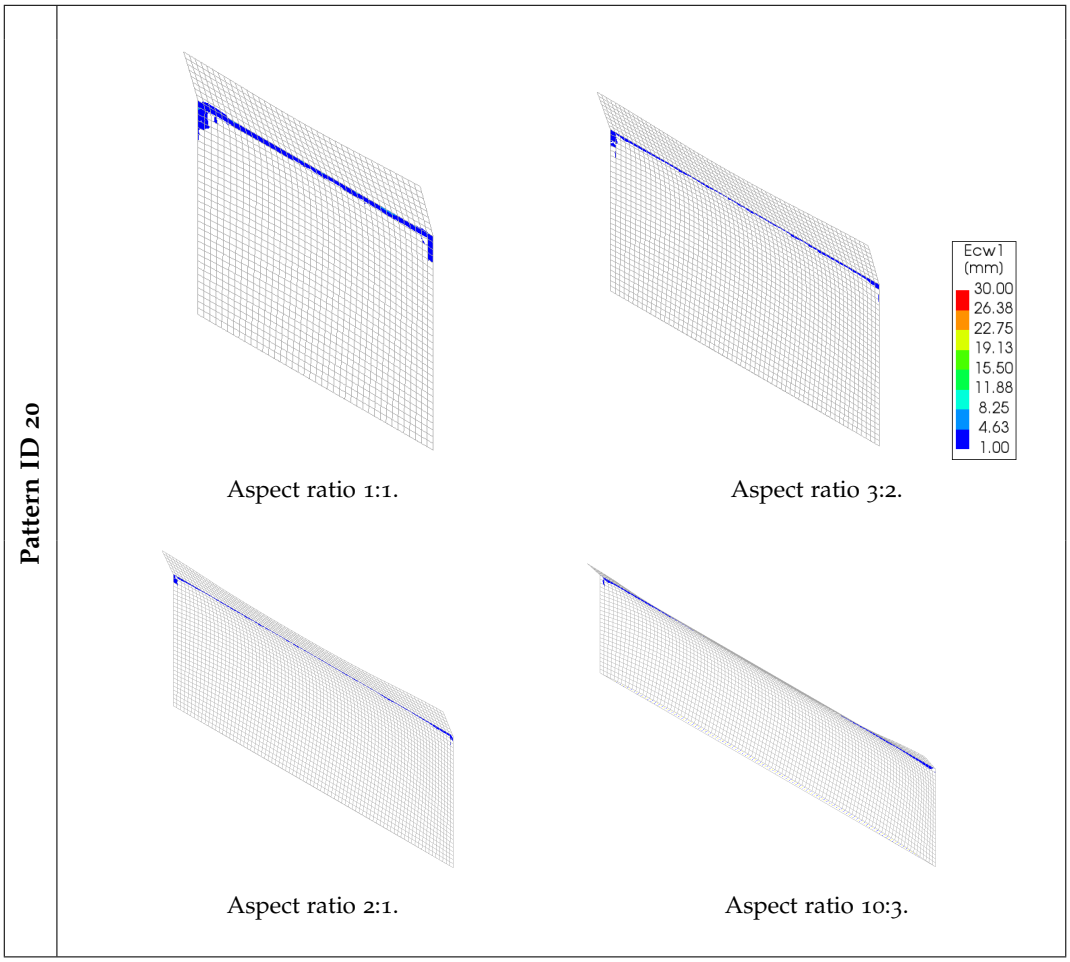
Figure C.2: Boundary and loading conditions of Pattern ID 20.

façades with thickness of 0.2 m and 0.3 m. For façades with a thickness of 0.4 m and 0.5 m, the magnitude of the moment was randomly chosen between 250 kNm/m and 350 kNm/m. This magnitude was enough to produce cracks with widths between 1 and 10 mm. Larger loads tend to cause numerical issues with the modelling and led to difficulties in achieving equilibrium in the finite element models. Smaller loads did not produce sufficiently large crack widths.

Crack Patterns

A nonlinear phased analysis of each of the models was completed using the procedure explained in Section 4.2.5. It was seen that even with 200 steps, this model has quite a number of points that cannot achieve convergence with 50 iterations per step. However, at the end of 200 steps, it was seen that the models are able to exhibit the expected crack patterns and deformations, as shown in Table C.1.

Table C.1: The deformed shape of the finite element model of Pattern ID 20 at the last step (deformation factor = 0.05)



C.2 PATTERN ID 21

In the crack pattern denoted by Pattern ID 20, a horizontal crack extends almost entirely along the length at location close to the middle of the height of the façade [7]. It is not too close the floor or the roof level, as shown in Fig. C.3. These types of cracks may be caused by vertical settlement or hindered dimensional changes due to chances in moisture content or temperature [7].

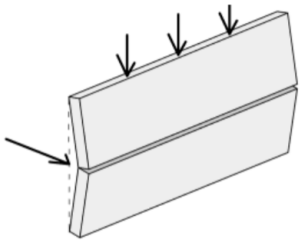


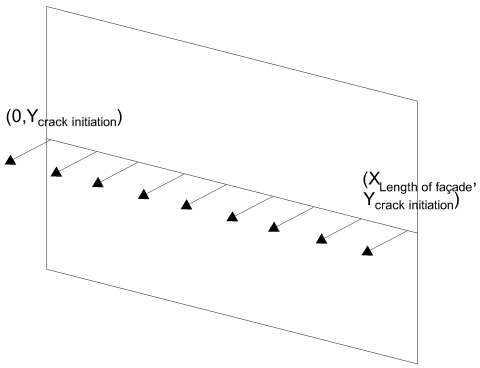
Figure C.3: Pattern ID 21 [7].

Boundary and loading conditions

A horizontal line was imprinted along the length of the façade near the mid-height of the façade to simulate this crack pattern. An out-of-plane load was applied along this line to produce an out-of-plane deformation that can cause such crack patterns. The edge of the façade connected to the rood and to the floor is assumed to be fixed against all translational motion (X, Y, Z), as shown in Fig. **??**.

After applying the self-weight and the pre-compression load in the first phase as explained in Section 4.2.4, the crack inducing load was applied in the second phase of the phased analysis. The crack inducing load applied to produce this type of crack pattern is shown in Fig. C.4. The crack-inducing load was applied as a distributed force along the horizontally imprinted line using the force control. The location of this horizontal line was randomly chosen to lie somewhere in the 30 - 70% of the façade height. The magnitude of the distributed force applied on the horizontally imprinted line varies with the thickness of the façade. For façades with a thickness of 0.2 m, the magnitude of the load was randomly chosen between 50 kN/m and 100 kN/m. For façades with thickness of 0.3 m, the load was randomly chosen between 100 kN/m and 150 kN/m, while façades with a thickness of 0.4 m have a magnitude of force between 150 kN/m and 250 kN/m. Finally, for façades with a thickness of 0.5 m, the magnitude of the force was randomly chosen between 250 kN/m and 350 kN/m. These loads produced crack widths in the range of 1 mm to 5 mm. If the magnitude of the load is larger, it tends to cause numerical issues with the modelling. Smaller loads did not result in large enough crack widths.

(a) Schematic representation of the boundary conditions of Pattern ID 21.



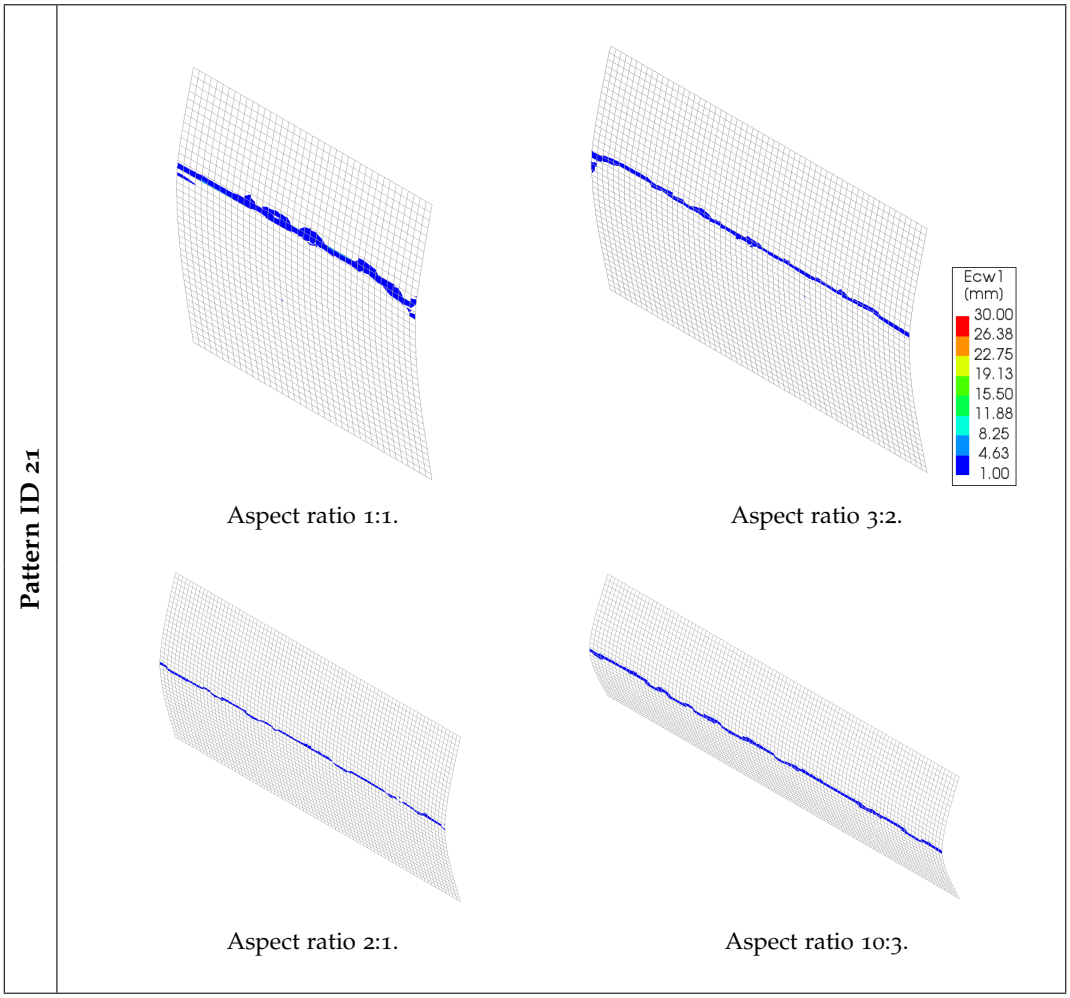
(b) Schematic representation of the crack inducing load given to Pattern ID 21.

Figure C.4: Boundary and loading conditions of Pattern ID 21.

Crack Patterns

The cracked façades produced from the non-linear phased analysis are given in Table C.2.

Table C.2: The deformed shape of the finite element model of Pattern ID 21 at the last step (Deformation factor = 0.05). The contour plot shows the crack width obtained in millimetres.



C.3 PATTERN ID 23

Pattern ID 23 is shown in Fig. C.5. It is typically denoted by a vertical crack that starts at the top of the façade and tapers towards the end of the crack [7]. These types of cracks maybe cause by vertical settlements or due to overloading. Hindered dimensional changes can also cause this type of cracks.

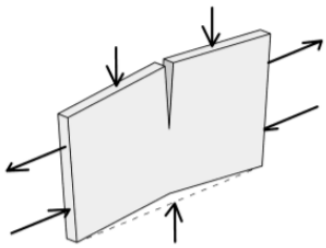


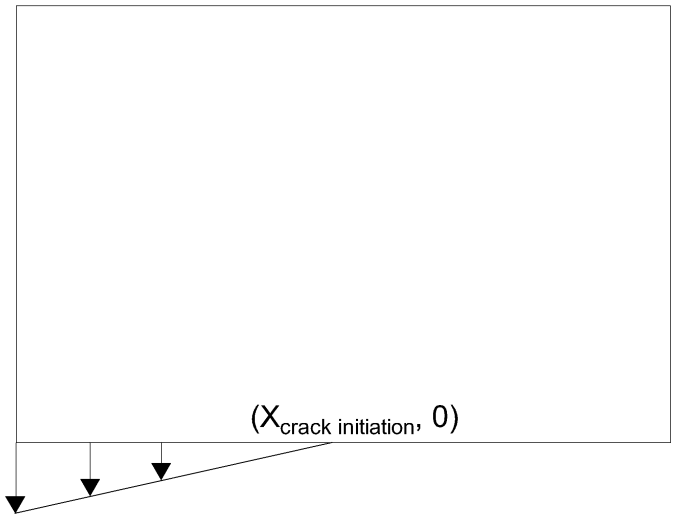
Figure C.5: Pattern ID 23 [7].

Boundary and loading conditions

To model the façades that exhibit this crack pattern, the bottom edge of the façade was assumed to be fixed in all translational directions (X, Y, Z). Additionally, to initiate the crack, a portion of the top edge was assumed to be restrained against motion in the vertical (Y) direction. These boundary conditions are shown in Fig. C.6a. The length of the top edge that was assumed to be restrained vertically was randomly chosen between 40% to 60% of the façade length.

The general loading conditions have been explained in Section 4.2.4. The crack inducing load used to produce this type of crack pattern is shown in Fig. C.6b. The cracks of this type were assumed to have been caused by soil subsidence. The load was applied using displacement control and was assumed to follow the geometry of the subsidence profile. The soil subsidence profile comprises of a linear function with its maximum at the vertex of the façade which eventually becomes zero along the façade length. The point at which the settlement profile becomes zero lies on the vertical axis at which the vertical restraint of the top edge begins. This can be understood from Fig. C.7. The magnitude of the settlement induced load was randomly chosen between 3 mm and 5 mm for all façades.

(a) Schematic representation of the boundary conditions of Pattern ID 23.



(b) Boundary and loading conditions of Pattern ID 23.

Figure C.6: The crack inducing load given to Pattern ID 23.

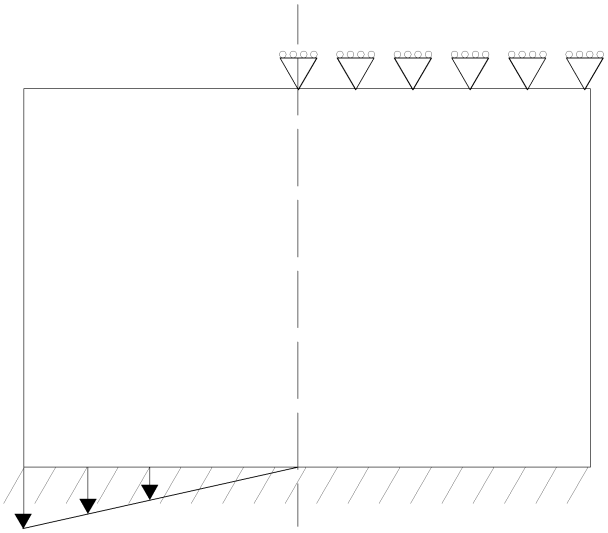
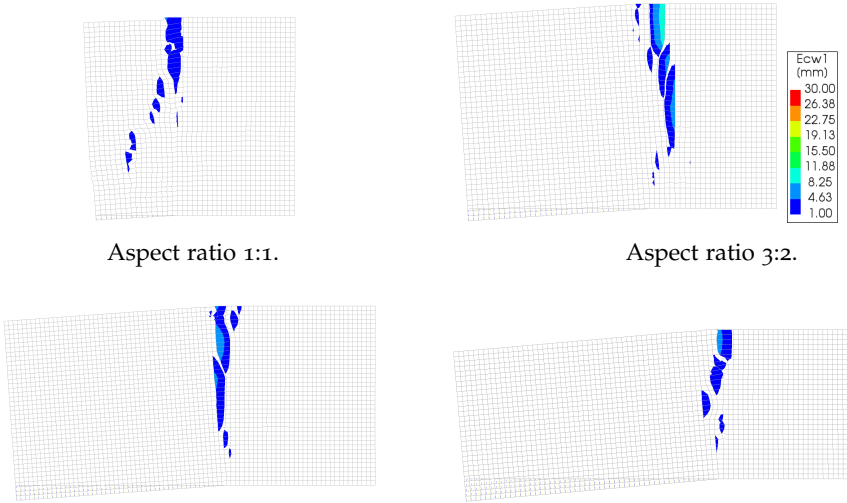


Figure C.7: Schematic representation of the boundary condition and the settlement profile of Pattern ID 23.

Crack Patterns

A non-linear phased analysis of each of the models was completed using the procedure explained in Section 4.2.5. The deformed façades showing the expected crack pattern are shown in Table C.3.

Table C.3: The deformed shape of the finite element model of Pattern ID 23 at the last step (deformation factor = 0.05)

Pattern ID 23	 Aspect ratio 1:1. Aspect ratio 3:2. Aspect ratio 2:1. Aspect ratio 10:3.

C.4 PATTERN ID 24

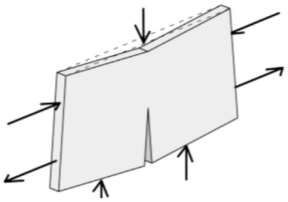
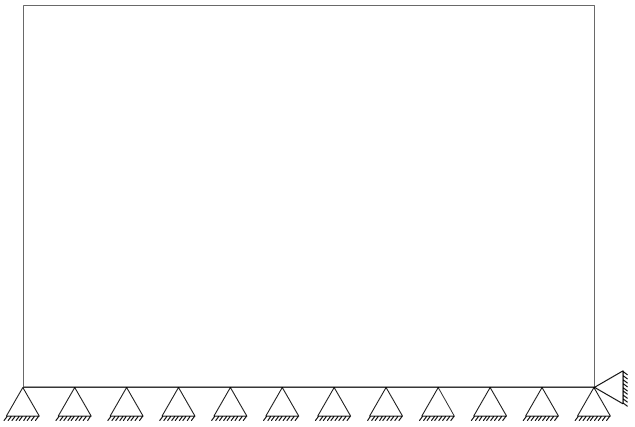


Figure C.8: Crack pattern of Pattern ID 24 [7].

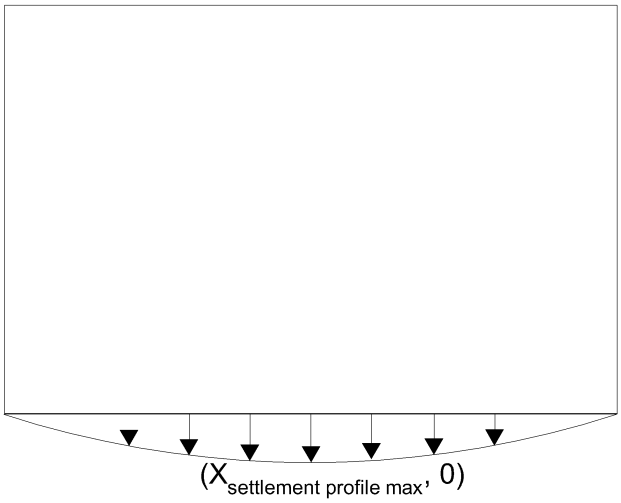
In Fig. C.8, Pattern ID 24 is shown. It typically shows a vertical crack beginning close to the floor of the façade which proceeds to taper along the crack length [7]. These cracks are widest at the bottom and they may be caused by vertical loading or hindered dimensional chances caused by changes induced by moisture or temperature [7].

Boundary and loading conditions

To simulate this type of crack pattern, the bottom edge of the façade was assumed to be fixed in translation for the vertical and horizontal directions (X, Y), as shown in Fig. C.9a. The general loading conditions have been explained in Section 4.2.4. The crack inducing load applied to produce this type of crack pattern is shown in Fig. C.9b. Like with the previous crack pattern type, the crack pattern of Pattern ID 24 was also assumed to have been caused by soil subsidence. A settlement induced load was applied using displacement control. The subsidence profile was assumed to be a Gaussian curve, with its maximum located close the middle of the length of the façade, as shown in Fig. C.9b. The Gaussian curve as a function of distance along the length of the façade is defined as per Eq. C.1.



(a) Schematic representation of the boundary conditions of Pattern ID 24.



(b) Schematic representation of the crack inducing load given to Pattern ID 24.

Figure C.9: Boundary and loading conditions of Pattern ID 24.

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right) \tag{C.1}$$

Where,

$\mu =$ Mean of the Gaussian curve. Here, it was used to specify the location of the maximum of the subsidence profile. Thus, the magnitude of μ was randomly chosen such that it occurs between 40% and 60% of the façade length.

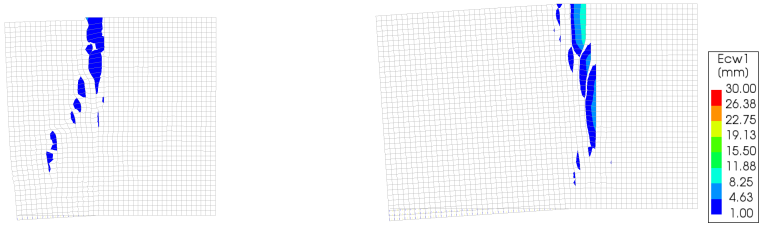
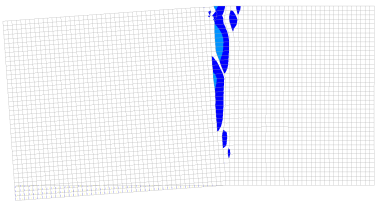
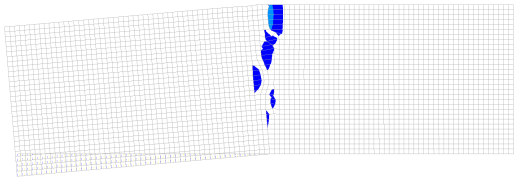
$\sigma =$ Standard deviation of the Gaussian curve. Here, it was used to specify the distribution of the subsidence profile about its maximum (μ). Thus, the magnitude of σ was randomly chosen such that it is between 20% and 30% of the façade length.

This type of load causes bending of the façades and result in tapered flexural cracks. The façades were subjected to a displacement of magnitude between 1 mm and 3 mm. Larger magnitudes of the loads sometimes led to small dispersed flexural cracks rather than one prominent crack.

Crack Patterns

The cracked façades produced from the non-linear phased analysis are given in Fig. C.4.

Table C.4: The deformed shape of the finite element model of Pattern ID 24 at the last step (deformation factor = 0.05)

Pattern ID 24	
	 Aspect ratio 1:1.
	Aspect ratio 3:2.
	 Aspect ratio 2:1.
	 Aspect ratio 10:3.

C.5 PATTERN ID 101

Figure C.10: Pattern ID 101 [6].

In the crack pattern denoted by Pattern ID 101 [6], a diagonal crack can be seen to initiate near an opening in the façade and moves towards the façade edge. These types of cracks may be caused by vertical settlement, one-end-heaves or hindered dimensional changes due to chances in moisture content or temperature. This type of crack pattern has been modelled on façades with openings.

Geometry

Pattern ID 101 was modelled on masonry façades with openings. This crack type can be simulated to initiate from any kind of openings, doors or windows. However, in this thesis, the modelling of Pattern ID 101 was limited to cracks initiating from doors. The position of the doors were kept at a distance of 1 m from the edge. Similar crack patterns opening from windows were modelled in Section C.6. For each aspect ratio of masonry façades considered in this study, there is a possibility to parametrise the location and number of openings (Door 1, Window 1 and Window 2). The dimensions of these openings are as explained in Section 4.2.1. After

considering the dimension of the façade, the dimensions of the openings, and the required position of the door to model this type of cracks, the following ten geometries were adopted to model Pattern ID 101.

For a façade of dimension 4 m by 4 m, only one geometry was possible after accommodating Door 1 and leaving sufficient distance between the edge. This is denoted by façade ID 1 and shown in Fig. C.11. In façades of dimensions 6 m by 4 m, 2 types of geometry were chosen, as shown in Fig.C.12. Door 1 was kept near the edge in both orientations. façade ID 2 also has Window 1, at a location randomly chosen such that it is at a distance of 1 m from the edge of the door and the wall as shown in Fig. C.12a.

façades of dimensions 8 m by 4 m show 3 types of geometry, as shown in Fig.C.13. Door 1 was kept near the edge in all three geometries. façade IDs 4 and 5 have additional openings: Window 1 and Window 2 respectively. These windows were placed at a location randomly chosen such that it is at a distance of 1 m from the edge of the door and the wall as shown in Fig. C.13a and Fig. C.13b. Masonry façades of dimensions 10 m by 3 m were generated to show 4 types of geometries, as shown in Fig.C.14. Door 1 was kept near the edge in all the geometries. façade IDs 7 and 8 had two additional openings, Window 1 and Window 2 as shown in Fig. C.14a and Fig. C.14b. The distance between the openings were kept constant as shown in Fig. 4.1. façade IDs 9 and 10 had only one extra opening: Window 1 and Window 2 respectively, as shown in Fig. C.14c and Fig. C.14d. These windows were placed at a location randomly chosen such that it was at a distance of 1 m from the edge of the door and the wall.

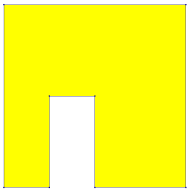
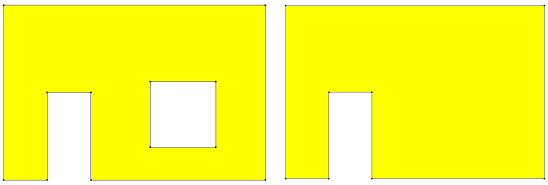
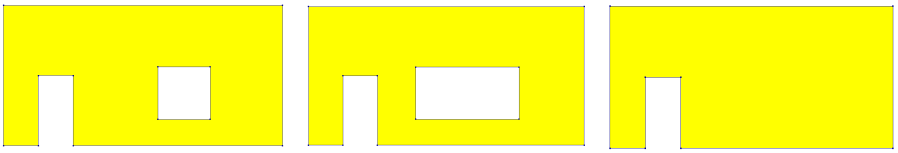


Figure C.11: Geometry of a 4 m x 4 m façade exhibiting façade ID 1.



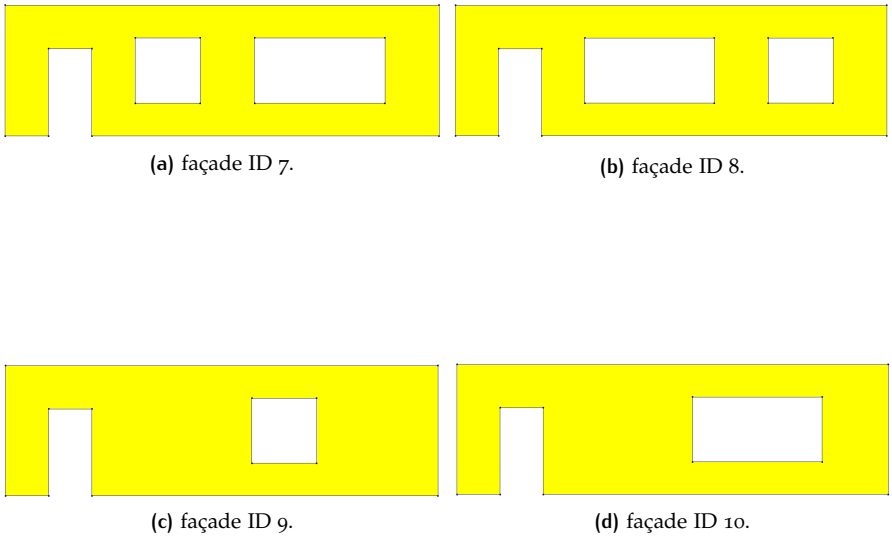
(a) façade ID 2. **(b)** façade ID 3.

Figure C.12: Geometry of a 6 m x 4 m façade exhibiting façade ID 2 and 3.



(a) façade ID 4. **(b)** façade ID 5. **(c)** façade ID 6.

Figure C.13: Geometry of a 6 m x 4 m façade exhibiting façade ID 4, 5 and 6.

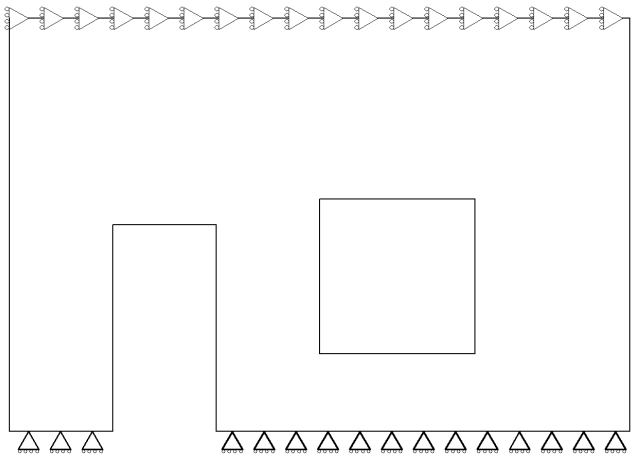


(a) façade ID 7. **(b)** façade ID 8. **(c)** façade ID 9. **(d)** façade ID 10.

Figure C.14: Geometry of a 10 m x 3 m façade exhibiting façade ID 7, 8, 9 and 10.

Boundary and loading conditions

To simulate Pattern ID 101, it was assumed that the bottom edge of the façade was fixed in translation in the vertical and horizontal directions (X, Y), as shown in Fig. C.15a. It was also assumed that the top edge was restrained against horizontal motion (X). The general loading conditions were explained in Section 4.2.4. For façades with openings, it was assumed that they were non-load bearing and only carried the self-weight before the application of the crack inducing load. The crack inducing load applied to produce this type of crack pattern is shown in Fig. C.15b. It was assumed that this type of crack patterns were caused by soil subsidence. A settlement induced load was simulated using displacement control. The subsidence profile was assumed to be a Gaussian curve, with its maximum located close to the edge of the façade closest to the door, as shown in Fig. C.15b, i.e. the left edge. The Gaussian curve as a function of distance along the length of the façade is defined as per Eq. C.1. The mean of this Gaussian curve was assumed to be at $x = 0\ m$ in order to obtain the maximum at the left edge of the façade, and the standard deviation was randomly chosen to lie in (1, 2) m.



(a) Schematic representation of the boundary conditions of Pattern ID 101.

(b) Schematic representation of the crack inducing load given to Pattern ID 101.

Figure C.15: Boundary and loading conditions of Pattern ID 101.

Crack Patterns

The cracked façades produced from the non-linear phased analysis are given in Table C.5.

Table C.5: The deformed shape of the finite element model of Pattern ID 101 at the last step (deformation factor = 0.05).

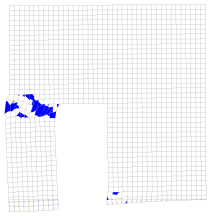
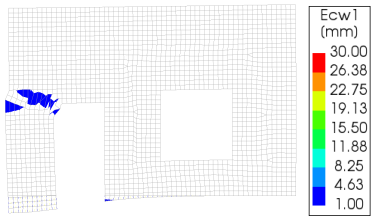
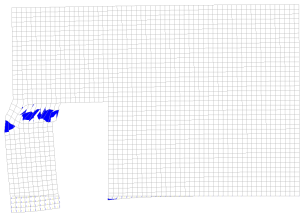
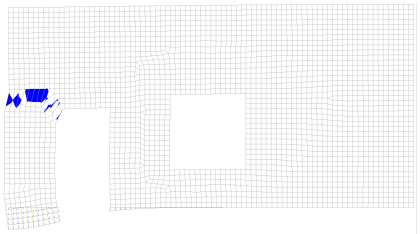
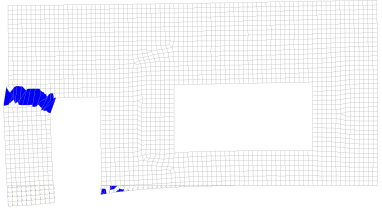
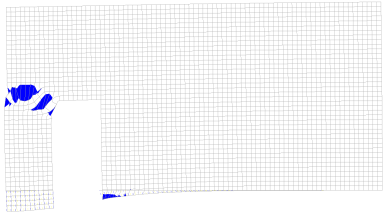
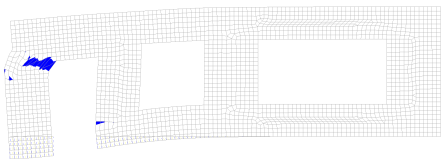
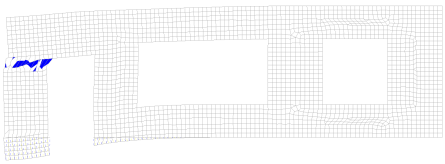
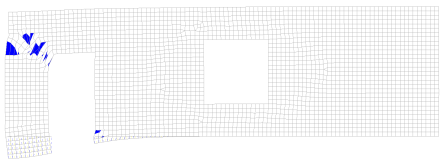
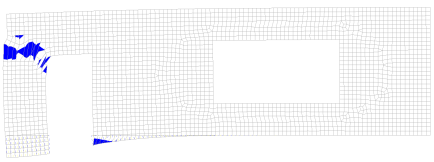
Pattern ID 101	
 Aspect ratio 1:1.	 Aspect ratio 3:2.
 Aspect ratio 3:2.	 Aspect ratio 2:1.
 Aspect ratio 2:1.	 Aspect ratio 2:1.

Table C.5: The deformed shape of the finite element model of Pattern ID 101 at the last step (deformation factor = 0.05).

Pattern ID 101		
	 Aspect ratio 10:3.	 Aspect ratio 10:3.
	 Aspect ratio 10:3.	 Aspect ratio 10:3.

C.6 PATTERN ID 102

Pattern ID 102 [6], was characterised by diagonal cracks initiating near an opening in the façade and moving towards the façade edges, similar to Pattern ID 101. These types of cracks may also be caused by vertical settlement, one-end-heaves or hindered dimensional changes due to chances in moisture content or temperature.

Figure C.16: Pattern ID 102 [6].

Geometry

This type of crack pattern was modelled on masonry façades with window openings. Even though this crack type can be simulated to initiate from any kind of openings, in this thesis, the modelling of Pattern ID 102 was limited to cracks initiating from windows. The position of the windows were kept at a distance of 1 m from the edge. This was very similar to the crack patterns modelled in Section C.5. There is a possibility to parametrise the location and number of openings (Door 1, Window 1 and Window 2) in each masonry façade. Section 4.2.1 explains the dimensions of these openings. The following 10 geometries were adopted to model Pattern ID 102. façade ID 11 was modelled on a masonry façade of dimension 4 m by 4 m with only Window 1. This was the only one geometry possible for a façade of this dimension after accommodating Window 1 and leaving sufficient distance between the edge, as shown in Fig. C.17. Two types of geometry were chosen for façades of dimension 6 m by 4 m, as shown in Fig.C.18. façade ID 12 has openings Door 1 and Window 1, and façade ID 13 has opening Window 2. Both windows were kept at a distance of 1 m from the edge. In façade ID 12, Door 1 was at a

location randomly chosen such that it was at a distance of 1 m from the edge of Window 1 and the edge as shown in Fig. C.18a.

façade IDs 14, 15 and 16 were the chosen geometries of masonry façades of dimensions 8 m by 4 m to model Pattern ID 102. This is shown in Fig.C.19. Window openings on the left edge was kept at distance of 1 m from the edge in all three geometries. façade IDs 14 and 15 have an additional opening: Door 1 placed at a location randomly chosen such that it was at a distance of 1 m from the edge of the door and the wall as shown in Fig. C.19a and Fig. C.19b. For the façades of dimensions 10 m by 3 m, façade IDs 17, 18, 19 and 20 were the chosen geometries, as shown in Fig.C.20. Window 1 and Window 2 were kept near the left edge in all the geometries. façade IDs 17 and 18 have two additional openings, as shown in Fig. C.20a and Fig. C.20b. The distance between the openings were kept constant as shown in Fig. 4.1. façade IDs 19 and 20 have an additional opening of Door 1, as shown in Fig. C.20c and Fig. C.20d. These doors were placed at a location randomly chosen such that it was at a distance of 1 m from the edge of the door and the wall.

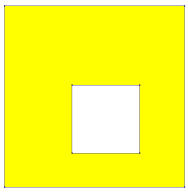
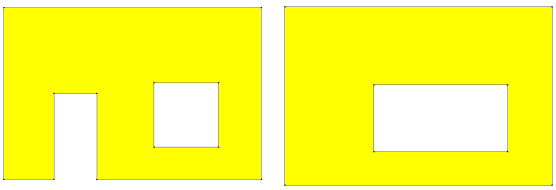
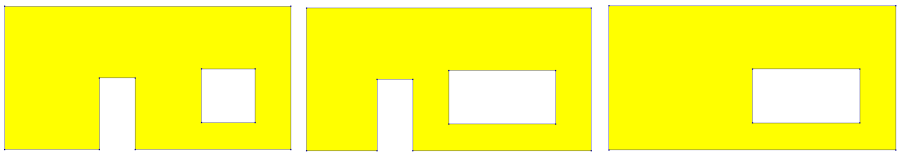


Figure C.17: Geometry of a 4 m x 4 m façade exhibiting façade ID 11.



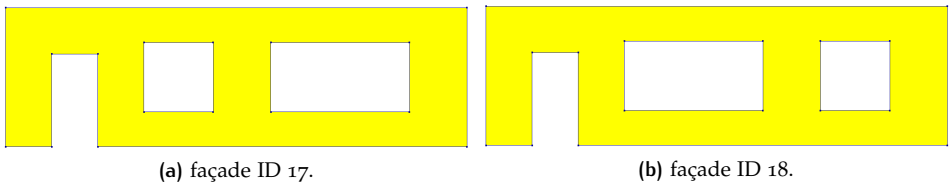
(a) façade ID 12. **(b)** façade ID 13.

Figure C.18: Geometry of a 6 m x 4 m façade exhibiting façade ID 2 and 3.

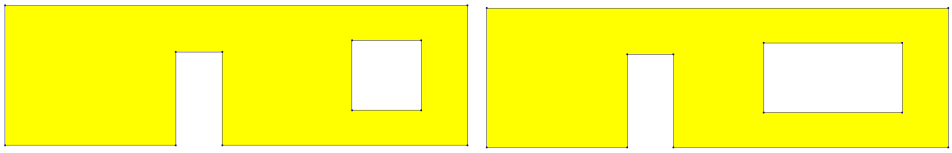


(a) façade ID 14. **(b)** façade ID 15. **(c)** façade ID 16.

Figure C.19: Geometry of a 6 m x 4 m façade exhibiting façade ID 14, 15 and 16.



(a) façade ID 17. **(b)** façade ID 18.

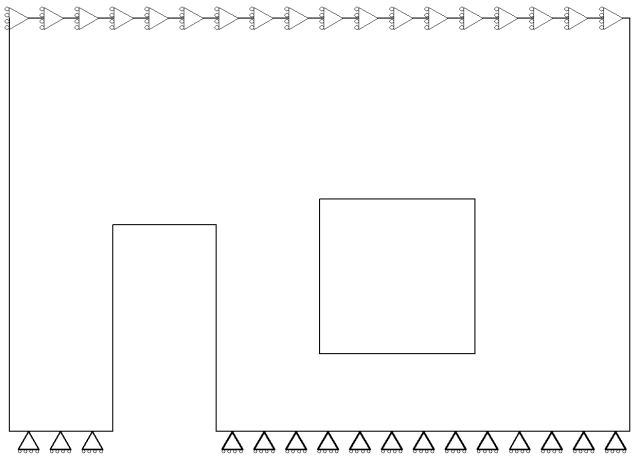


(c) façade ID 19. **(d)** façade ID 20.

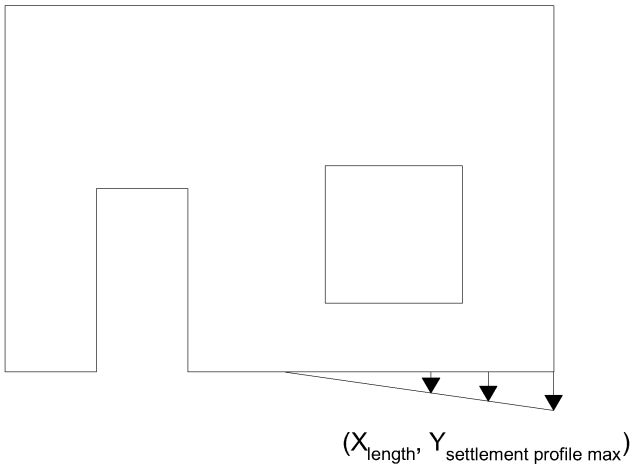
Figure C.20: Geometry of a 10 m x 3 m façade exhibiting façade ID 17, 18, 19 and 20.

Boundary and loading conditions

For this crack pattern type, it was assumed that the bottom edge of the façade was fixed in translation in the vertical and horizontal directions (X, Y), as shown in Fig. C.21a. It was also assumed that the top edge was restrained against horizontal motion (X). This is similar to Pattern ID 101. In Section 4.2.4, the general loading conditions are discussed. It was presumed that façades with openings are non-load bearing. The crack inducing load used to create this type of fracture pattern is depicted in C.21b. This type of crack pattern is thought to be created by soil subsidence simulated using displacement control. As shown in Fig. C.21b, the subsidence profile is considered to be a Gaussian curve, with the maximum positioned close to the edge of the façade closest to the window (the right edge). Eq. C.1 defines the Gaussian curve as a function of distance along the length of the façade. In order to produce the maximum at the right edge of the façade, the mean of this Gaussian curve was fixed at x = *length of the façade (m)*, and the standard deviation was arbitrarily determined to lie in (1, 2) m. The façades were subjected to a displacement of magnitude between 1 mm and 3 mm. Larger magnitudes of the loads sometimes led to small dispersed flexural cracks on the other side of the door. It was also seen that larger cracks tend to produce cracks at the edges of other window openings that was sufficiently close to the left edge.



(a) Schematic representation of the boundary conditions of Pattern ID 102.



(b) Schematic representation of the crack inducing load given to Pattern ID 102.

Figure C.21: Boundary and loading conditions of Pattern ID 102.

Crack Pattern

The cracked façades produced from the non-linear phased analysis are given in Table C.6.

Table C.6: The deformed shape of the finite element model of Pattern ID 102 at the last step (deformation factor = 0.05)

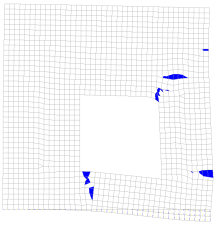
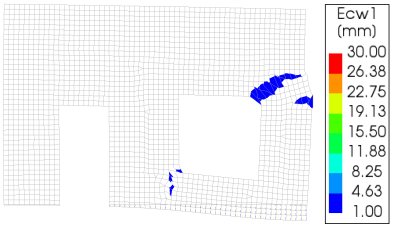
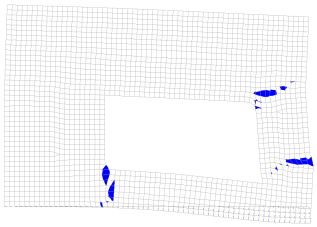
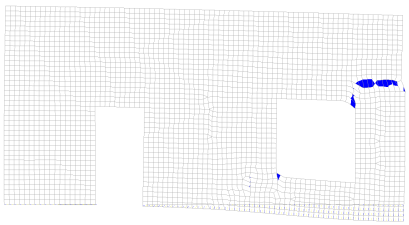
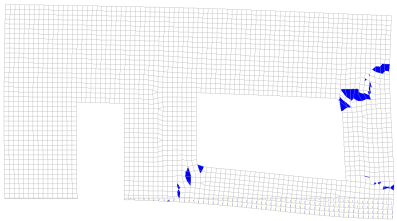
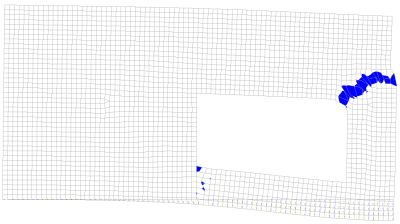
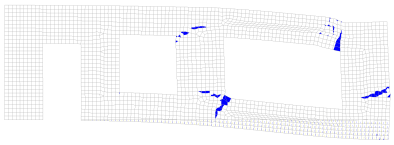
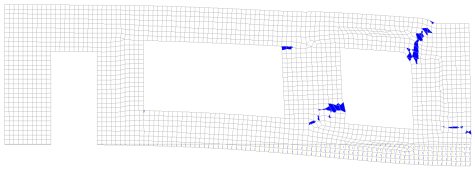
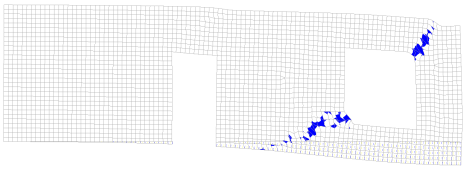
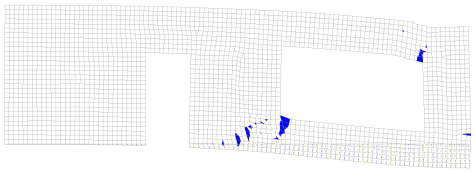
Pattern ID 102	 Aspect ratio 1:1.	 Aspect ratio 3:2.
	 Aspect ratio 3:2.	 Aspect ratio 2:1.
	 Aspect ratio 2:1.	 Aspect ratio 2:1.

Table C.6: The deformed shape of the finite element model of Pattern ID 102 at the last step (deformation factor = 0.05).

Pattern ID 101	 Aspect ratio 10:3.	 Aspect ratio 10:3.
	 Aspect ratio 10:3.	 Aspect ratio 10:3.

C.7 PATTERN ID 103

Pattern ID 103 [6], was described by a crack initiating close to the floor and moving towards an opening such as a window. These cracks were widest at the bottom and they may be caused by vertical loading or hindered dimensional chances caused by changes induced by moisture or temperature [7].

Figure C.22: Pattern ID 102 [6].

Geometry

Pattern ID 103 was modelled on masonry façades with window openings. As the crack was assumed to initiate from below window openings, all the selected geometries of this crack pattern have windows. The position of the all the openings were kept at a distance of 1 m from the edges and from each other. There is again a possibility to parametrise the location and number of openings (Door 1, Window 1 and Window 2), as explained in Section 4.2.1. The following 10 geometries were adopted to model Pattern ID 103. façade ID 21 was modelled on a masonry façade of dimension 4 m by 4 m with only Window 1. The location of Window 1 was randomly chosen such that there is a distance of 1 m from either edge, as shown in Fig. C.23. façade ID 22 and 23 was the two types of geometry chosen for façades of dimension 6 m by 4 m, as shown in Fig.C.24. façade ID 22 has openings Door 1 and Window 1, and façade ID 13 has opening Door 1 and Window 2. All the openings were at a location randomly chosen such that it was at a distance of 1 m from each other, as shown in Fig. C.24a.

façade IDs 24, 25 and 26 are the geometries of masonry façades of dimensions 8 m by 4 m to model Pattern ID 102. This is shown in Fig.C.25. All the openings were placed at a location randomly chosen such that it was at a distance of 1 m from each other and the edges, as shown in Fig. C.25a, Fig. C.25b and Fig. C.25c. Masonry façades of dimensions 10 m by 3 m were generated to show 4 types of geometries, as shown in Fig.C.26. façade IDs 27 and 28 have 3 openings: Door1, Window 1 and Window 2 as shown in Fig. C.26a and Fig. C.26b. The distance between the openings were kept constant as shown in Fig. 4.1. In façade IDs 29 and 30, all of the openings were randomly arranged at a distance of 1 m from each other and the edges, as shown in Fig. C.26c and Fig. C.26d.

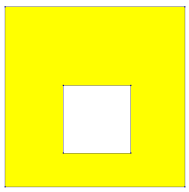
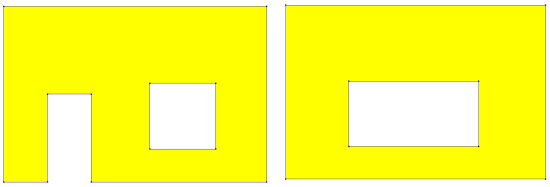
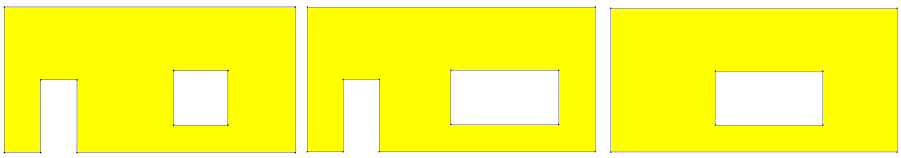


Figure C.23: Geometry of a 4 m x 4 m façade exhibiting façade ID 21.



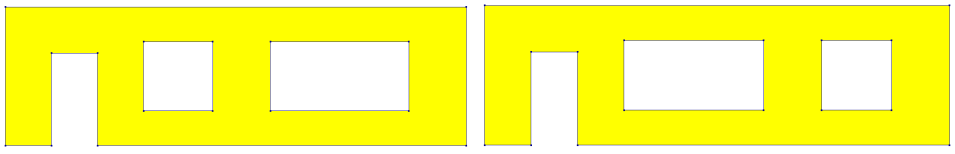
(a) façade ID 22. **(b)** façade ID 23.

Figure C.24: Geometry of a 6 m x 4 m façade exhibiting façade ID 22 and 23.

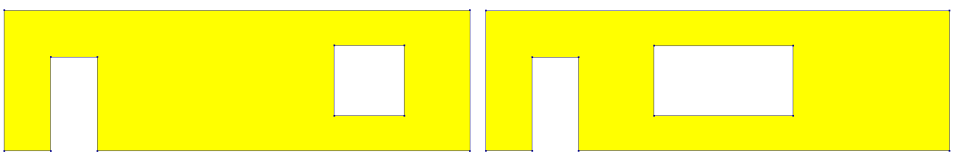


(a) façade ID 24. **(b)** façade ID 25. **(c)** façade ID 26.

Figure C.25: Geometry of a 6 m x 4 m façade exhibiting façade ID 24, 25 and 26.



(a) façade ID 27. **(b)** façade ID 28.

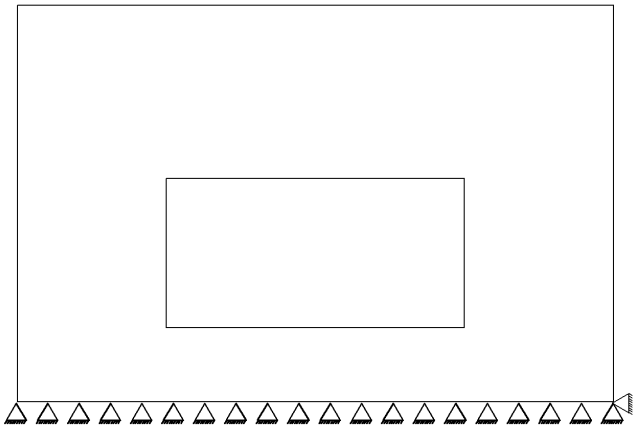


(c) façade ID 29. **(d)** façade ID 30.

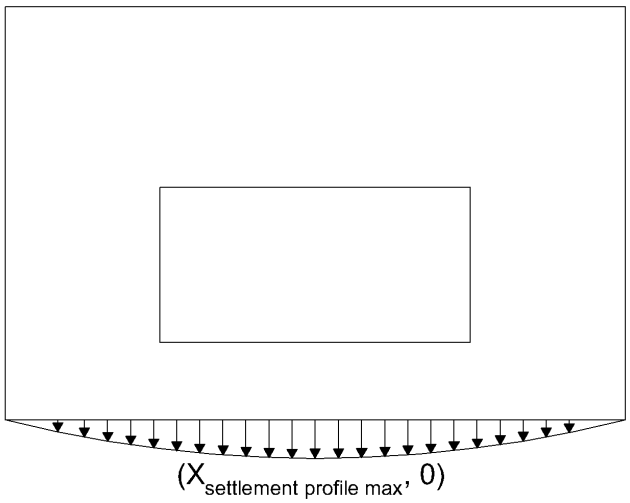
Figure C.26: Geometry of a 10 m x 3 m façade exhibiting façade ID 7, 8, 9 and 10.

Boundary and loading conditions

For this crack pattern type, it was assumed that the bottom edge of the façade was fixed in translation in the vertical and horizontal directions (X, Y), as shown in Fig. C.27a. The general loading conditions have been explained in Section 4.2.4. The crack inducing load applied to produce this type of crack pattern is shown in Fig. C.15a. This crack pattern was also assumed to have been caused by soil subsidence applied using displacement control. The subsidence profile was assumed to be a Gaussian curve, with its maximum located close to the middle of the length of the façade, as shown in Fig. C.8. The Gaussian curve as a function of distance along the length of the façade is defined as per Eq. C.1. The mean of this Gaussian curve was randomly chosen such that it occurred between 40% and 60% of the façade length. The magnitude of standard deviation was randomly chosen such that it was between 20% and 30% of the façade length. The façades bend under this type of load, resulting in tapering flexural cracks. The façades were subjected to displacements ranging from 1 to 3 mm. Larger loads resulted in small, distributed flexural cracks rather than a single conspicuous crack.



(a) Schematic representation of the boundary conditions of Pattern ID 103.



(b) Schematic representation of the crack inducing load given to Pattern ID 103.

Figure C.27: Boundary and loading conditions of Pattern ID 103.

Crack Pattern

The cracked façades produced from the non-linear phased analysis are given in Table C.7.

Table C.7: The deformed shape of the finite element model of Pattern ID 103 at the last step (deformation factor = 0.05).

Pattern ID 103

Aspect ratio 1:1. Aspect ratio 3:2.

Aspect ratio 3:2. Aspect ratio 2:1.

Aspect ratio 2:1. Aspect ratio 2:1.

Aspect ratio 10:3. Aspect ratio 10:3.

Aspect ratio 10:3. Aspect ratio 10:3.

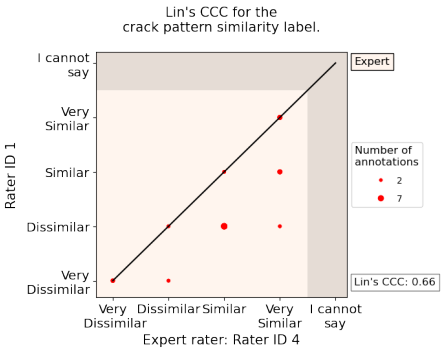
C.8 PATTERN IDS EXCLUDED FROM THIS THESIS

Pattern ID 30 is characterised by a diagonal crack that initiates close to the roof, as shown in Fig. 4.17a. The crack direction was diagonal and the width tapers towards the end. It is believed that these kind of cracks are caused by vertical settlements [7]. Similarly, Patter ID 31 appears in a column or at a corner, as shown in Fig. 4.17a. It is identified by a diagonal crack that is widest at the top and narrowest at the bottom. The crack runs diagonally, tapering along the length. It is thought to be caused by vertical settlement or hindered dimensional produced by changes in temperature or moisture [7]. Fig. 4.17a shows Pattern ID 32. It is recognised by a diagonal crack that imitates close to the floor and has diagonal crack direction. It is widest at the bottom and narrowest at the top. It is thought to be caused by vertical settlement or hindered dimensional changes [7]. Pattern ID 201 is characterised by cracks initiating near the roof on masonry façades with openings. These cracks are widest and the top and narrowest and the bottom and are assumed to be caused by hindered dimensional changes produced by changes in moisture or temperature [6].

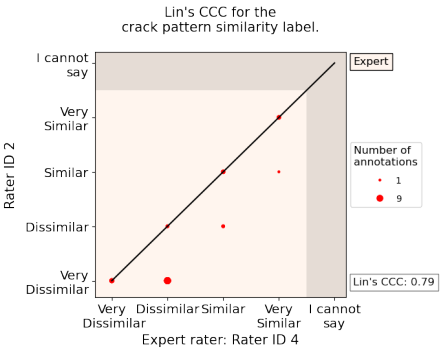
D AGREEMENT WITH THE STANDARD RATER

D.1 AGREEMENT WITH THE STANDARD RATER FOR THE DATA GENERATED FROM THE STATISTICS-BASED APPROACH

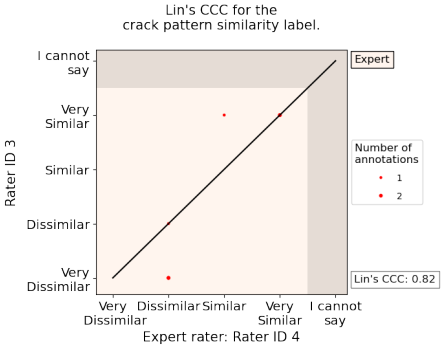
Crack Pattern Similarity Label



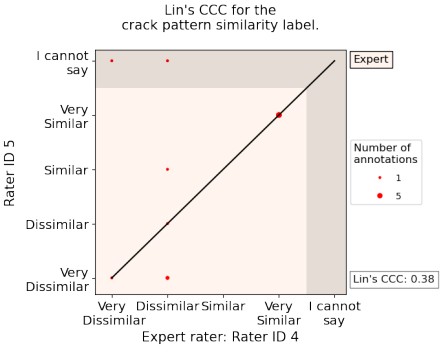
(a) Lin's CCC of Rater ID 1 with the standard rater.



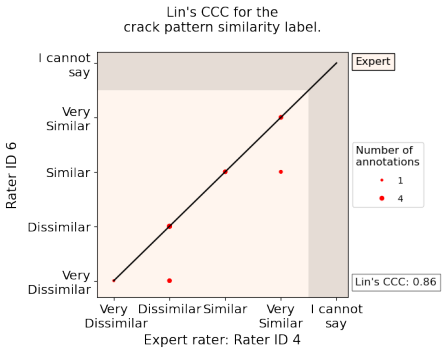
(b) Lin's CCC of Rater ID 2 with the standard rater.



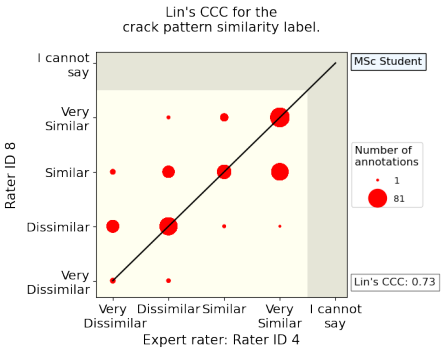
(c) Lin's CCC of Rater ID 3 with the standard rater.



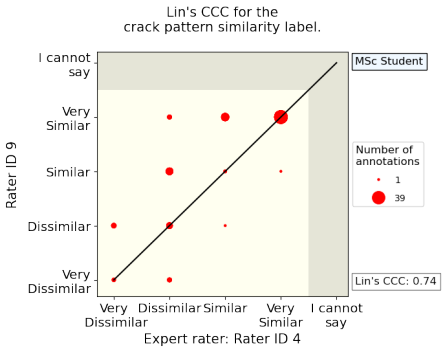
(d) Lin's CCC of Rater ID 5 with the standard rater.



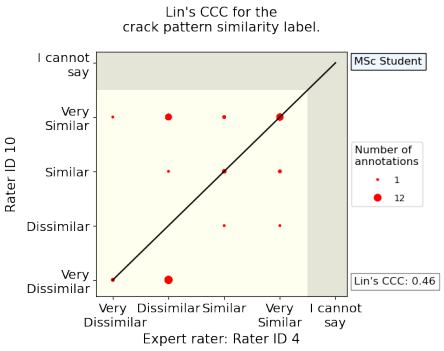
(e) Lin's CCC of Rater ID 6 with the standard rater.



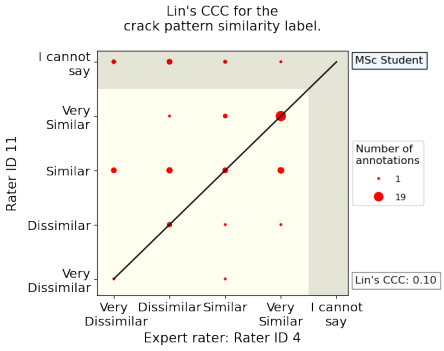
(f) Lin's CCC of Rater ID 8 with the standard rater.



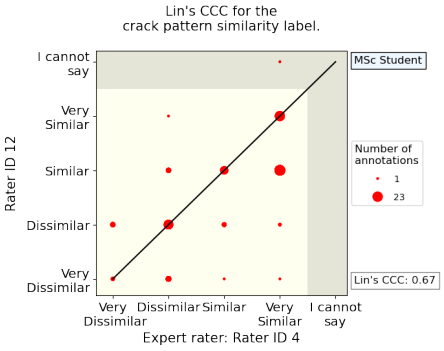
(g) Lin's CCC of Rater ID 9 with the standard rater.



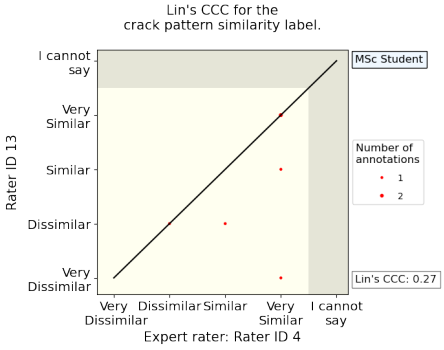
(h) Lin's CCC of Rater ID 10 with the standard rater.



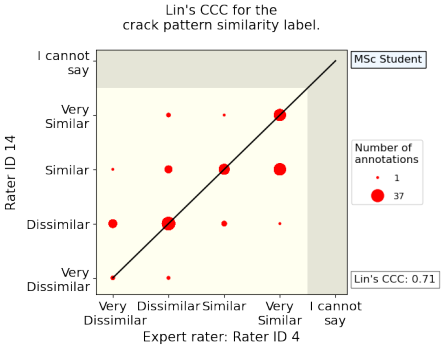
(i) Lin's CCC of Rater ID 11 with the standard rater.



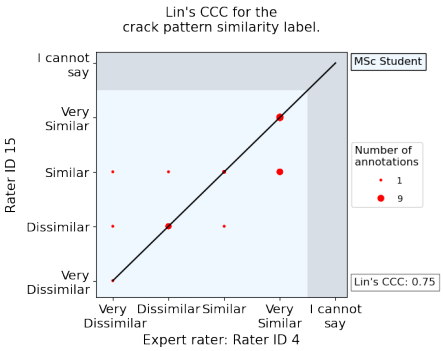
(j) Lin's CCC of Rater ID 12 with the standard rater.



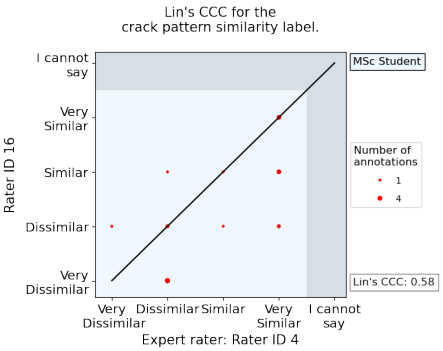
(k) Lin's CCC of Rater ID 13 with the standard rater.



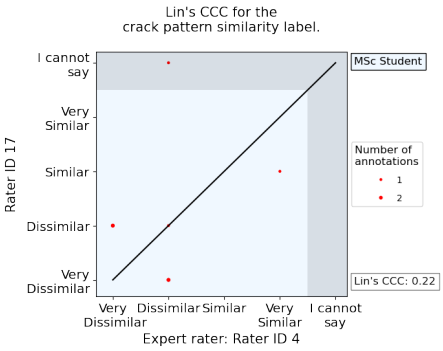
(l) Lin's CCC of Rater ID 14 with the standard rater.



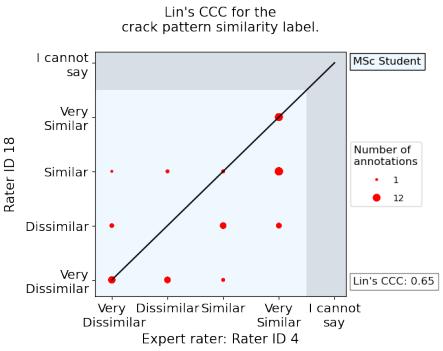
(m) Lin's CCC of Rater ID 15 with the standard rater.



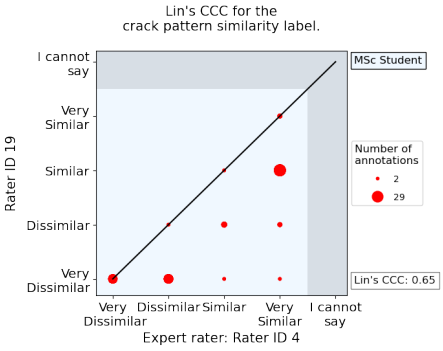
(n) Lin's CCC of Rater ID 16 with the standard rater.



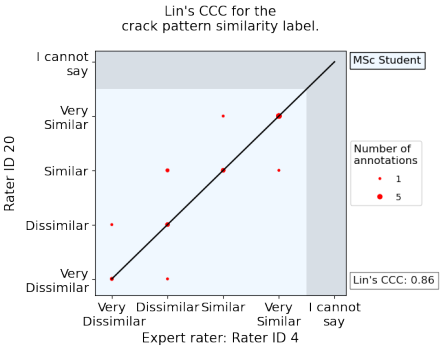
(o) Lin's CCC of Rater ID 17 with the standard rater.



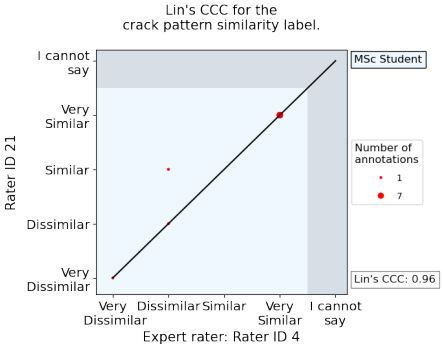
(p) Lin's CCC of Rater ID 18 with the standard rater.



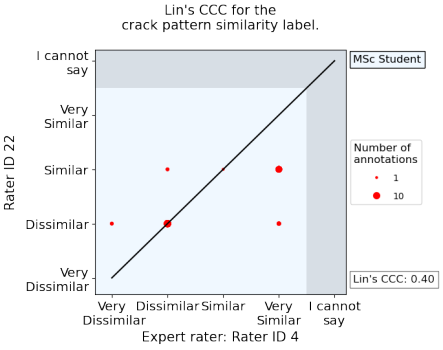
(q) Lin's CCC of Rater ID 19 with the standard rater.



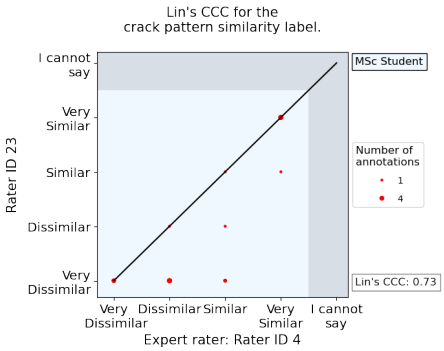
(r) Lin's CCC of Rater ID 20 with the standard rater.



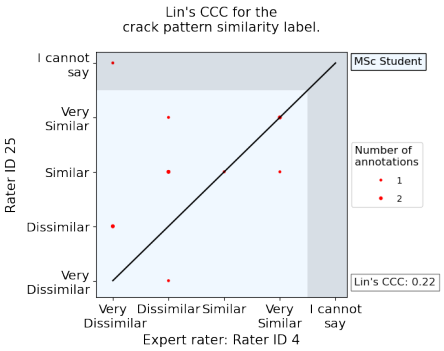
(s) Lin's CCC of Rater ID 21 with the standard rater.



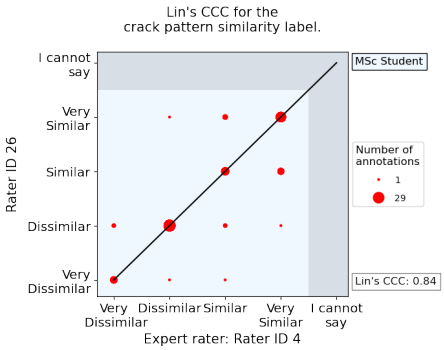
(t) Lin's CCC of Rater ID 22 with the standard rater.



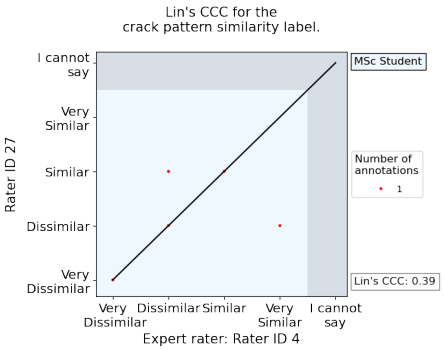
(u) Lin's CCC of Rater ID 23 with the standard rater.



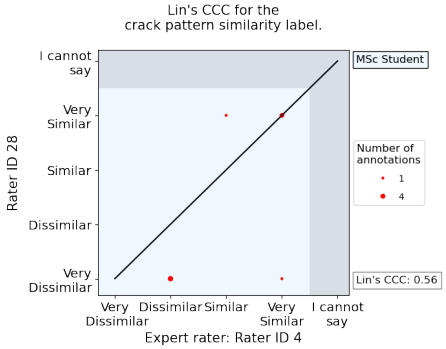
(v) Lin's CCC of Rater ID 25 with the standard rater.



(w) Lin's CCC of Rater ID 26 with the standard rater.



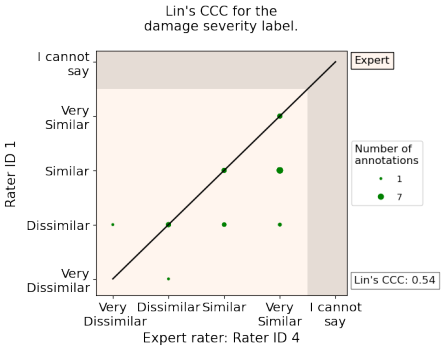
(x) Lin's CCC of Rater ID 27 with the standard rater.



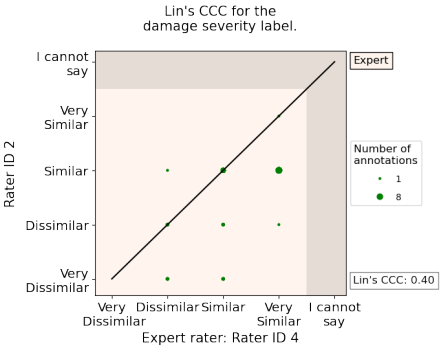
(y) Lin's CCC of Rater ID 28 with the standard rater.

Figure D.1: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the crack pattern similarity label for the data generated from the statistics-based approach.

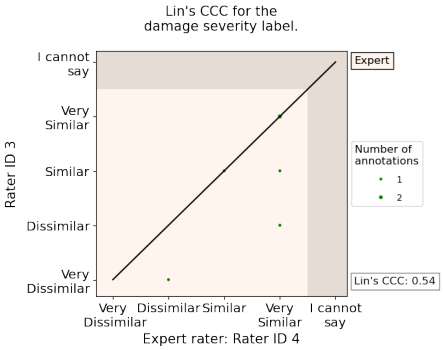
Damage Severity Label



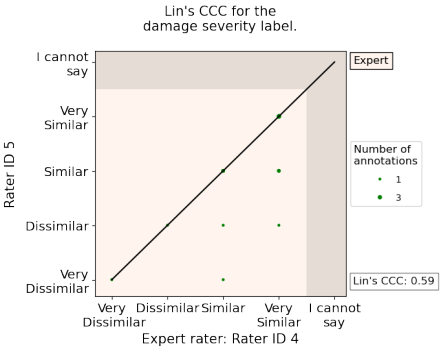
(a) Lin's CCC of Rater ID 1 with the standard rater.



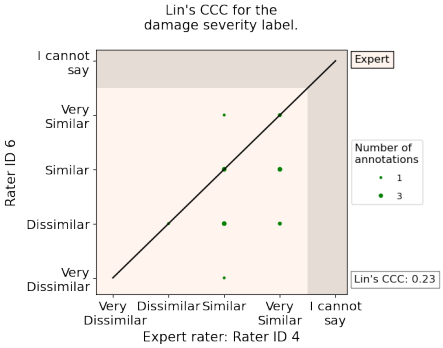
(b) Lin's CCC of Rater ID 2 with the standard rater.



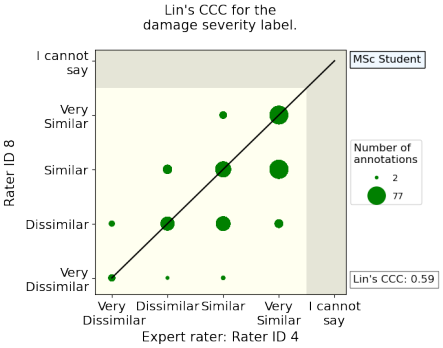
(c) Lin's CCC of Rater ID 3 with the standard rater.



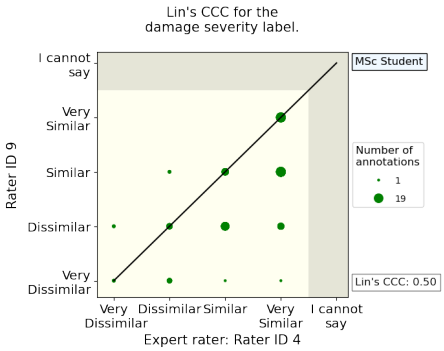
(d) Lin's CCC of Rater ID 5 with the standard rater.



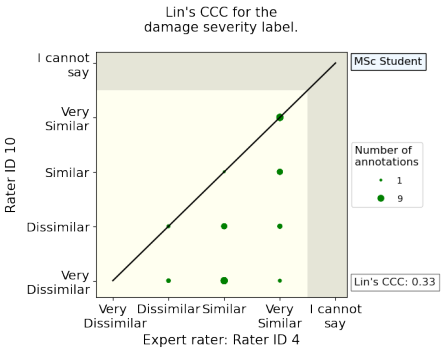
(e) Lin's CCC of Rater ID 6 with the standard rater.



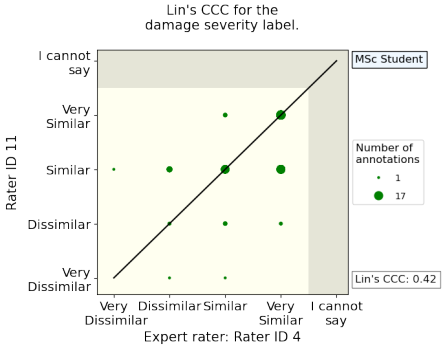
(f) Lin's CCC of Rater ID 8 with the standard rater.



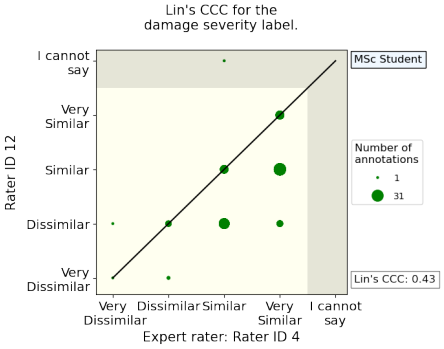
(g) Lin's CCC of Rater ID 9 with the standard rater.



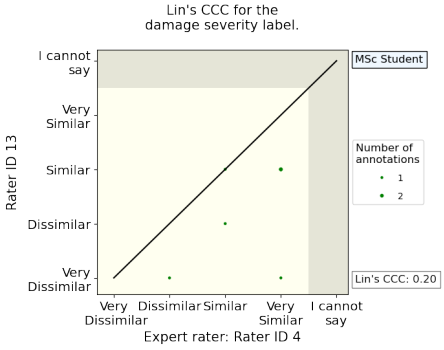
(h) Lin's CCC of Rater ID 10 with the standard rater.



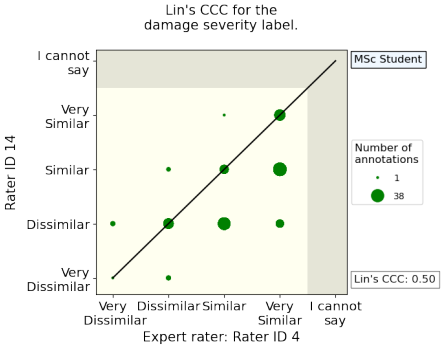
(i) Lin's CCC of Rater ID 11 with the standard rater.



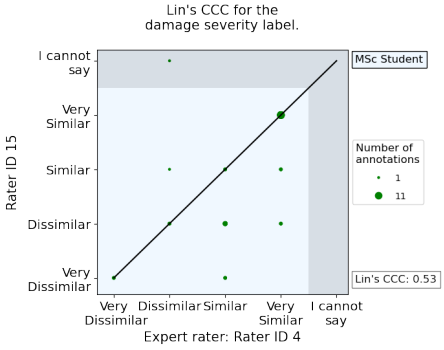
(j) Lin's CCC of Rater ID 12 with the standard rater.



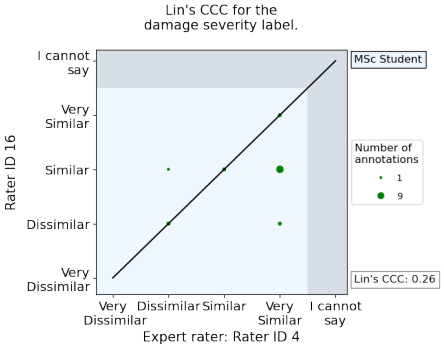
(k) Lin's CCC of Rater ID 13 with the standard rater.



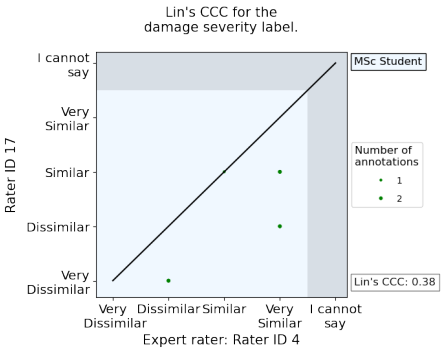
(l) Lin's CCC of Rater ID 14 with the standard rater.



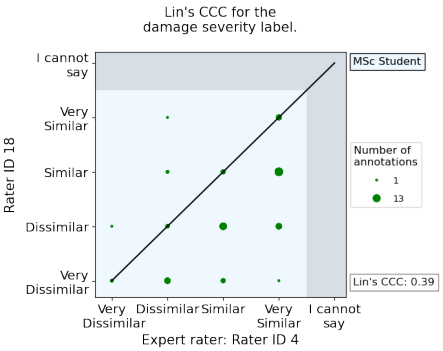
(m) Lin's CCC of Rater ID 15 with the standard rater.



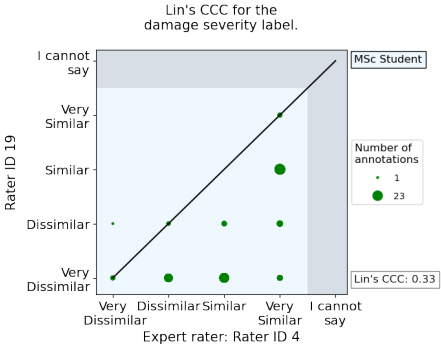
(n) Lin's CCC of Rater ID 16 with the standard rater.



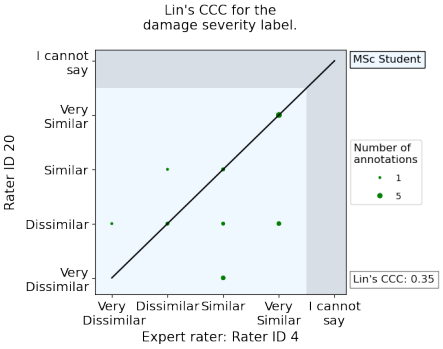
(o) Lin's CCC of Rater ID 17 with the standard rater.



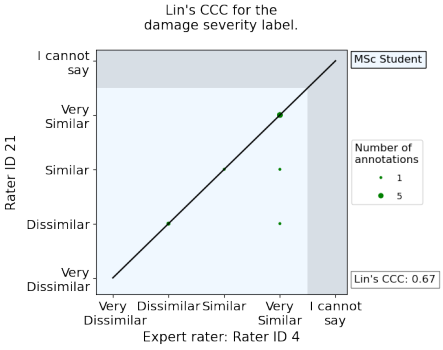
(p) Lin's CCC of Rater ID 18 with the standard rater.



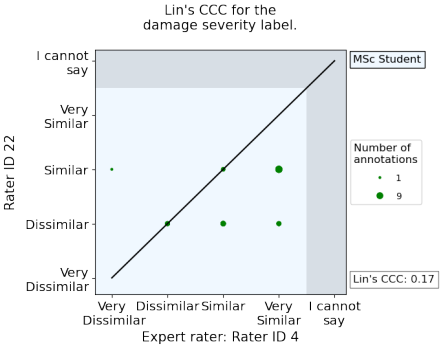
(q) Lin's CCC of Rater ID 19 with the standard rater.



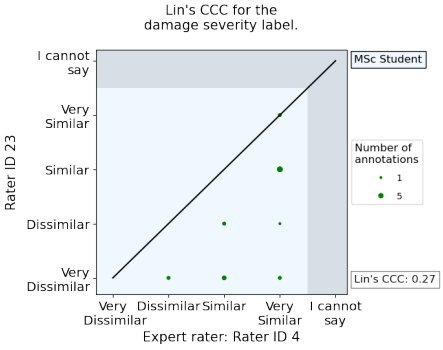
(r) Lin's CCC of Rater ID 20 with the standard rater.



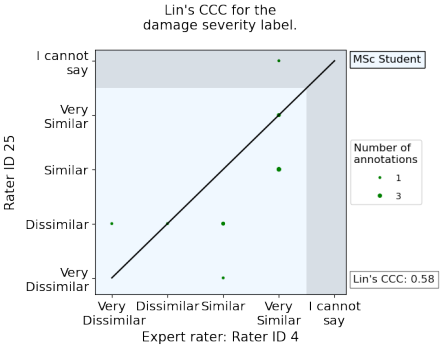
(s) Lin's CCC of Rater ID 21 with the standard rater.



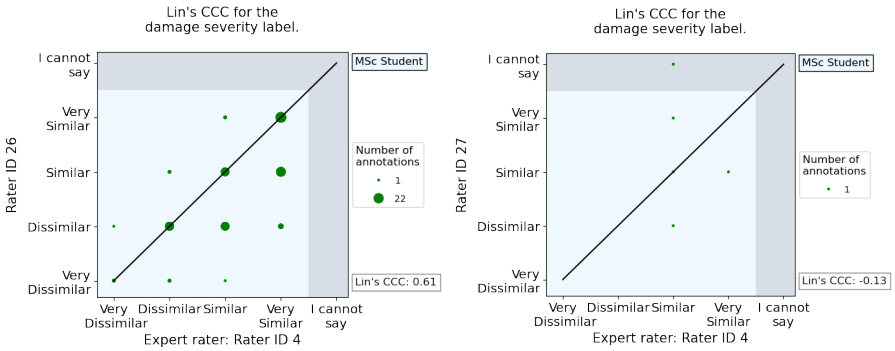
(t) Lin's CCC of Rater ID 22 with the standard rater.



(u) Lin's CCC of Rater ID 23 with the standard rater.

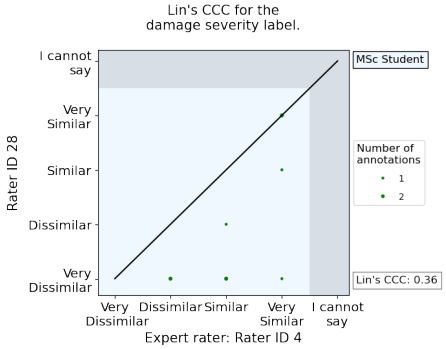


(v) Lin's CCC of Rater ID 25 with the standard rater.



(w) Lin's CCC of Rater ID 26 with the standard rater.

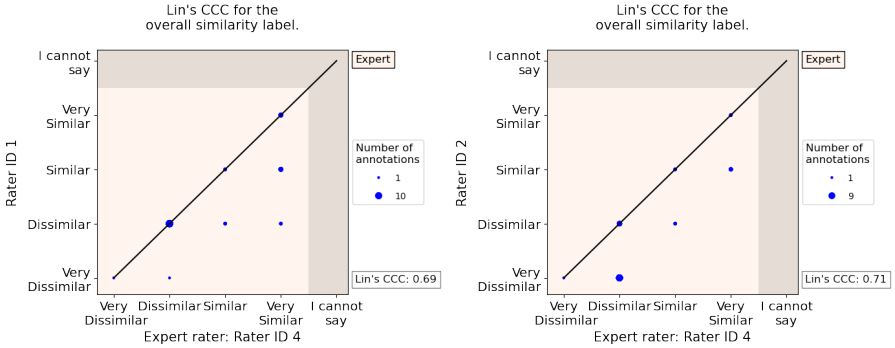
(x) Lin's CCC of Rater ID 27 with the standard rater.



(y) Lin's CCC of Rater ID 28 with the standard rater.

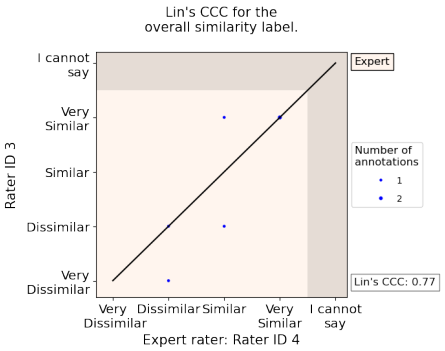
Figure D.2: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the damage severity label for data generated from the statistics-based approach.

Overall Similarity Label

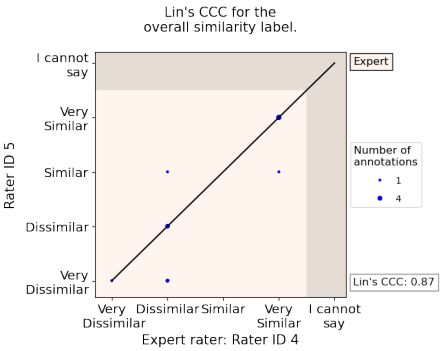


(a) Lin's CCC of Rater ID 1 with the standard rater.

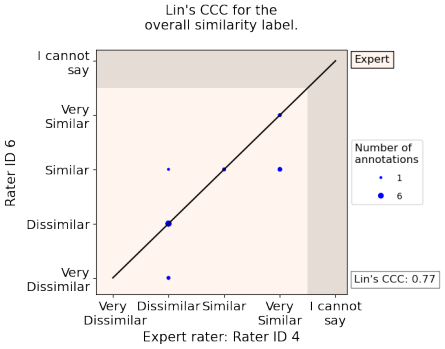
(b) Lin's CCC of Rater ID 2 with the standard rater.



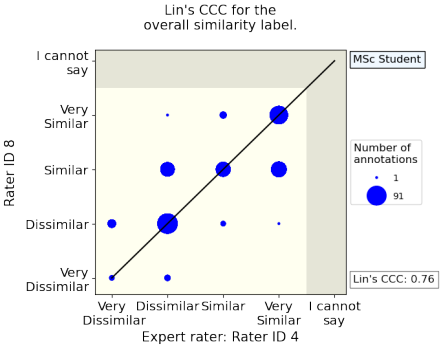
(c) Lin's CCC of Rater ID 3 with the standard rater.



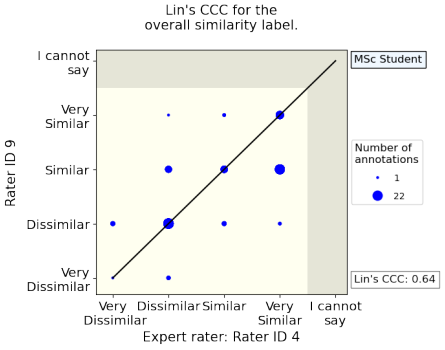
(d) Lin's CCC of Rater ID 5 with the standard rater.



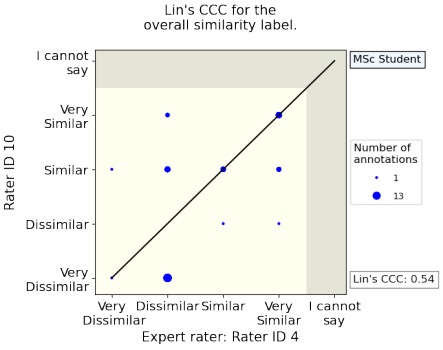
(e) Lin's CCC of Rater ID 6 with the standard rater.



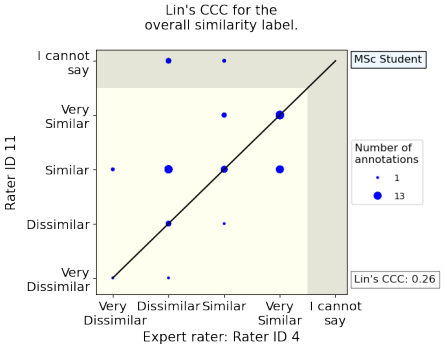
(f) Lin's CCC of Rater ID 8 with the standard rater.



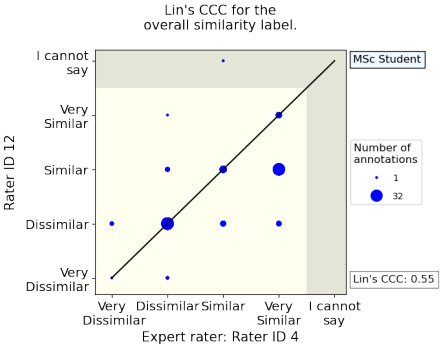
(g) Lin's CCC of Rater ID 9 with the standard rater.



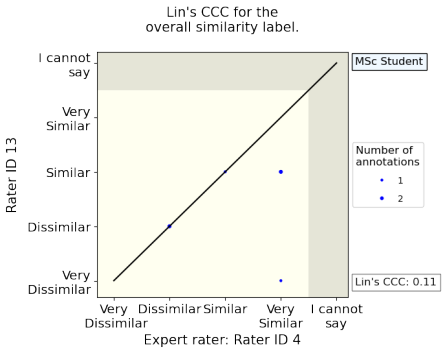
(h) Lin's CCC of Rater ID 10 with the standard rater.



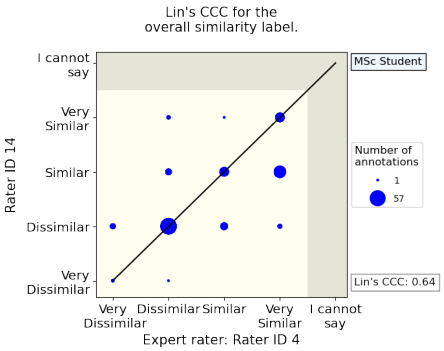
(i) Lin's CCC of Rater ID 11 with the standard rater.



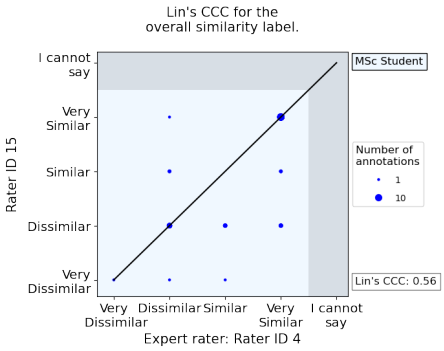
(j) Lin's CCC of Rater ID 12 with the standard rater.



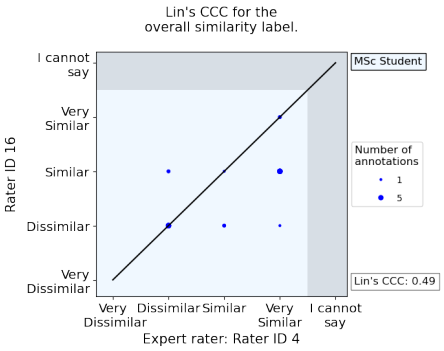
(k) Lin's CCC of Rater ID 13 with the standard rater.



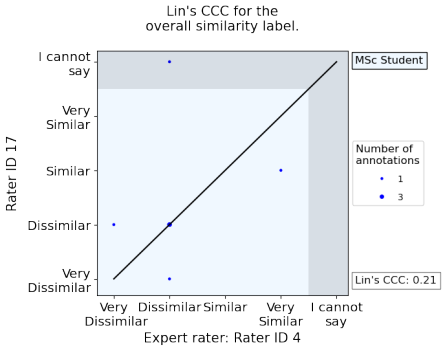
(l) Lin's CCC of Rater ID 14 with the standard rater.



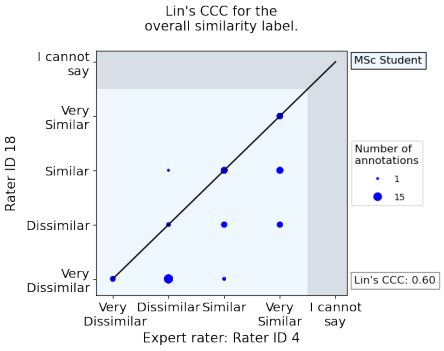
(m) Lin's CCC of Rater ID 15 with the standard rater.



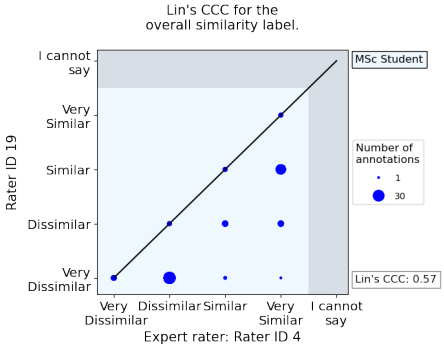
(n) Lin's CCC of Rater ID 16 with the standard rater.



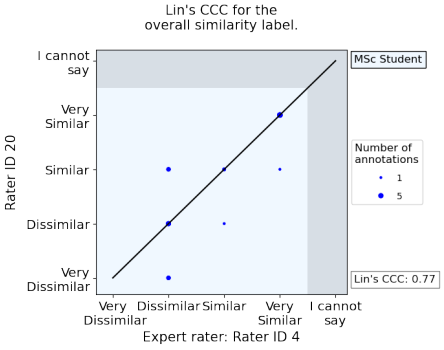
(o) Lin's CCC of Rater ID 17 with the standard rater.



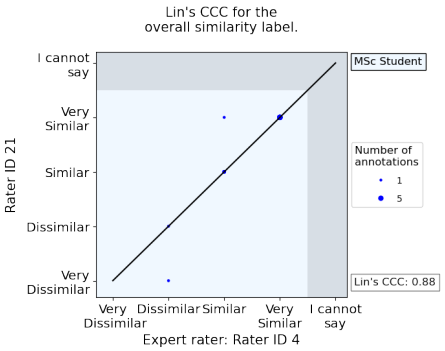
(p) Lin's CCC of Rater ID 18 with the standard rater.



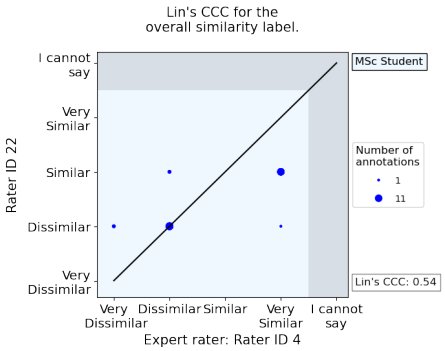
(q) Lin's CCC of Rater ID 19 with the standard rater.



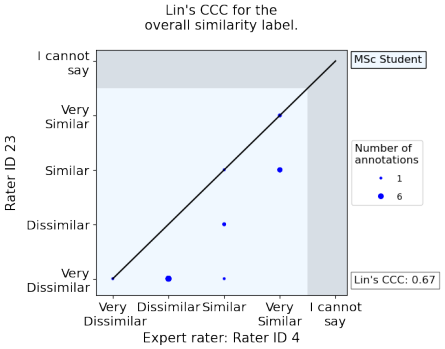
(r) Lin's CCC of Rater ID 20 with the standard rater.



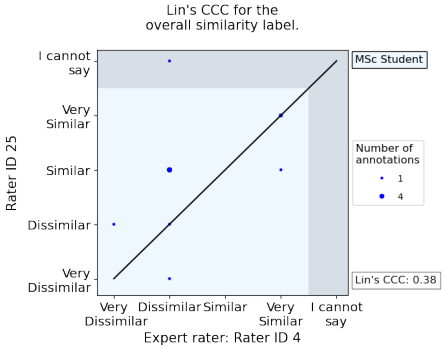
(s) Lin's CCC of Rater ID 21 with the standard rater.



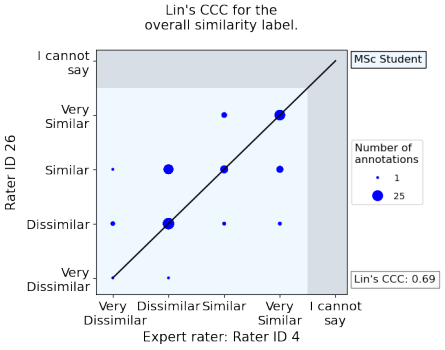
(t) Lin's CCC of Rater ID 22 with the standard rater.



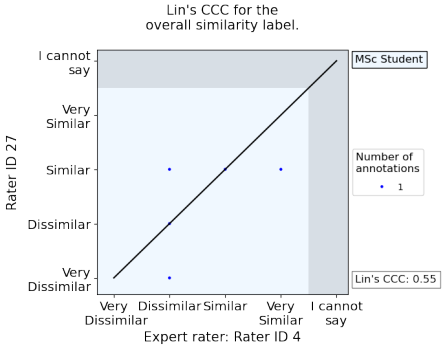
(u) Lin's CCC of Rater ID 23 with the standard rater.



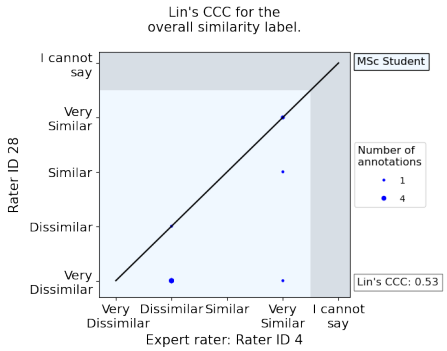
(v) Lin's CCC of Rater ID 25 with the standard rater.



(w) Lin's CCC of Rater ID 26 with the standard rater.



(x) Lin's CCC of Rater ID 27 with the standard rater.

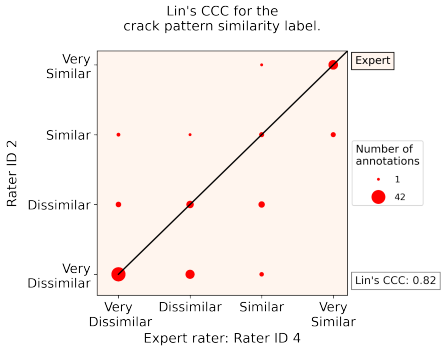


(y) Lin's CCC of Rater ID 28 with the standard rater.

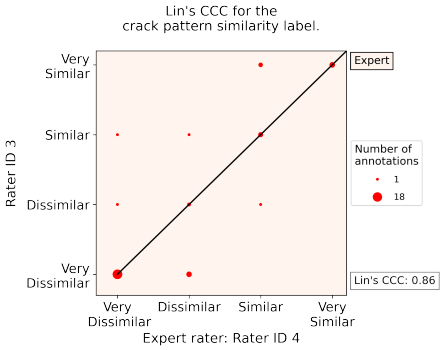
Figure D.3: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the overall similarity label for data generated from the statistics-based approach.

D.2 AGREEMENT WITH THE STANDARD RATER FOR THE DATA GENERATED FROM THE COMPUTATIONAL PHYSICS-BASED APPROACH

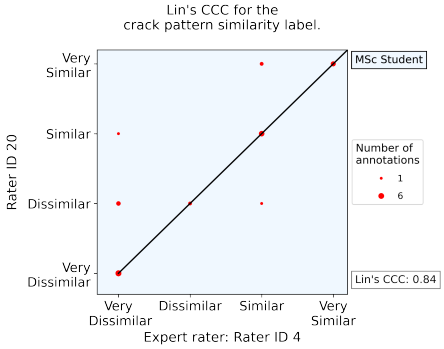
Crack Pattern Similarity Label



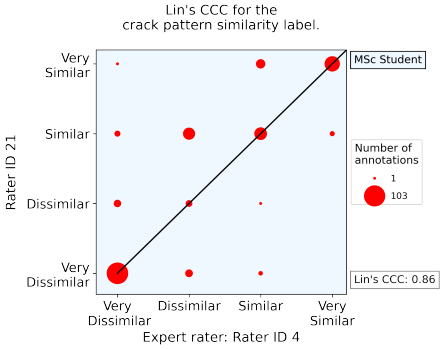
(a) Lin's CCC of Rater ID 2 with the standard rater.



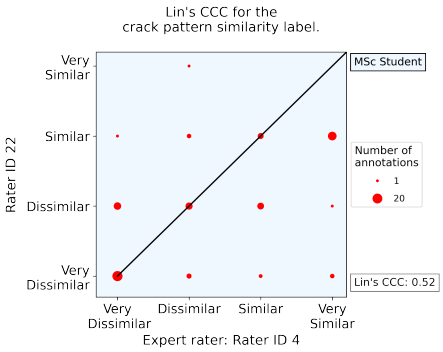
(b) Lin's CCC of Rater ID 3 with the standard rater.



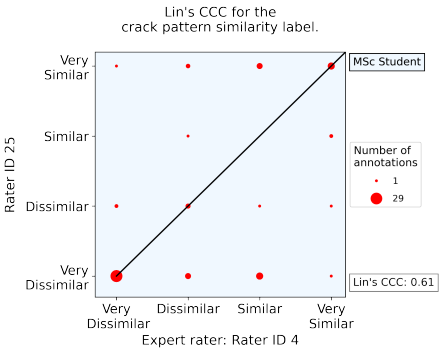
(c) Lin's CCC of Rater ID 20 with the standard rater.



(d) Lin's CCC of Rater ID 21 with the standard rater.



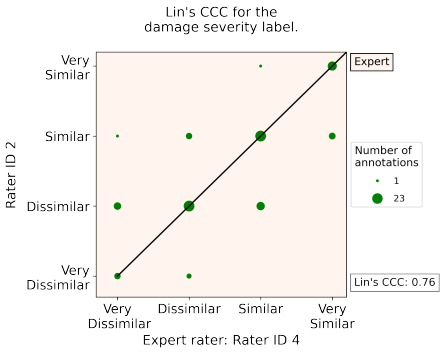
(e) Lin's CCC of Rater ID 22 with the standard rater.



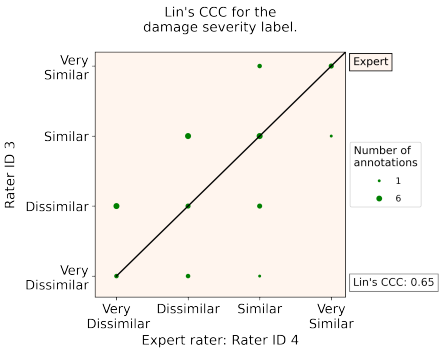
(f) Lin's CCC of Rater ID 25 with the standard rater.

Figure D.4: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the crack pattern similarity label for the data generated from the computational-physics based approach.

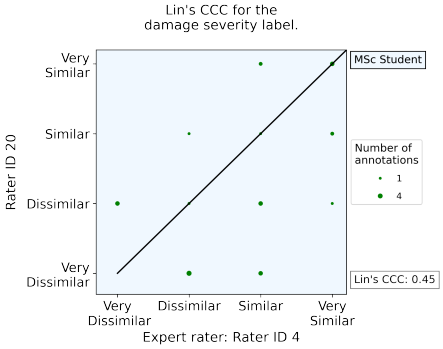
Damage Severity Label



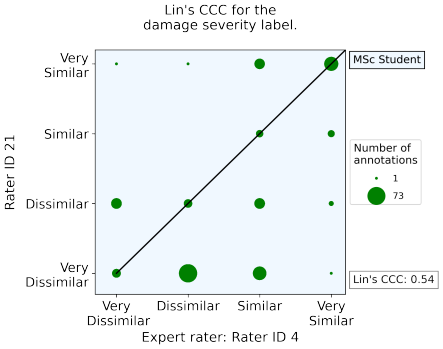
(a) Lin's CCC of Rater ID 2 with the standard rater.



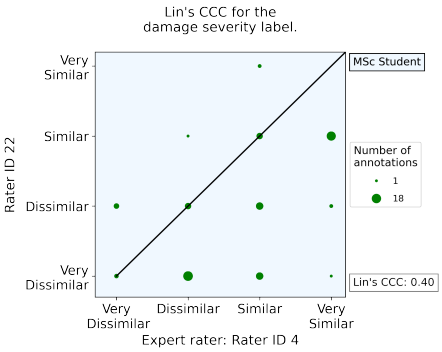
(b) Lin's CCC of Rater ID 3 with the standard rater.



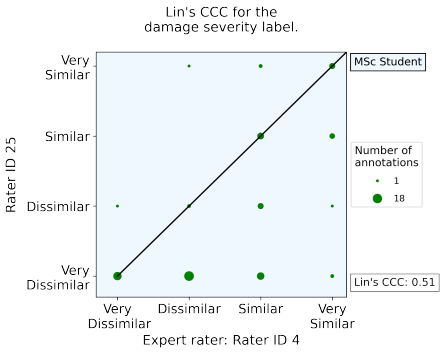
(c) Lin's CCC of Rater ID 20 with the standard rater.



(d) Lin's CCC of Rater ID 21 with the standard rater.



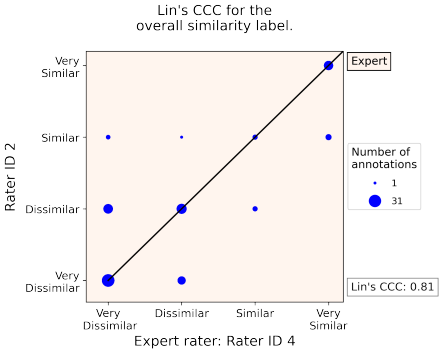
(e) Lin's CCC of Rater ID 22 with the standard rater.



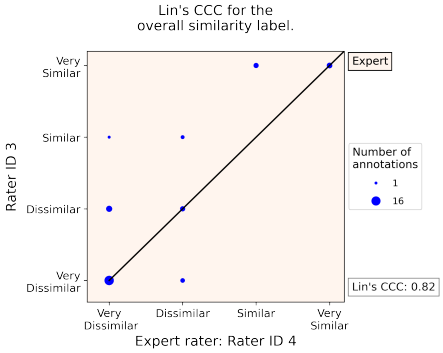
(f) Lin's CCC of Rater ID 23 with the standard rater.

Figure D.5: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the damage severity label for the data generated from the computational-physics based approach.

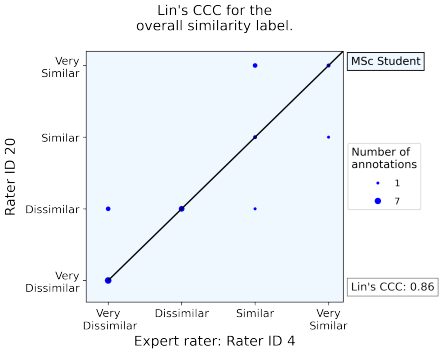
Overall Similarity Label



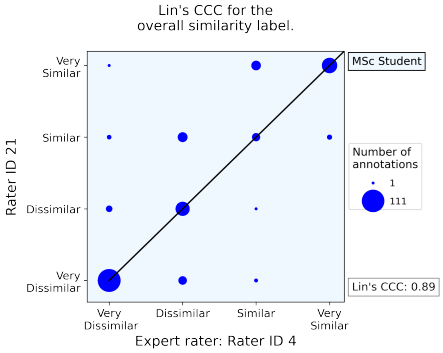
(a) Lin's CCC of Rater ID 2 with the standard rater.



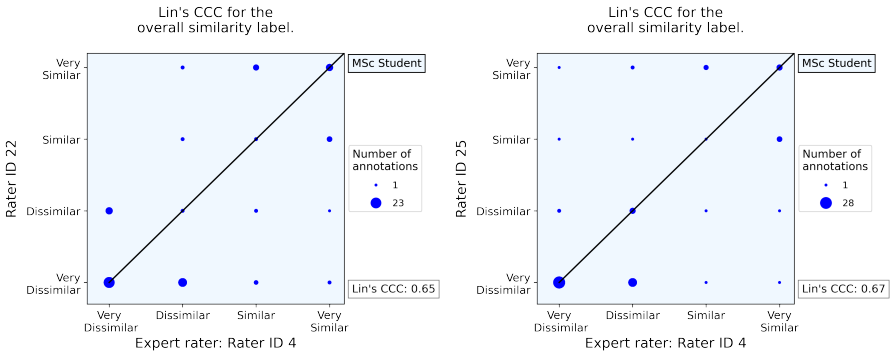
(b) Lin's CCC of Rater ID 3 with the standard rater.



(c) Lin's CCC of Rater ID 21 with the standard rater.



(d) Lin's CCC of Rater ID 22 with the standard rater.



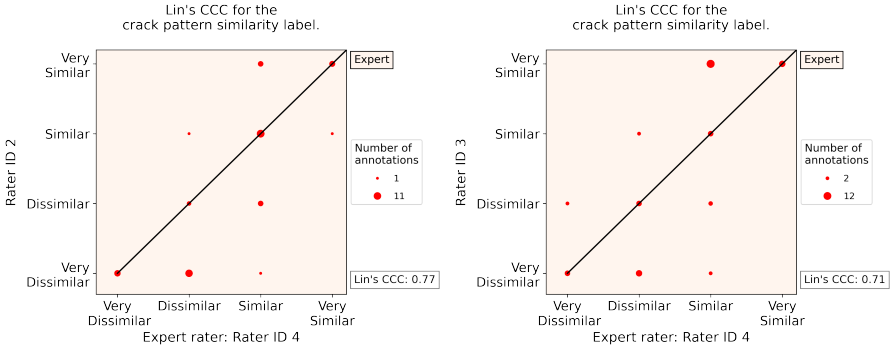
(e) Lin's CCC of Rater ID 23 with the standard rater.

(f) Lin's CCC of Rater ID 25 with the standard rater.

Figure D.6: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the overall similarity label for the data generated from the computational-physics based approach.

D.3 AGREEMENT WITH THE STANDARD RATER FOR THE DATA GENERATED FROM THE ALL THREE APPROACHES

Crack Pattern Similarity Label

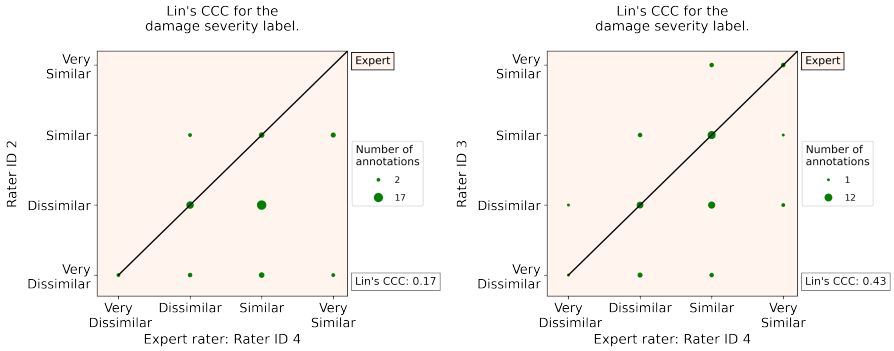


(a) Lin's CCC of Rater ID 2 with the standard rater.

(b) Lin's CCC of Rater ID 3 with the standard rater.

Figure D.7: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the crack pattern similarity label for the data generated from the statistics-based approach, computational-physics based approach and real-world images.

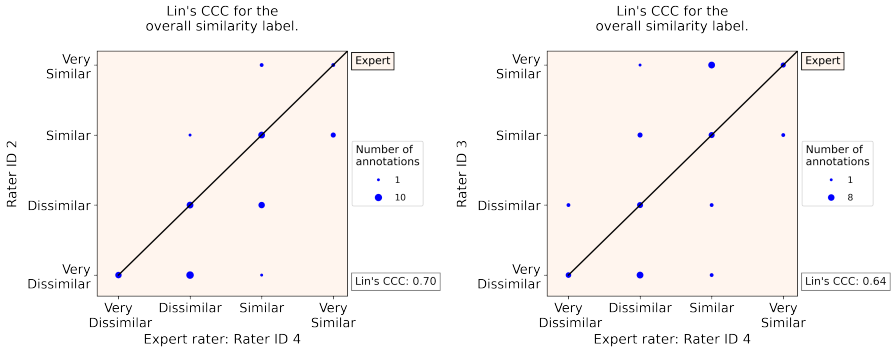
Damage Severity Label



(a) Lin's CCC of Rater ID 2 with the standard rater. **(b)** Lin's CCC of Rater ID 3 with the standard rater.

Figure D.8: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the damage severity label for the data generated from the statistics-based approach, computational-physics based approach and real-world images.

Overall Similarity Label



(a) Lin's CCC of Rater ID 2 with the standard rater. **(b)** Lin's CCC of Rater ID 3 with the standard rater.

Figure D.9: Lin's concordance correlation coefficient (CCC) of each rater with the standard rater for the overall similarity label for the data generated from the statistics-based approach, computational-physics based approach and real-world images.

E RECALL, PRECISION, AND F-SCORE OF THE SIMILARITY PREDICTION DONE BY THE CNN

Four different subsets of labelled crack pattern image pairs were generated and used to fit and test a neural network. Each of these subsets are further divided into smaller subsection of data, as discussed in Section 5.3. The precision, recall, and F-score of these fitted neural networks are discussed in this appendix. The definitions of these metrics are given in Section 3.7.

E.1 SUBSET 1: NO GENERALISATION

Precision

CRACK PATTERN SIMILARITY LABEL From Fig. E.1, it was seen that 93% was the highest magnitude of precision for the crack pattern similarity label, obtained when the neural network was fitted and tested with Subset 1.2. The lowest precision was seen for Subset 1.5, with a precision of 79%. This was obtained when the neural network was fitted and tested with the image pairs annotated by the MSc students. This also meant that, when the neural network was fitted with any of the subsets under Subset 1, the precision of the neural network in predicting the similarity rating of the crack pattern similarity label was always above 79%. This means that, of all the times that a similarity rating was predicted for an image pair, it was correctly predicted more than 79% of the time.

DAMAGE SEVERITY LABEL According to Fig. E.1, the neural network fitted and tested with Subset 1.2 again had the maximum magnitude of precision for the damage severity label, with a magnitude of 86%. With a precision of 74%, Subset 1.3

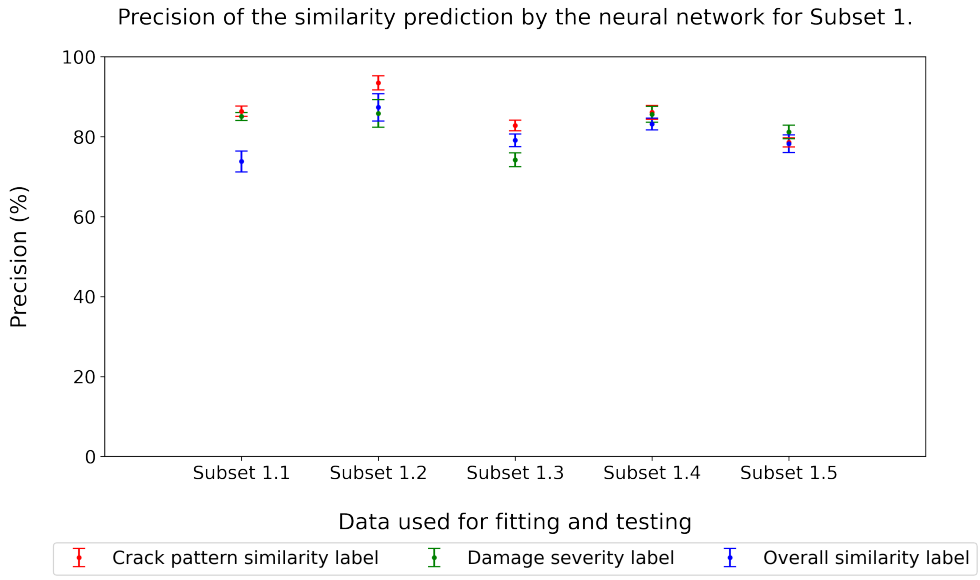


Figure E.1: Precision of all the labels for Subset 1.

had the lowest precision for the same task. The precision of the damage severity label was, in general, lower than that of the crack pattern similarity label, except in Subset 1.5 where it outperformed the other two labels.

OVERALL SIMILARITY LABEL Again, it was seen that the neural network fitted and tested with Subset 1.2 had the largest magnitude of precision for the overall similarity label, with a magnitude of 87%, according to Fig. E.1. The precision of the neural network fitted with Subset 1.1, had the least magnitude of 74%. Thus, both damage severity label and overall similarity labels were able to achieve a minimum precision of 79% with this subset.

Recall

CRACK PATTERN SIMILARITY LABEL As per Fig. E.2, the neural network fitted and tested with Subset 1.2 had the maximum magnitude of recall for the crack pattern similarity label, with a magnitude of 95%. This was the largest recall that any label could attain when the neural network was fitted with any annotated data in Subset 1. The recall of the neural network in predicting the similarity rating of the crack pattern similarity label was always above 81% , the lowest being for Subset 1.5. This means that, of all the times that a similarity rating was present in the test sample, it was correctly predicted more than 81% of the time.

DAMAGE SEVERITY LABEL The neural network fitted and tested with Subset 1.1 had the greatest magnitude of recall for the damage severity label, with a magnitude of 85%, as shown in Fig. E.2. This was closely followed by the recall achieved by Subsets 1.2 and 1.4, respectively. Only Subset 1.3 had a recall less than 80% when predicting the damage severity label.

OVERALL SIMILARITY LABEL As shown in Fig. E.2, the neural network fitted and tested using Subset 1.2 had the greatest magnitude of recall for the overall similarity label, with a magnitude of 86%. Subsets 1.4 and 1.5 were also able to achieve a recall rate of more than 80%. With a recall of 74%, Subset 1.1 had the lowest recall for the neural network's ability to predict the similarity rating of the overall similarity label.

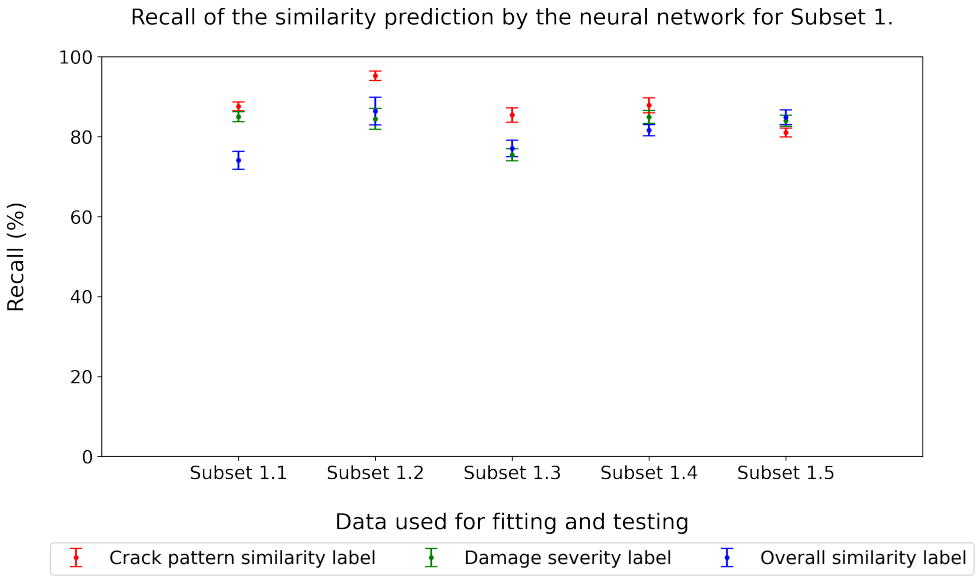


Figure E.2: Recall of all the labels for Subset 1.

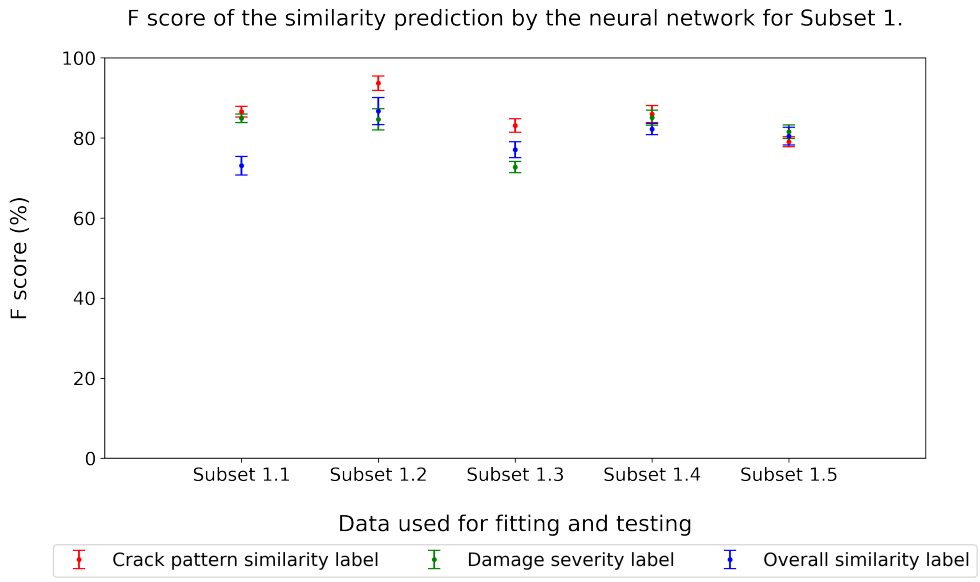


Figure E.3: F-score of all the labels for Subset 1.

F-score

CRACK PATTERN SIMILARITY LABEL The neural network fitted and tested with Subset 1.2 had the largest magnitude of F-score for the crack pattern similarity label, with a magnitude of 94%, as shown in Fig. E.3. The precision and recall achieved with this set were 95% and 93% respectively. A high F-score was achieved with a high precision and a high recall value. Subset 1.5 had the lowest F-score for the neural network's ability to predict the similarity rating of the crack pattern similarity label, with a magnitude of 79%. The corresponding precision and recall for this subset were 78% and 81%, respectively. Thus, it was seen that almost all data sets in Subset 1 were able to perform reasonably well for this label with a minimum F-score of 80%.

DAMAGE SEVERITY LABEL The damage severity label had the biggest magnitude of F-score for the neural network when fitted and tested with Subset 1.4, with a magnitude of 85% as shown in Fig. E.3. This F-score was achieved with a precision of 84% and a recall of 85%. Except for Subset 1.4, all the other subsets had an F-score greater than 80%.

OVERALL SIMILARITY LABEL In the neural network fitted and evaluated using Subset 1.2, the overall similarity label achieves the largest magnitude of F-score, with a magnitude of 86%. This is shown in Fig. E.3. This F-score was obtained with a precision of 87% and a recall of 86%. Subset 1.1 had the lowest F-score for the neural network's ability to predict the similarity rating of the overall similarity label, with a value of 73%. The related precision and recall for this subset were 73% and 74%, respectively.

E.2 SUBSET 2: GENERALISABILITY WITH RESPECT TO PATTERN ID

Precision

CRACK PATTERN SIMILARITY LABEL From Fig. E.4, it was seen that the highest magnitude of precision for the crack pattern similarity label was obtained the neural

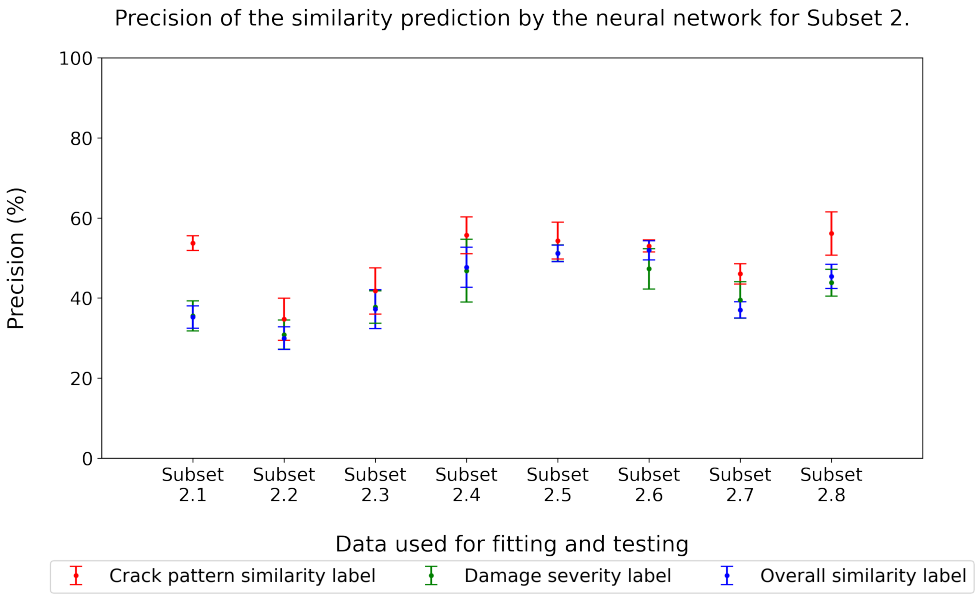


Figure E.4: Precision of all the labels for Subset 2.

network fitted and tested with Subset 2.8, with a magnitude of 56%. The lowest precision was seen for Subset 2.2, with a precision of only 34%. The standard deviations of the precision for each data set in Subset 2 was also seen to be very high.

DAMAGE SEVERITY LABEL According to Fig. E.4, the neural network fitted and tested with Subset 2.5 had the maximum magnitude of precision for the damage severity label, with a magnitude of 51%. The precision of the damage severity label was, in general, lower than that of the already poorly performing crack pattern similarity label. With a precision of only 31%, Subset 2.2 had the lowest precision for the same task. The standard deviations of precision for each data set in Subset 2 were also very high.

OVERALL SIMILARITY LABEL The precision of the overall similarity label fitted and tested with Subset 2 showed the same behaviour as that of the damage severity label, according to Fig. E.4. With Subset 2.5, the largest magnitude of precision for the overall similarity label was obtained at 52%. The least precision for the label was 30%, obtained with Subset 2.2. The precision standard deviations for each data set in Subset 2 were, similarly, extremely high.

Recall

CRACK PATTERN SIMILARITY LABEL As per Fig. E.5, the neural network fitted and tested with Subset 2.1 had the maximum magnitude of recall for the crack pattern similarity label, with a magnitude of 60%. Subset 2.3 was seen to have the lowest recall with a magnitude of 26%. In general, this label shows a very poor recall value.

DAMAGE SEVERITY LABEL The neural network fitted and tested with Subset 2.7 had the greatest magnitude of recall for the damage severity label, with a magnitude of 45%, as shown in Fig. E.5. The lowest recall for this label was seen when the neural network was fitted with Subset 2.3, with a magnitude of 22%.

OVERALL SIMILARITY LABEL As shown in Fig. E.5, the overall similarity label mimics the behaviour of the damage severity label closely. The largest recall for the overall similarity label was obtained when the neural network was fitted and tested

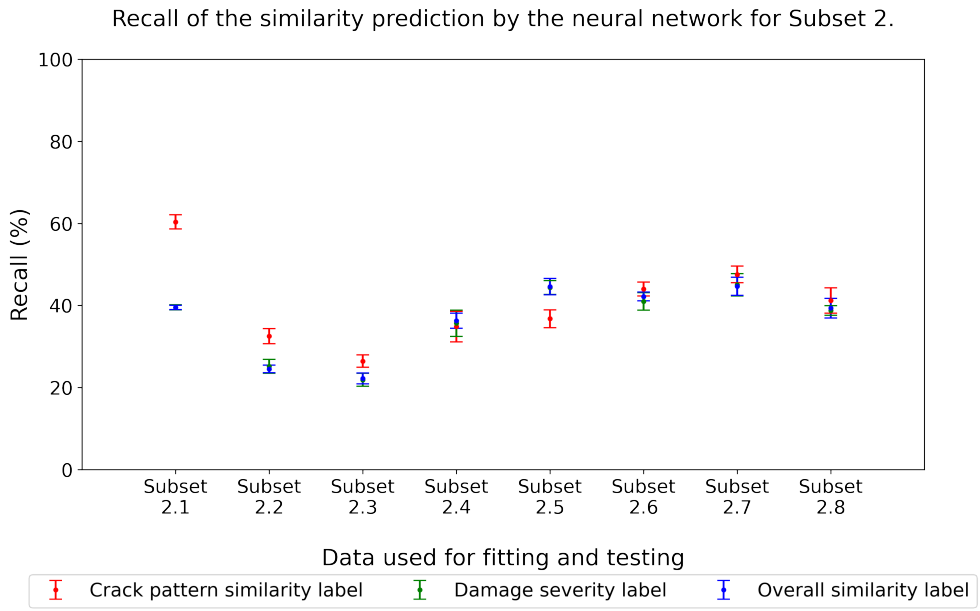


Figure E.5: Recall of all the labels for Subset 2.

using Subset 2.7, with a magnitude of 46%. As with the damage severity label, the lowest recall was seen for Subset 2.3, with a value of 22%.

F-score

CRACK PATTERN SIMILARITY LABEL The neural network fitted and tested with Subset 2.1 had the largest magnitude of F-score for the crack pattern similarity label, with a magnitude of 55%, as shown in Fig. E.6. The precision and recall achieved with this set were 53% and 60%. Subset 2.3 had the lowest F-score for the neural network's ability to predict the similarity rating of the crack pattern similarity label, with a F-score of 22%. The corresponding precision and recall for these subsets were 42% and 26%, respectively.

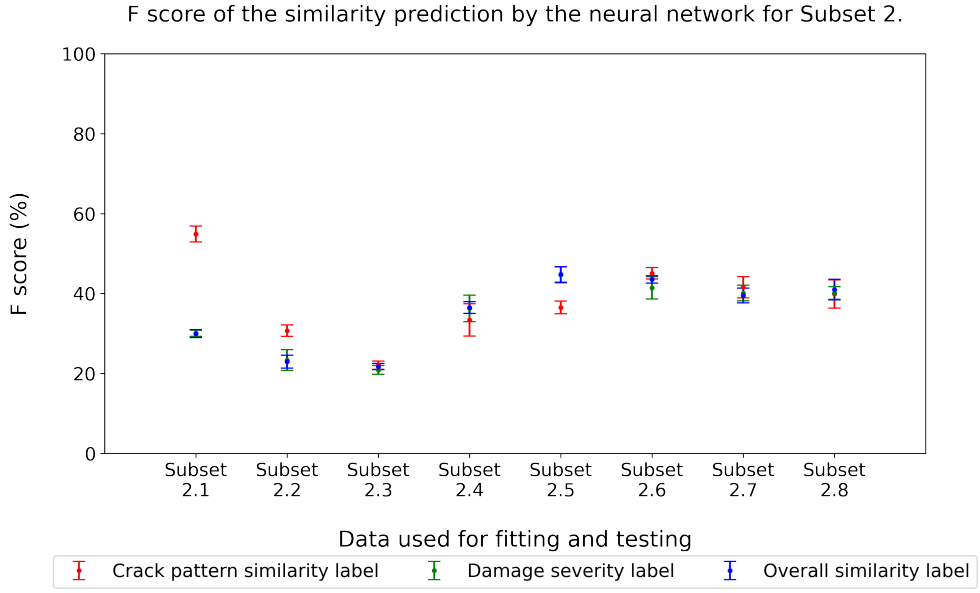


Figure E.6: F-score of all the labels for Subset 2.

DAMAGE SEVERITY LABEL The damage severity label had the biggest magnitude of F-score for the neural network fitted and tested with Subset 2.5, with a magnitude of 45%, as shown in Fig. E.6. This F-score was achieved with a precision of 51% and a recall of 44%. With a F-score of 21%, Subset 2.3 had the lowest F-score for the neural network's ability to predict the similarity rating of the damage severity label. For this subset, the associated precision and recall were 37% and 22%.

OVERALL SIMILARITY LABEL In the neural network fitted and evaluated using Subset 2.5, the overall similarity label had the largest magnitude of F-score, with a magnitude of 45%, as shown in Fig. E.6. This F-score was obtained with a precision of 51% and a recall of 45%. It was seen that the overall similarity label showed a behaviour similar to that of the damage severity label, with Subset 2.3 attaining the lowest F-score of 22% for this label. The related precision and recall for this subset were 37% and 22%, respectively.

E.3 SUBSET 3: GENERALISABILITY WITH RESPECT TO FAÇADE GEOMETRY

Precision

CRACK PATTERN SIMILARITY LABEL From Fig. E.7, it was seen that the highest magnitude of precision for the crack pattern similarity label was obtained the neural network fitted and tested with Subset 3.1, with a magnitude of 63%. The lowest precision was seen for Subset 3.5, with a precision of 43%.

DAMAGE SEVERITY LABEL According to Fig. E.7, the neural network fitted and tested with Subset 3.1 had the maximum magnitude of precision for the damage severity label, with a magnitude of 60%. With a precision of 32%, Subset 3.5 had the lowest precision for the same task.

OVERALL SIMILARITY LABEL The neural network fitted and tested with Subset 3.1 had the largest magnitude of precision for the overall similarity label, with a magnitude of 62%, according to Fig. E.7. Mirroring the damage severity label,

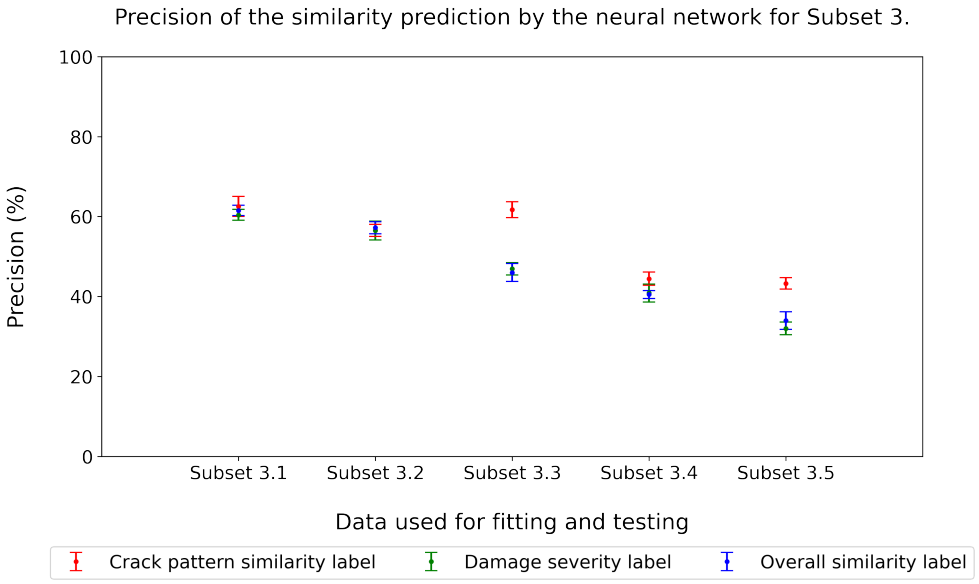


Figure E.7: Precision of all the labels for Subset 3.

Subset 3.5 had the lowest precision for the same assignment, with a precision of 34%.

Recall

CRACK PATTERN SIMILARITY LABEL As per Fig. E.8, the neural network fitted and tested with Subset 3.1 had the maximum magnitude of recall for the crack pattern similarity label, with a magnitude of 59%. With a recall of 38%, Subset 3.4 had the lowest recall for the neural network's ability to predict the similarity rating of the crack pattern similarity label.

DAMAGE SEVERITY LABEL The neural network fitted and tested with Subset 3.1 had the greatest magnitude of recall for the damage severity label, with a magnitude of 55%, as shown in Fig. E.8. Subset 3.5 had the lowest recall for the neural network's ability to predict the similarity rating of the damage severity label, with a recall of just 34%.

OVERALL SIMILARITY LABEL As shown in Fig. E.8, the neural network fitted and tested using Subset 3.1 had the greatest magnitude of recall for the overall similarity label, with a magnitude of 56%. With a recall of only 36%, Subset 3.5 had the lowest recall for the neural network's ability to predict the similarity rating of the overall similarity label.

F-score

CRACK PATTERN SIMILARITY LABEL The neural network fitted and tested with Subset 3.1 had the largest magnitude of F-score for the crack pattern similarity label, with a magnitude of 58%, as shown in Fig. E.9. The precision and recall achieved with this set were 58% and 62%, respectively. Subset 3.4 had the lowest F-score for the neural network's ability to predict the similarity rating of the crack pattern similarity label, with a F-score of 36%. The corresponding precision and recall for this subset were 44% and 37%, respectively.

DAMAGE SEVERITY LABEL The damage severity label had the biggest magnitude of F-score for the neural network fitted and tested with Subset 3.1 with a magnitude of 55% , as shown in Fig. E.9. This F-score was achieved with a precision of 55%

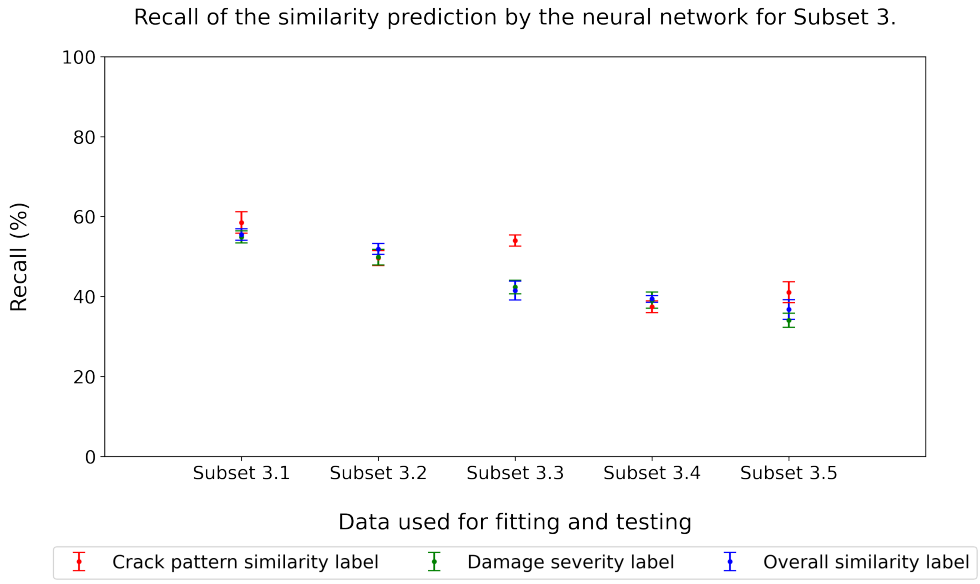


Figure E.8: Recall of all the labels for Subset 3.

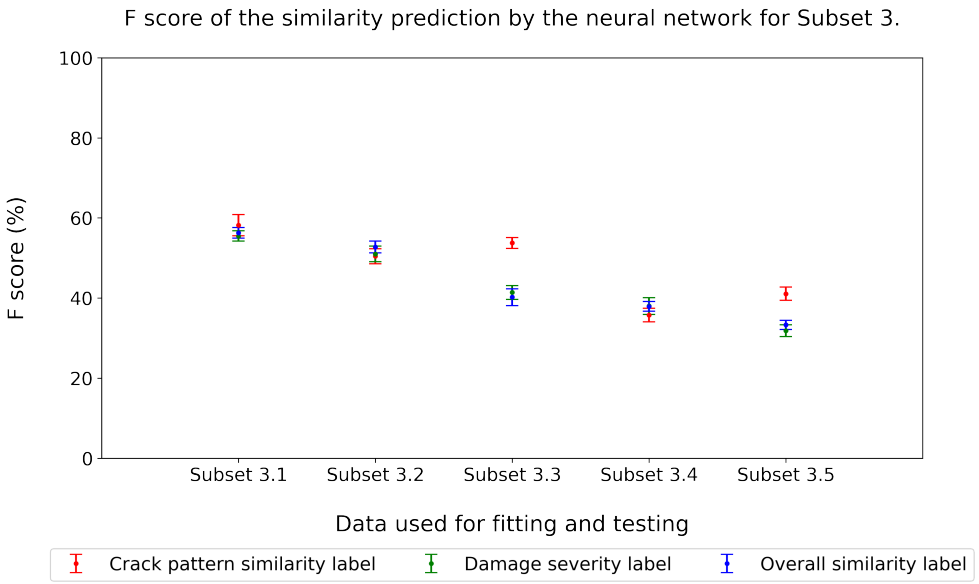


Figure E.9: F-score of all the labels for Subset 3.

and a recall of 60%. With a F-score of 32%, Subset 3.5 had the lowest F-score for the neural network's ability to predict the similarity rating of the damage severity label. For this subset, the associated precision and recall were 32% and 34%.

OVERALL SIMILARITY LABEL In the neural network fitted and evaluated using Subset 3.1, the overall similarity label had the largest magnitude of F-score, with a magnitude of 56%, as shown in Fig. E.9. This F-score was obtained with a precision of 62% and a recall of 55%. Subset 3.5 had the lowest F-score for the neural network's ability to predict the similarity rating of the overall similarity label, with a F-score of 33%. The related precision and recall for this subset were 34% and 37%, respectively.

E.4 SUBSET 4: GENERALISABILITY WITH RESPECT TO STATISTICALLY GENERATED IMAGES AND REAL-WORLD IMAGES

Precision

CRACK PATTERN SIMILARITY LABEL From Fig. E.10, it was seen that the highest magnitude of precision for the crack pattern similarity label was obtained the neural network fitted and tested with Subset 4.1, with a magnitude of 28%. The lowest precision was seen for Subset 4.3, with a precision of 21%.

DAMAGE SEVERITY LABEL According to Fig. E.10, the neural network fitted and tested with Subset 4.2 had the maximum magnitude of precision for the damage severity label, with a magnitude of 43%. With a precision of 28%, Subset 4.5 had the lowest precision for the same task.

OVERALL SIMILARITY LABEL The neural network fitted and tested with Subset 4.2 had the largest magnitude of precision for the overall similarity label with a magnitude of 44%, according to Fig. E.10. Subset 4.4 had the lowest precision for the same assignment, with a precision of 25%.

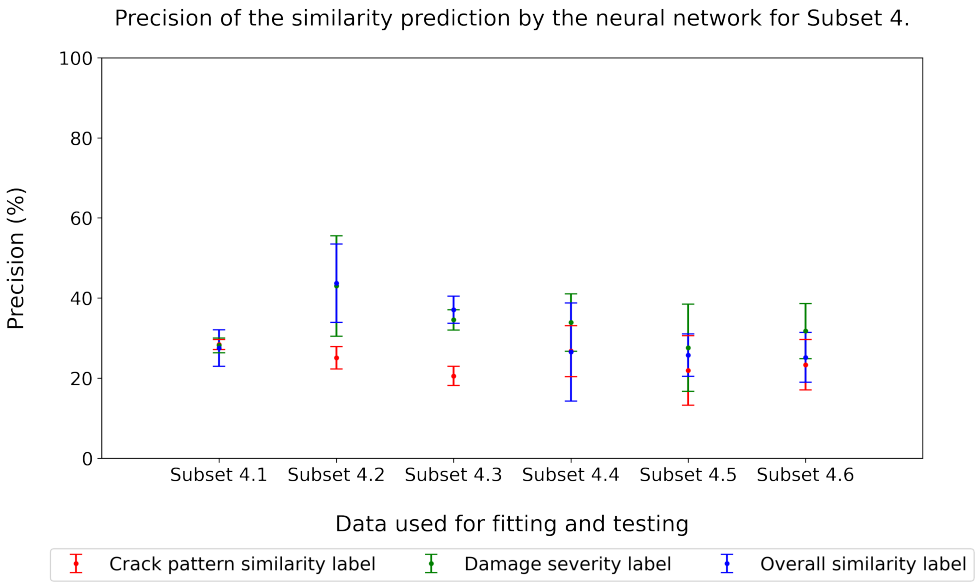


Figure E.10: Precision of all the labels for Subset 4.

Recall

CRACK PATTERN SIMILARITY LABEL As per Fig. E.11, the neural network fitted and tested with Subset 4.4 had the maximum magnitude of recall for the crack pattern similarity label, with a magnitude of 31%. With a recall of 16%, Subset 4.5 had the lowest recall for the neural network's ability to predict the similarity rating of the crack pattern similarity label.

DAMAGE SEVERITY LABEL The neural network fitted and tested with Subset 4.2 had the greatest magnitude of recall for the damage severity label, with a magnitude of 38%, as shown in Fig. E.11. Subset 4.6 had the lowest recall for the neural network's ability to predict the similarity rating of the damage severity label, with a recall of just 19%.

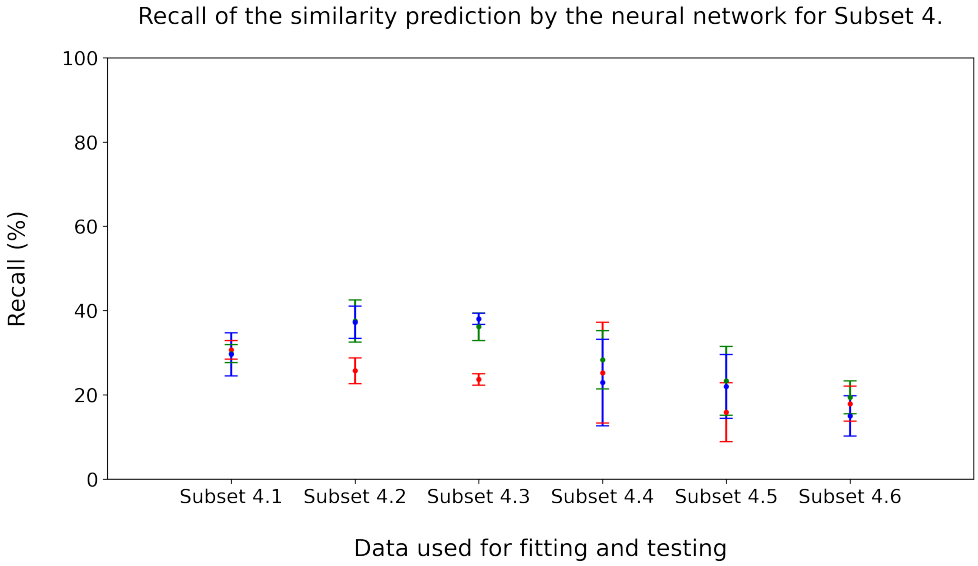


Figure E.11: Recall of all the labels for Subset 4.

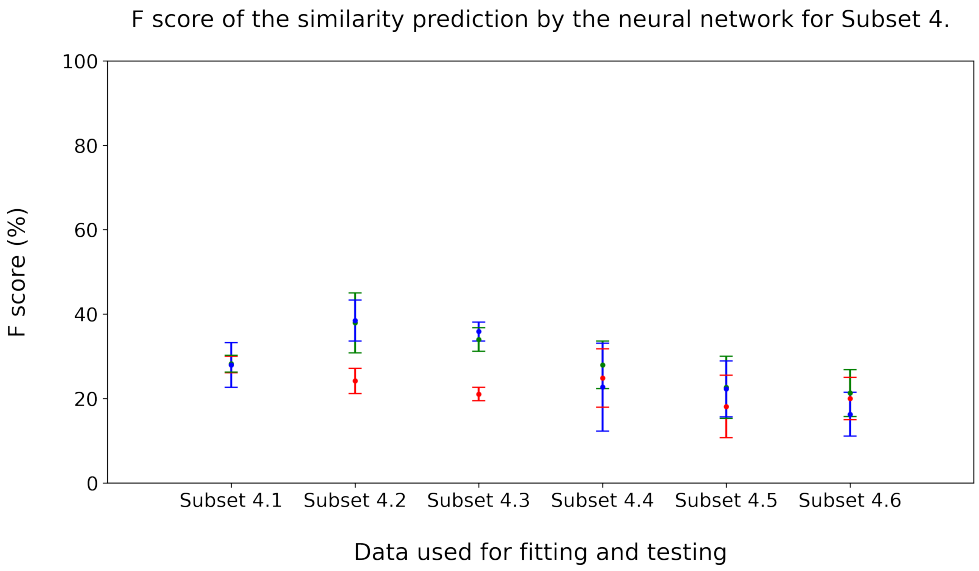


Figure E.12: F-score of all the labels for Subset 3.

OVERALL SIMILARITY LABEL As shown in Fig. E.11, the neural network fitted and tested using Subset 4.3 had the greatest magnitude of recall for the overall similarity label, with a magnitude of 38%. With a recall of only 15%, Subset 4.6 had the lowest recall for the neural network's ability to predict the similarity rating of the overall similarity label.

F-score

CRACK PATTERN SIMILARITY LABEL The neural network fitted and tested with Subset 4.1 had the largest magnitude of F-score for the crack pattern similarity label, with a magnitude of 28%, as shown in Fig. E.12. The precision and recall achieved with this set were 28% and 31%, respectively. Subset 4.5 had the lowest F-score for the neural network's ability to predict the similarity rating of the crack pattern similarity label, with a F-score of 18%. The corresponding precision and recall for this subset were 22% and 16%, respectively.

DAMAGE SEVERITY LABEL The damage severity label had the biggest magnitude of F-score for the neural network fitted and tested with Subset 4.2 with a magnitude of 38% , as shown in Fig. E.12. This F-score was achieved with a precision of 43% and a recall of 38%. With a F-score of 21%, Subset 4.6 had the lowest F-score for the neural network's ability to predict the similarity rating of the damage severity label. For this subset, the associated precision and recall were 31% and 19%.

OVERALL SIMILARITY LABEL In the neural network fitted and evaluated using Subset 4.2, the overall similarity label had the largest magnitude of F-score, with a magnitude of 38%, as shown in Fig. E.12. This F-score was obtained with a precision of 43% and a recall of 37%. Subset 4.6 had the lowest F-score for the neural network's ability to predict the similarity rating of the overall similarity label, with a F-score of 16%. The related precision and recall for this subset were 25% and 15%, respectively.