DYNAMIC DETECTION OF SURFACE CRACKS USING DEEP LEARNING AND COMPUTER VISION

PARSAPU KAMAL RAJ

TUHIN ANAND

SHASHANK PANDEY

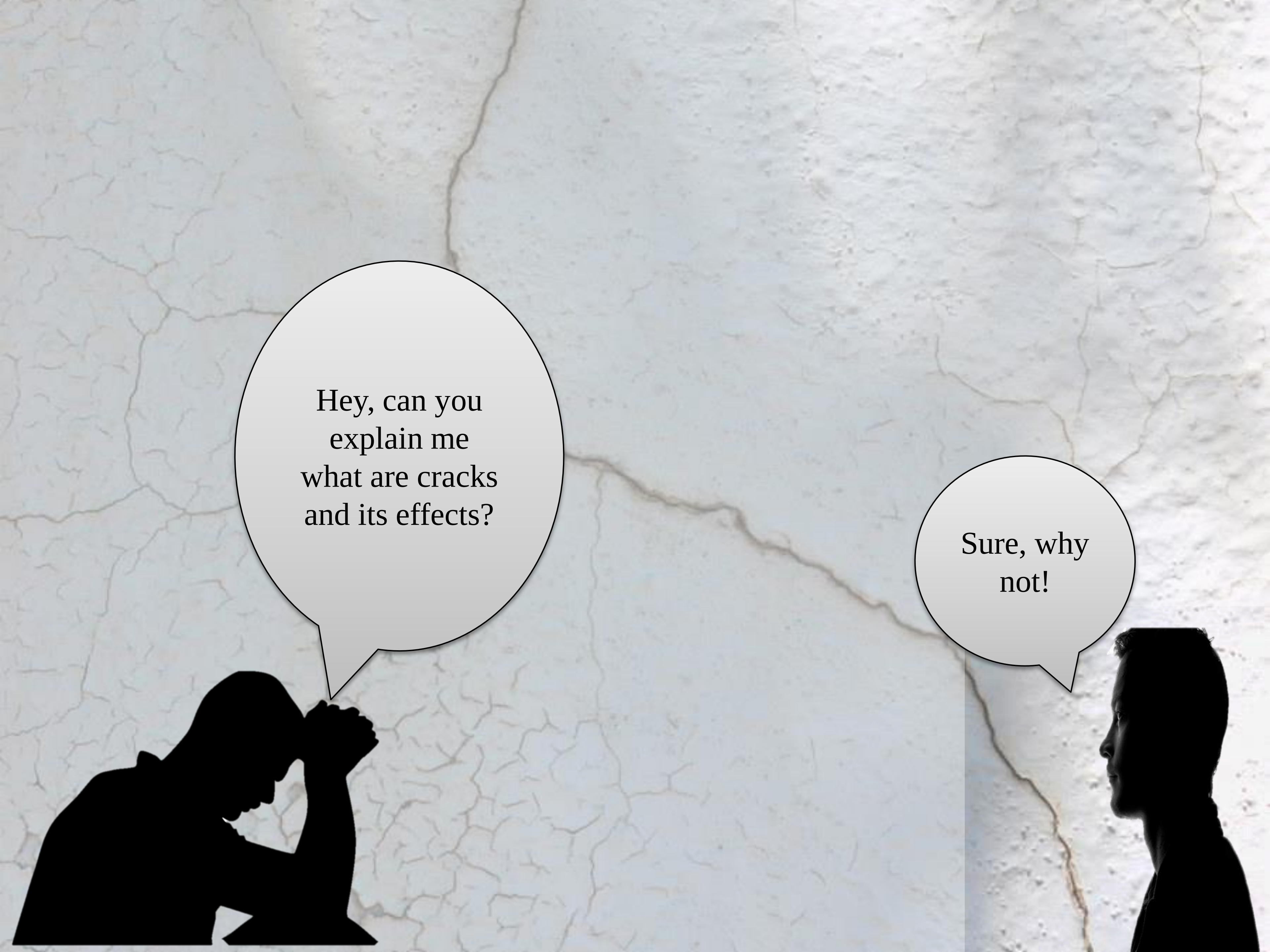
CONTENTS

OBJECTIVE

Our objective is to create a model that can be able to dynamically detect cracks in any engineering structure using the concept of Deep learning and Computer vision.

This can prove to be very helpful in the future when made commercially viable with few modifications and more extensive model training.

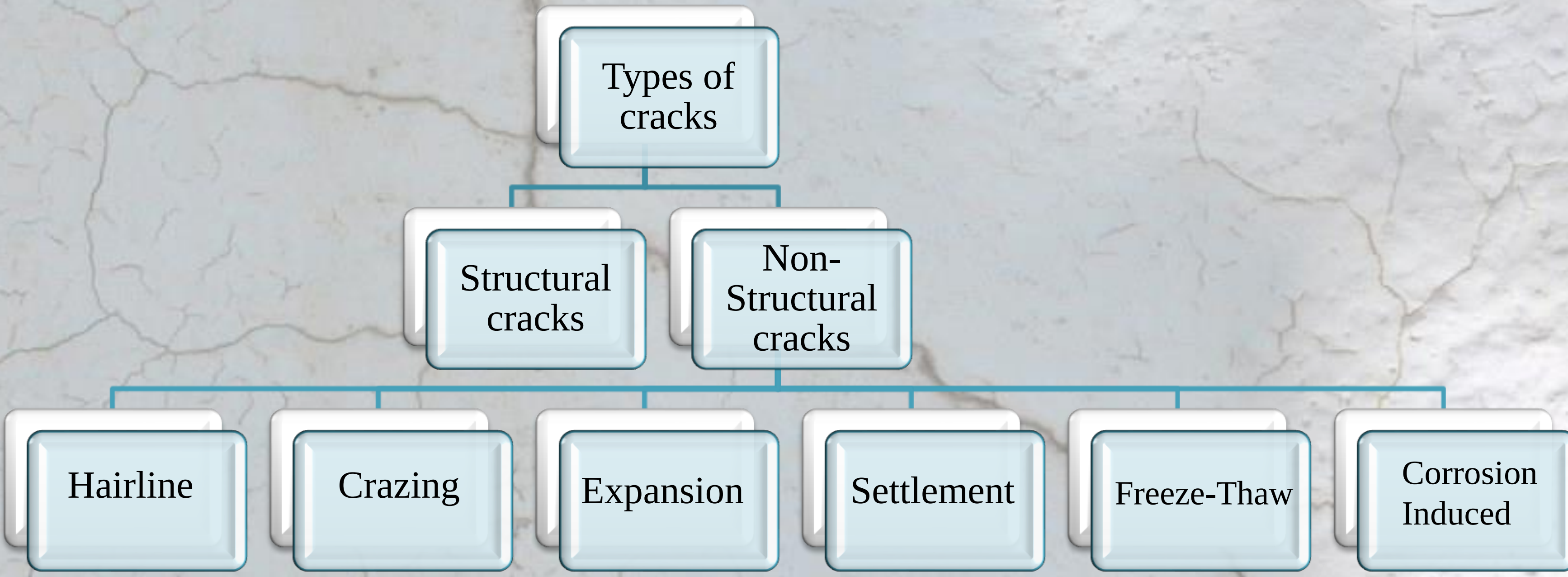
On our way to create such a model we also learnt about cracks extensively from its types to various impacts.


Hey, can you explain me what are cracks and its effects?
Sure, why not!


In civil engineering, a crack refers to a visible fissure or separation in a structure, such as a building or pavement, often caused by factors like stress, settlement, or material deterioration. Cracks can impact the integrity of the structure and may require evaluation and repair to ensure safety and longevity.
Oh……Got it! What about its impacts?

Impact of cracks on structure

- Structural Integrity
- Durability
- Corrosion of steel reinforcement
- Serviceability issues
- Safety concerns
- Repair and maintenance costs
- Impact on functionality
- Environmental impact


Types of cracks
Structural cracks
Non-Structural cracks
Hairline
Crazing
Expansion
Settlement
Freeze-Thaw
Corrosion Induced

Summary of the literature reviews

The literature reviews highlight various methodologies and techniques employed by different researchers for the detection of cracks in structures, particularly focusing on digital image processing technology. Each study introduces a unique approach, combining different methods and technologies to achieve accurate and efficient crack detection. Here are some key conclusions drawn from the literature reviews:

- **Diverse Approaches**
- **Integration of Technologies**
- **Pre-processing and Feature Extraction**
- **Adaptation of Algorithms**
- **Efficiency and Computation Time**
- **Use of Machine Learning and deep learning**
- **Application Specificity**
- **Advancements in Image-Based Techniques**
- **Automation and Unsupervised Training**

MATERIALS AND METHODS

In this project we have utilised Deep learning, computer vision to a great extent. Although project is not perfect and certainly requires some more efficient modifications to be commercially or industrial used, it definitely can serve the testing purpose in a short run.

We have put our whole hearted efforts going into this project and have obtained some great results with the different models.

We classify the surface cracks using 3 different base models with transfer learning:

- **ResNet v2**
- **VGG16**
- **Xception**

Finally, we will display the activation maps using Grad-Cam.

The best feature is, it can be used in real time using phone camera and results can be simultaneously displayed or viewed on the laptop with high accuracy.

There is very slight time lag of 1 or 2 seconds when you are continuously moving the camera and suddenly focus on it on some specific crack or surface.

This issue could be addressed with a bit powerful computer system that has higher RAM.

- ***ResNet v2***, known for its exceptional ability to handle the challenges of training very deep networks, is employed to capture intricate features within images. The introduction of skip connections in ResNet v2 facilitates the learning of residual functions, enabling more effective representation of complex crack patterns.

- ***VGG16***, with its straightforward architecture comprising multiple convolutional layers, is chosen for its simplicity and efficiency. The small (3x3) convolutional filters in VGG16 aid in learning hierarchical features, making it adept at capturing nuanced details in images of structural elements.

- ***Xception***, a model inspired by the Inception architecture, brings efficiency to the project through depth-wise separable convolutions. By separating spatial and channel-wise operations, Xception optimizes parameter usage, making it well-suited for crack detection tasks where resource efficiency is crucial.

In this project, we take a dataset comprising 40,000 images with positive and negative labels for cracks. The dataset includes 20,000 images labeled as crack-positive and 20,000 images labeled as crack-negative. This diverse dataset enables comprehensive training and evaluation of the models across various crack types and structural materials.

In this project, we aim to compare the performance of ResNet v2, VGG16, and Xception in the context of crack detection, considering factors such as accuracy, speed, and generalization to different types of cracks and structural materials. The insights gained from this study can contribute to the development of robust and efficient crack detection systems, enhancing the overall safety and resilience of critical infrastructure.

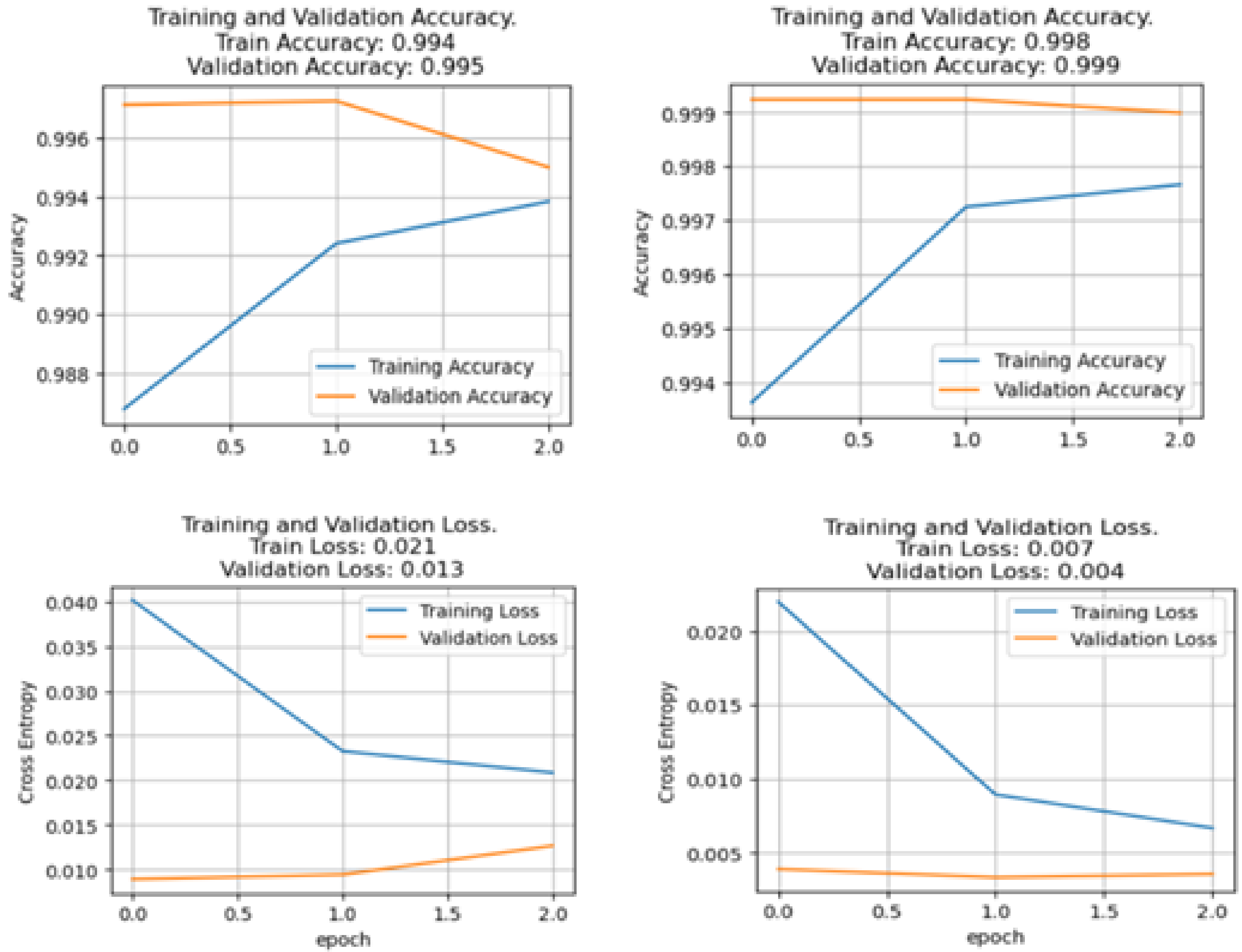
Results and Discussion

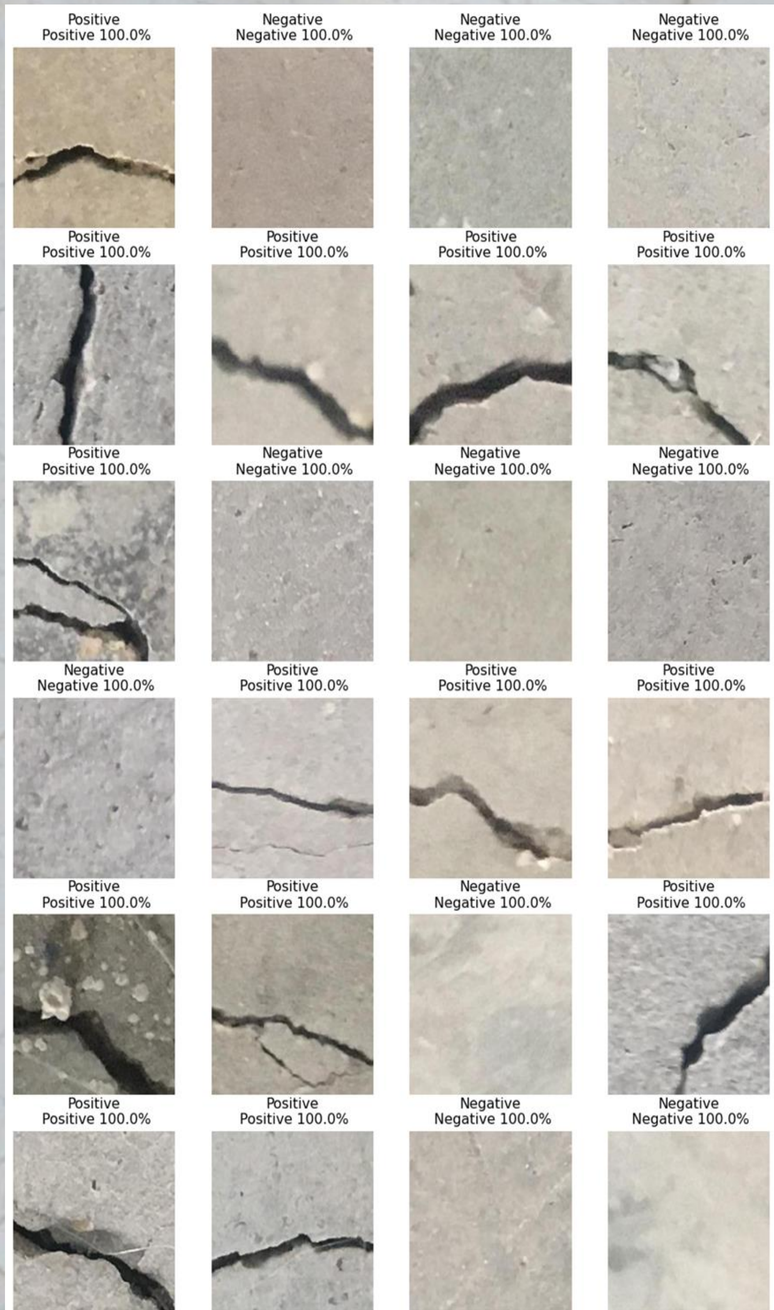
- *ResNet v2*

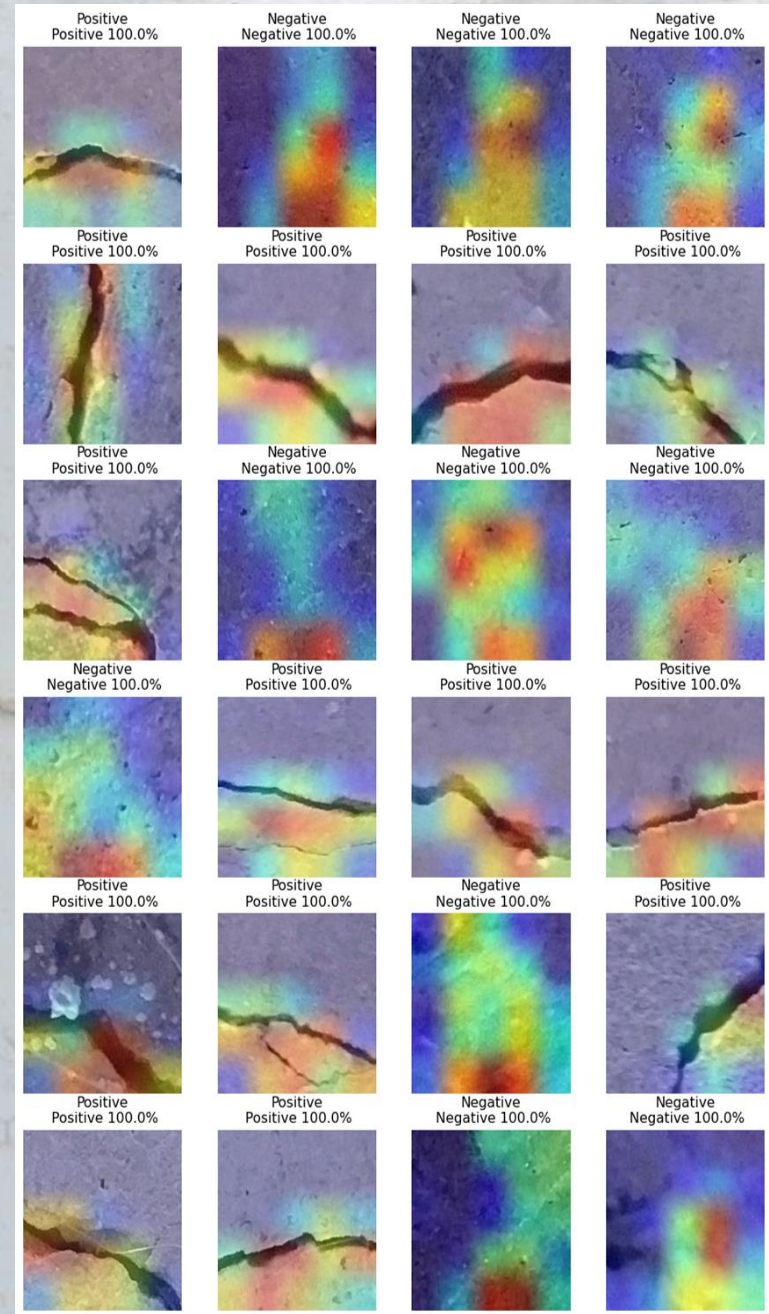
ResNet v2, an evolution of the original ResNet architecture, represents a groundbreaking approach in deep neural networks. Introduced as a solution to the vanishing gradient problem, ResNet v2 incorporates a residual learning framework that utilizes identity connections to address the challenges associated with training extremely deep networks. Unlike its predecessor, ResNet v2 places the activation and batch normalization prior to weight layers, creating a pre-activation residual block. This strategic modification enables smoother gradient flow during backpropagation, enhancing the model's training efficiency. The architecture comprises bottleneck blocks, integrating 1x1, 3x3, and 1x1 convolutions, optimizing computational efficiency without compromising expressive power. The residual connections facilitate the training of exceedingly deep networks, with ResNet-152 being a notable example, achieving remarkable accuracy in image classification tasks. Furthermore, ResNet v2 introduces the concept of skip connections, allowing the direct flow of information between layers, thereby mitigating the risk of vanishing gradients. This innovative design not only enhances training but also contributes to improved model generalization and accuracy. Overall, ResNet v2 stands as a testament to the continual refinement of deep neural network architectures, emphasizing the importance of residual connections in facilitating the training of exceptionally deep networks with superior performance across various computer vision tasks.

- ***Results***

The ResNet v2 model, with a base of 780 layers, demonstrates remarkable performance in the given task. After the training process, a strategic decision to unfreeze 195 layers in the base model was made, allowing the model to fine-tune its learned representations for the specific dataset. This fine-tuning significantly contributes to the impressive accuracy achieved, reaching an astounding 99.89%. The minimal loss value of 0.004 further underscores the model's effectiveness in minimizing prediction errors. These results underscore the capability of the Inception architecture to generalize well to the complexities of the dataset, showcasing its robust feature extraction and classification capabilities. The high accuracy and low loss indicate a well-optimized model that can confidently and accurately classify images within the specified categories, affirming the success of the training and fine-tuning process.


Training and Validation Accuracy.
Train Accuracy: 0.994
Validation Accuracy: 0.995
Accuracy
Training Accuracy
Validation Accuracy
Training and Validation Accuracy.
Train Accuracy: 0.998
Validation Accuracy: 0.999
Accuracy
Training Accuracy
Validation Accuracy
Training and Validation Loss.
Train Loss: 0.021
Validation Loss: 0.013
Cross Entropy
Training Loss
Validation Loss
epoch
Training and Validation Loss.
Train Loss: 0.007
Validation Loss: 0.004
Cross Entropy
Training Loss
Validation Loss
epoch


Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%


Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%

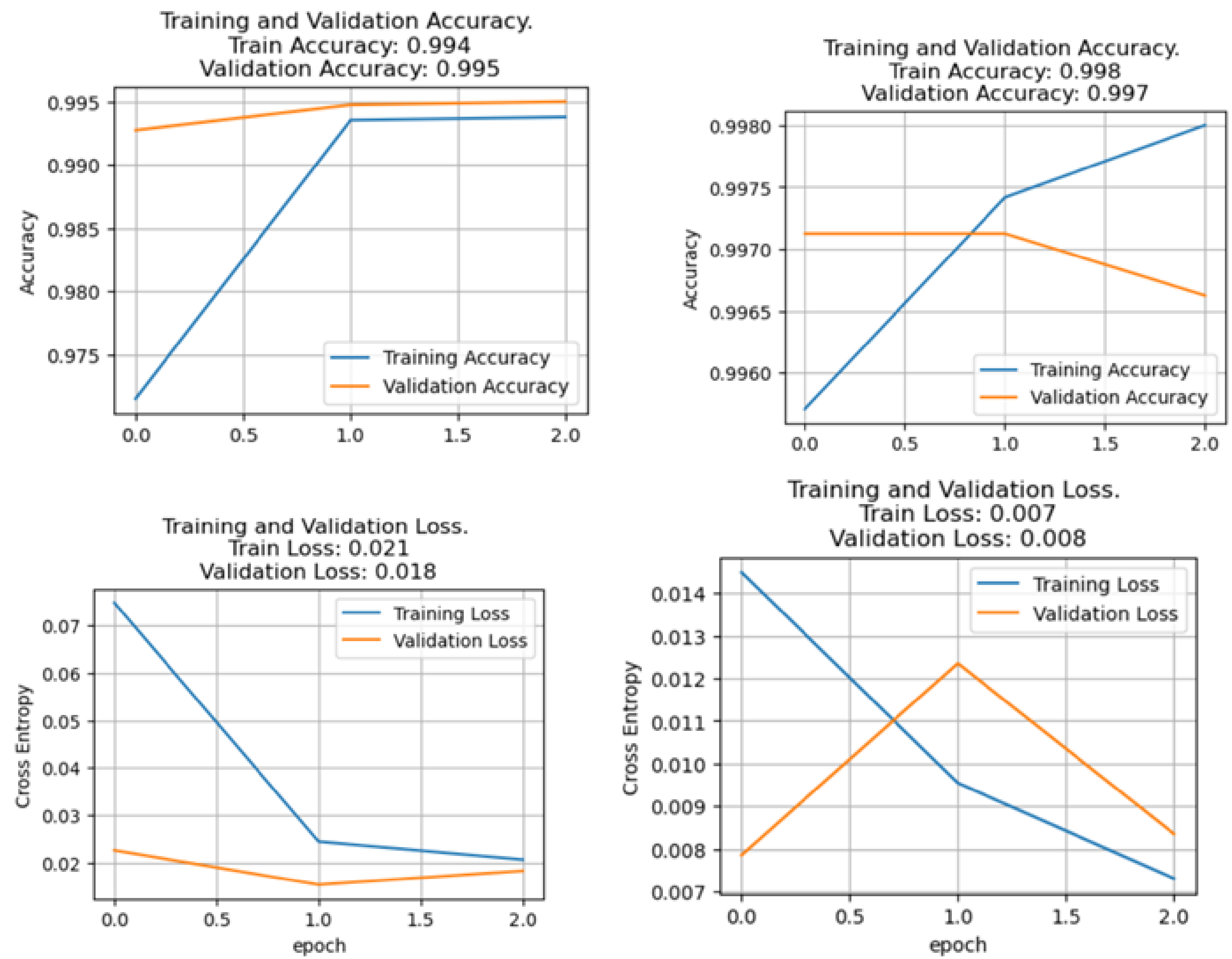
- ***VGG16***

VGG-16, or Very Deep Convolutional Networks for Large-Scale Image Recognition with 16 layers, stands as a pioneering convolutional neural network (CNN) architecture designed by K. Simonyan and A. Zisserman. Introduced to address the shortcomings of previous models like AlexNet, VGG-16 employs a unique structure characterized by small 3x3 receptive fields and a 1-pixel stride. This design choice, deviating from larger filters used in earlier models, enhances the network's ability to create powerful decision functions with a higher number of non-linear activation layers, facilitating faster convergence during training.

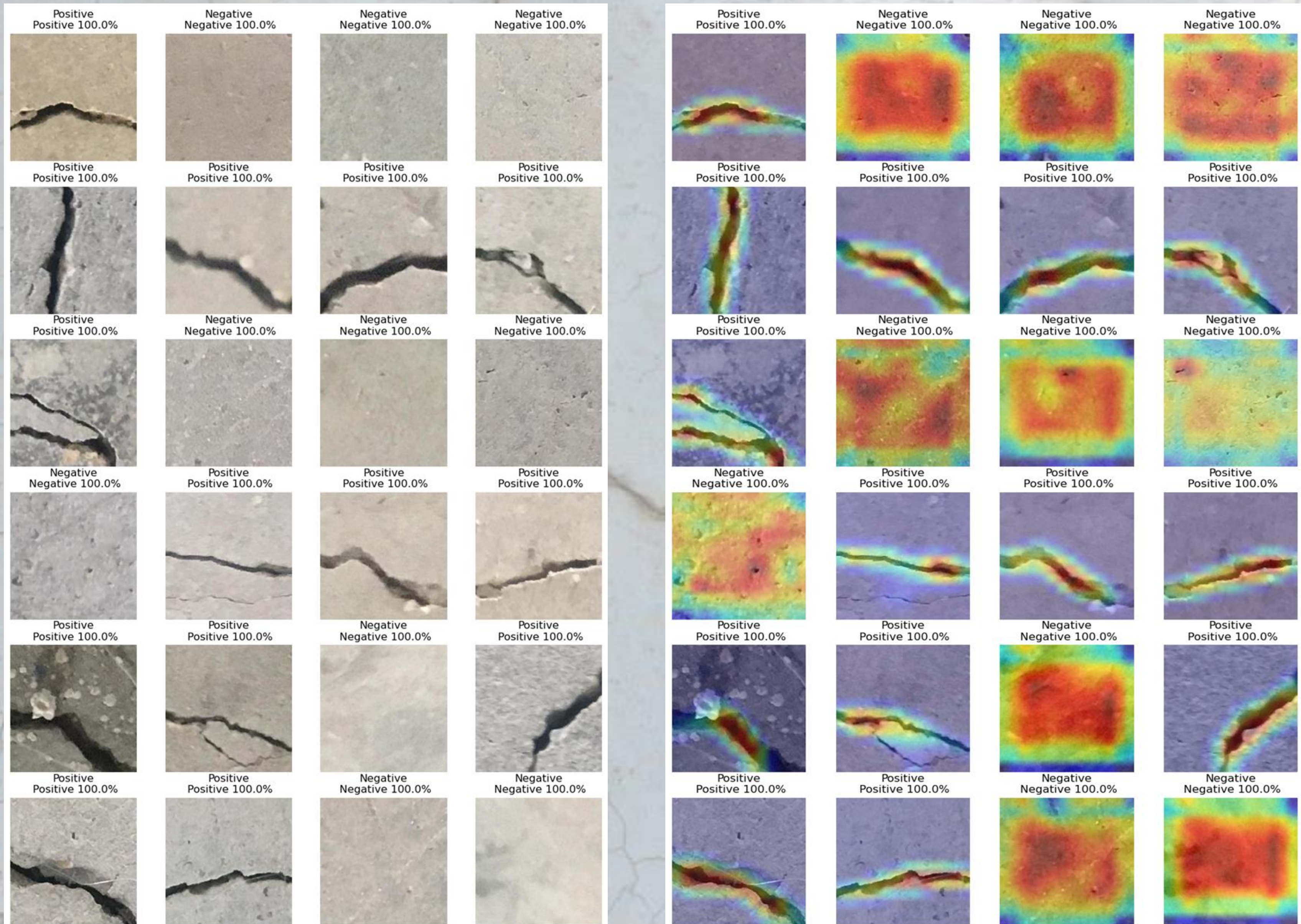
In comparison to its predecessor, VGG-16 improves accuracy by embracing more layers. However, the model's extensive depth could lead to convergence challenges. The introduction of the Residual Network (ResNet) architecture further enhanced training efficiency by employing skip connections, allowing inputs to bypass certain convolutional layers and reducing training time. Despite its remarkable performance, achieving a 7.0% test error in the ILSVRC competitions, VGG-16 is associated with extended training times and a large model size of 500MB. Modern architectures have since adopted innovations like skip connections and inceptions, contributing to improved efficiency in terms of both accuracy and training time.

- ***Results***

The results of the trained VGG16 model demonstrate a remarkable accuracy of 99.81% on the given task. This signifies the effectiveness of the model in correctly classifying the input data. The corresponding loss value of 0.007 indicates a minimal discrepancy between the predicted and actual values, further validating the model's robust performance. The utilization of 19 base layers in the VGG16 architecture, with a strategic decision to unfreeze only 4 layers during training, highlights a careful balance between model complexity and fine-tuning. This approach is essential for preventing overfitting and ensuring generalization to unseen data. The unfreezing of specific layers allows the model to adapt to task-specific features while retaining the knowledge gained from the pre-trained layers. These results collectively underscore the success of the training process, showcasing the model's ability to capture intricate patterns and features within the dataset.



Training and Validation Accuracy.
Train Accuracy: 0.994
Validation Accuracy: 0.995
Accuracy
Training Accuracy
Validation Accuracy
Training and Validation Accuracy.
Train Accuracy: 0.998
Validation Accuracy: 0.997
Accuracy
Training Accuracy
Validation Accuracy
Training and Validation Loss.
Train Loss: 0.021
Validation Loss: 0.018
Cross Entropy
epoch
Training Loss
Validation Loss
Training and Validation Loss.
Train Loss: 0.007
Validation Loss: 0.008
Cross Entropy
epoch
Training Loss
Validation Loss



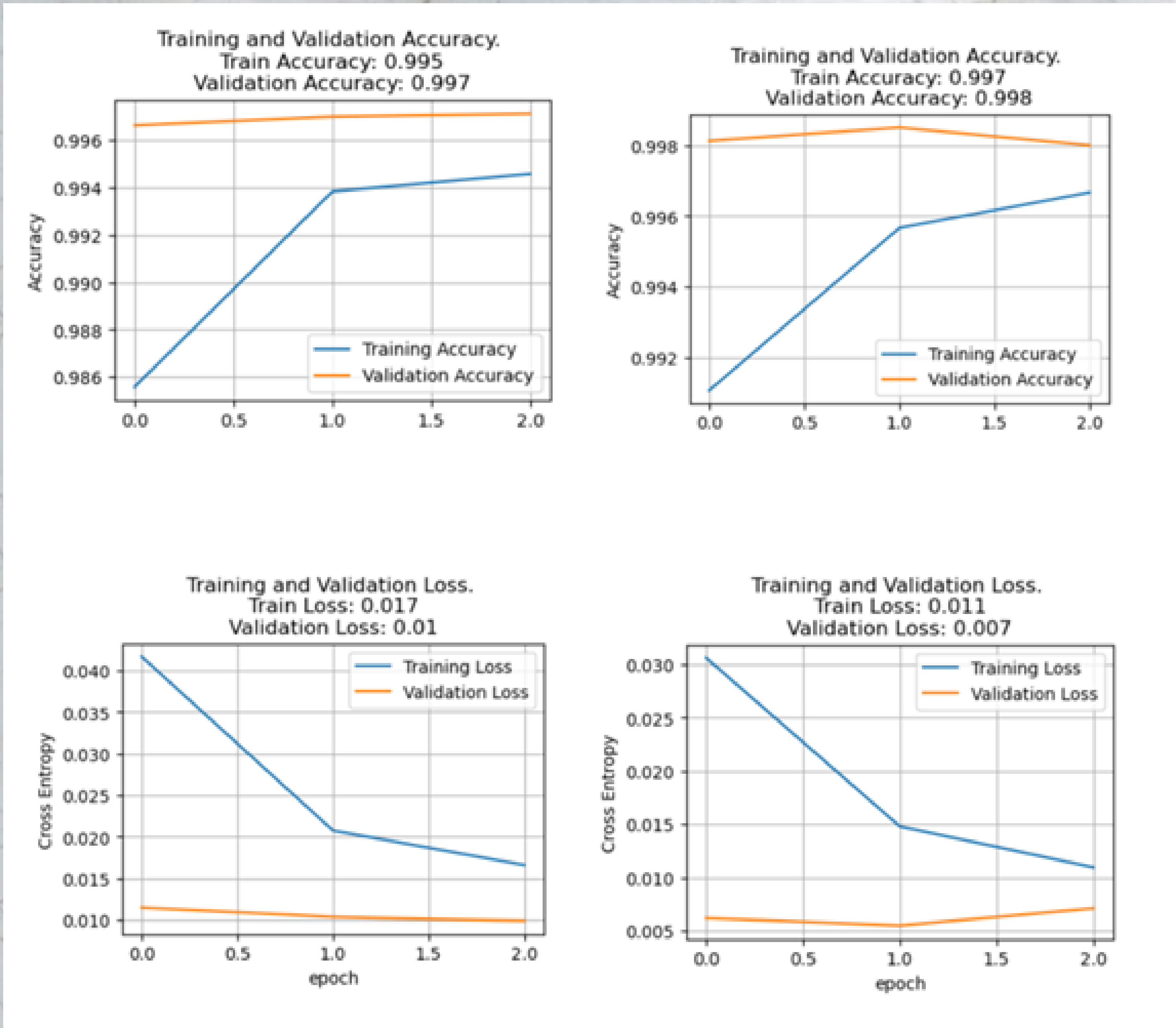
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%

- ***Xception***

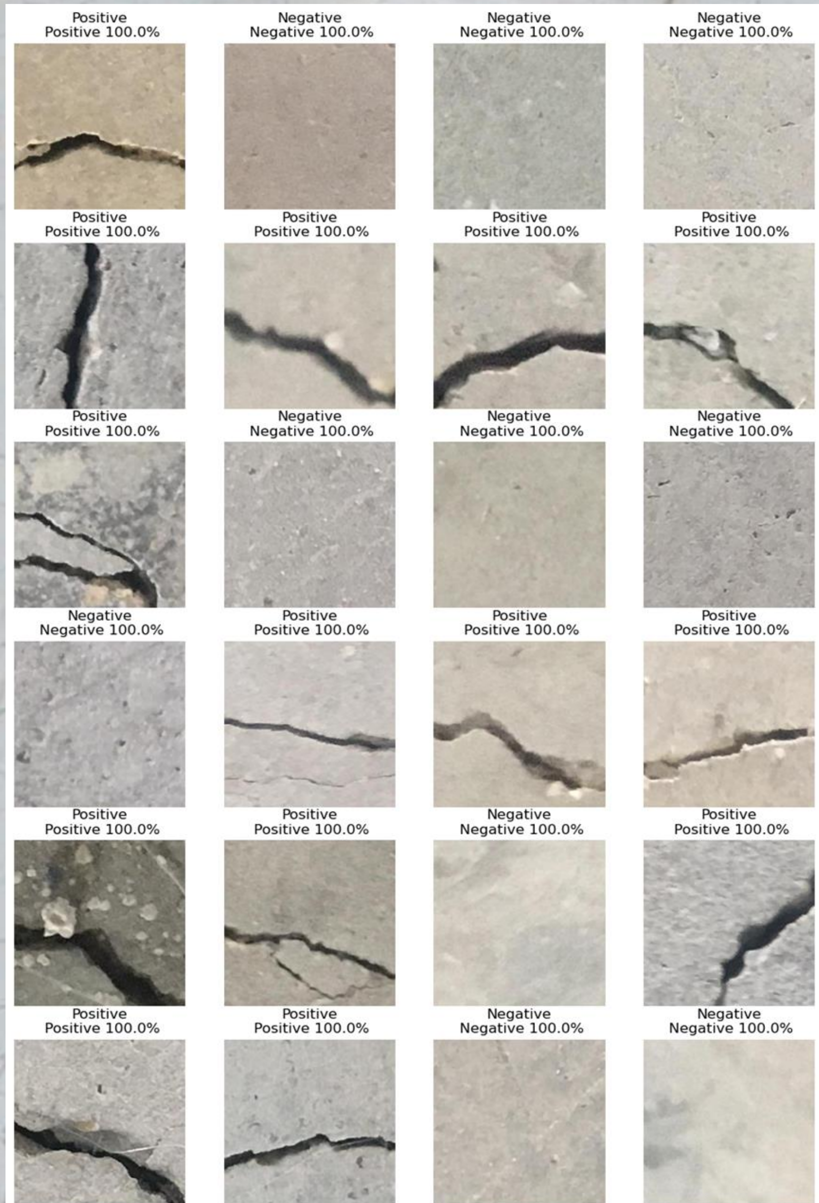
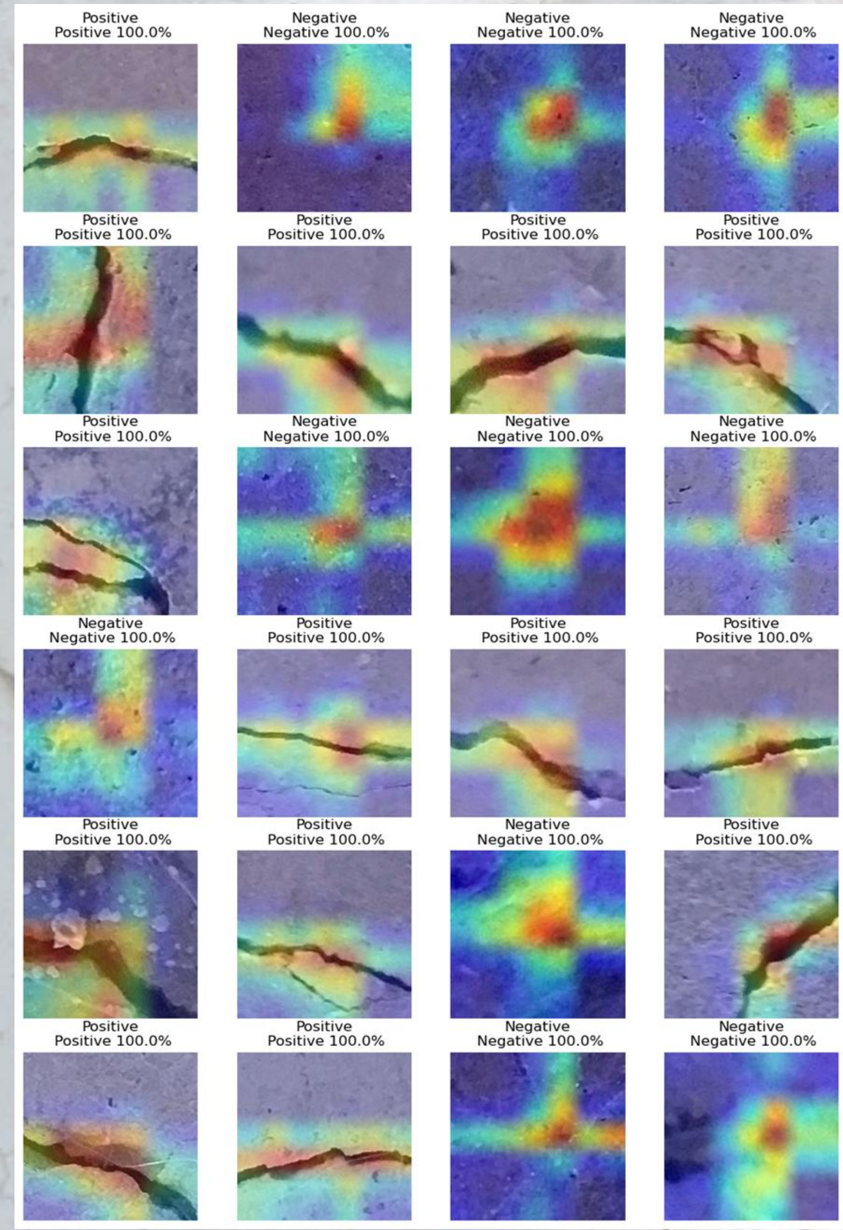
Xception, a ground-breaking convolutional neural network architecture introduced by François Chollet, represents a significant advancement in deep learning, specifically designed to enhance efficiency in image classification tasks. The innovation lies in interpreting Inception modules as an intermediate step between conventional convolutions and depth-wise separable convolutions. In essence, Xception replaces Inception modules with depth wise separable convolutions, demonstrating that this novel architecture, often referred to as "Extreme Inception," outperforms Inception V3 on various datasets. Unlike its predecessor, Xception achieves superior results not due to increased model capacity but rather through a more effective utilization of existing parameters. The core idea behind Xception is to decouple cross-channel and spatial correlations in feature maps, leading to a linear stack of depth-wise separable convolution layers with residual connections for streamlined definition and modification. The architecture's 36 convolutional layers are organized into 14 modules, fostering simplicity in implementation without sacrificing performance. Xception's experimental evaluations on both ImageNet and the expansive JFT dataset reveal marginal improvements over Inception V3 on ImageNet and a substantial 4.3% relative enhancement in multi-label classification on JFT. This performance gain, coupled with comparable model sizes and training speeds, underscores the efficiency and promise of Xception in the realm of deep learning architectures. Moreover, the study explores the impact of residual connections and the necessity of intermediate non-linearities, providing valuable insights into the architecture's design principles.

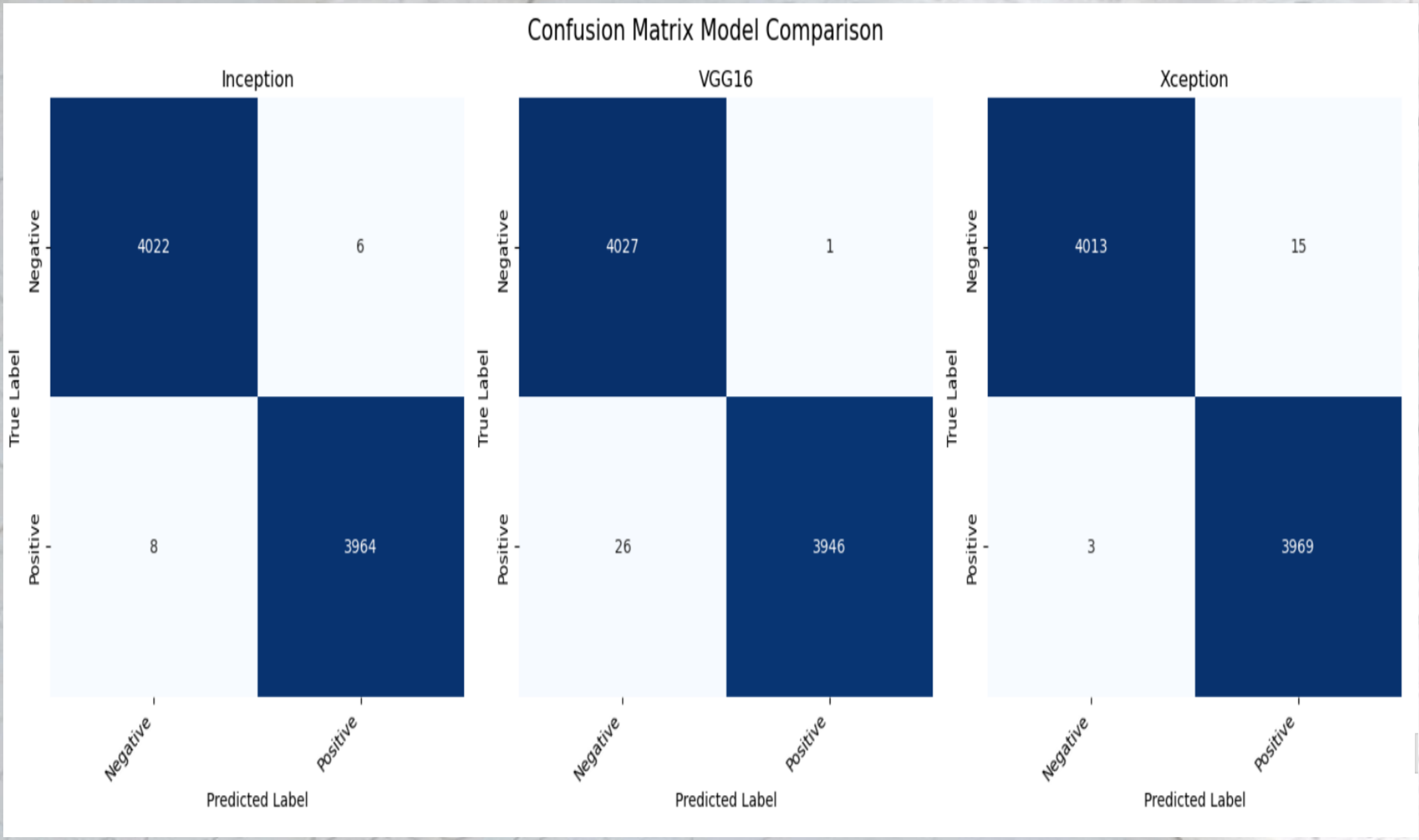
- ***Results***

The meticulously trained Xception-based model demonstrates outstanding performance on the task at hand, achieving an impressive accuracy of 99.84%. The corresponding loss is exceptionally low, reaching 0.007, indicative of the model's ability to make highly accurate predictions with minimal error. Notably, the base layers of the Xception architecture, totalling 132, play a crucial role in feature extraction and representation learning. For further refinement and fine-tuning, a strategic decision was made to unfreeze a subset of the base model layers. Specifically, 33 layers were unfrozen, allowing the model to adapt and specialize in capturing task-specific patterns. This selective unfreezing enhances the model's capacity to learn intricate details from the dataset, contributing to the impressive accuracy observed. The meticulous tuning of these parameters and the utilization of the powerful Xception architecture collectively yield a model that excels in accuracy, making it well-suited for demanding image classification tasks.



Training and Validation Accuracy.
Train Accuracy: 0.995
Validation Accuracy: 0.997
Accuracy
0.996
0.994
0.992
0.990
0.988
0.986
0.0
0.5
1.0
1.5
2.0
Training Accuracy
Validation Accuracy
Training and Validation Accuracy.
Train Accuracy: 0.997
Validation Accuracy: 0.998
Accuracy
0.998
0.996
0.994
0.992
0.0
0.5
1.0
1.5
2.0
Training Accuracy
Validation Accuracy
Training and Validation Loss.
Train Loss: 0.017
Validation Loss: 0.01
Cross Entropy
0.040
0.035
0.030
0.025
0.020
0.015
0.010
0.0
0.5
1.0
1.5
2.0
epoch
Training Loss
Validation Loss
Training and Validation Loss.
Train Loss: 0.011
Validation Loss: 0.007
Cross Entropy
0.030
0.025
0.020
0.015
0.010
0.005
0.0
0.5
1.0
1.5
2.0
epoch
Training Loss
Validation Loss


Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%

Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Positive
Positive 100.0%
Negative
Negative 100.0%
Negative
Negative 100.0%

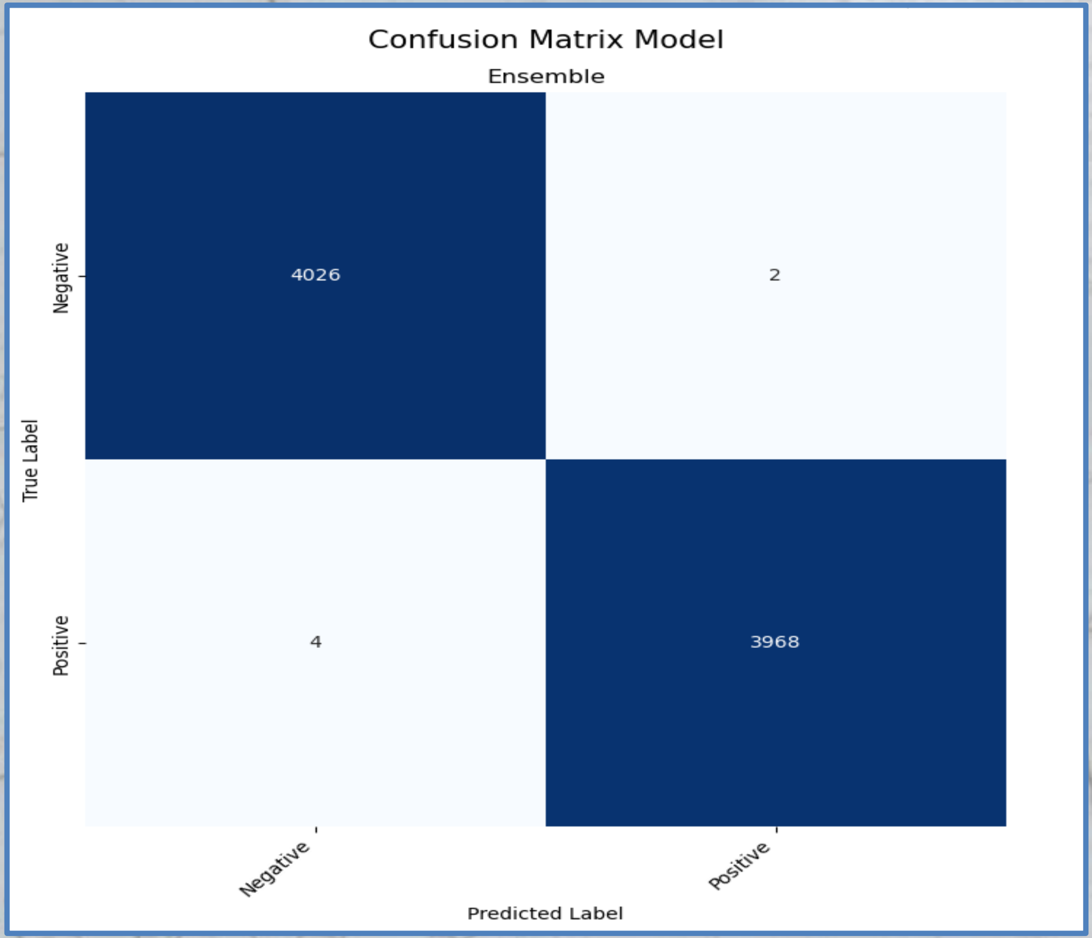


Confusion Matrix Model Comparison
Inception
VGG16
Xception
True Label
Negative
Positive
Predicted Label
4022
6
8
3964
4027
1
26
3946
4013
15
3
3969

Ensemble Model Integration: Ensembling the ResNet v2 , VGG16 & Xception to form one new model

- **The fusion of VGG-16, ResNet-50, and Xception into a unified ensemble model aims to leverage the strengths of each architecture, fostering a comprehensive approach to image classification. By aggregating the outputs of these individual models, the ensemble model strives to enhance predictive accuracy, mitigate potential weaknesses inherent in any single model, and provide a robust framework for diverse feature extraction.**
- **Each model in the ensemble brings a unique perspective to the classification task. VGG-16, with its deep structure, captures intricate details in images, while ResNet-50's innovative residual connections aid in addressing gradient-related challenges. Xception, with its depthwise separable convolutions, excels in efficient feature extraction.**
- **The ensemble model amalgamates predictions from VGG-16, ResNet-50, and Xception, harnessing their collective intelligence to produce a final classification output. Leveraging this ensemble approach, the model endeavours to provide a more nuanced and accurate understanding of the input data, potentially outperforming individual models and offering a robust solution to complex image classification scenarios.**

Confusion Matrix for the Ensemble model



MODEL / METRICS	Resnet v2	VGG16	Xception	Ensemble
Accuracy	0.99825	0.996625	0.99775	0.99925
Precision	0.9982501184441626	0.9966442551382927	0.9977545270724332	0.9992501221943734
Recall	0.99825	0.996625	0.99775	0.99925
F1-Score	0.9982499968279582	0.9966248932019421	0.9977500185630616	0.9992499986405535

Real life application of the model in NITK

***Images are taken at Old Gym, Satpura Hostel Fountain base, Our Hostel room in MT2 respectively**

Conclusion

In this on crack detection in structural elements, we explored a diverse range of methodologies, technologies, and advanced deep learning models, including ResNet v2, VGG-16, and Xception. The literature review revealed a rich landscape of crack detection techniques, emphasizing the importance of integration, efficiency, and adaptability to various applications. As technology evolves, there's a clear trend towards automation and the integration of machine learning techniques, showcasing a promising future for robust and accurate crack detection systems.

The deep learning-based approach focused on leveraging ResNet v2, VGG-16, and Xception for crack detection. Each model brought unique strengths to the table, with ResNet v2 excelling in handling extremely deep networks, VGG-16 showcasing efficiency with its straightforward architecture, and Xception introducing depthwise separable convolutions for enhanced resource efficiency.

The ensemble model, integrating predictions from ResNet v2, VGG-16, and Xception, emerged as a robust solution for crack detection. The fusion of these models enhanced predictive accuracy and addressed potential limitations associated with individual architectures. The numerical insights from the confusion matrix for the ensemble model provided a detailed breakdown of its performance across various classes, showcasing its ability to provide nuanced and accurate predictions.

Beyond accuracy, the project emphasized the significance of computational efficiency for real-world deployment. The models were evaluated not only on their predictive capabilities but also on their inference time and resource utilization, ensuring practical feasibility in diverse scenarios.

In conclusion, this project represents a significant advancement in the field of crack detection in structural elements. The integration of state-of-the-art deep learning models in an ensemble framework offers a holistic and efficient solution, contributing to proactive maintenance and ensuring the longevity of critical infrastructure. The insights gained from this project have implications not only for crack detection but also for the broader domain of image-based analysis in infrastructure safety and maintenance. As technology continues to progress, the methodologies and models presented in this project pave the way for future innovations in the pursuit of safer and more resilient structures.

THANK YOU