School of Collective Intelligence

Mohammed VI Polytechnic University

Thesis of the end-of-studies project

To obtain a Master's degree of Collective Intelligence

Deploy Services of Network Infrastructure: AI/ML Digital Deploy

Authored by :

Ms. Chaymae Bennouri

Pr.Mark Klein - SCI Supervisor

Mr. Oussama Ghandari - Nokia Supervisors

Class of: 2023/2024

Dedicated

This thesis is dedicated to my mother, Aicha, and my siblings, Yassine, Hafsa, Hajae, and Oussama. Who was always there for me, even on the tough days.

This thesis is dedicated to the memory of my father Ibrahim, I miss him every day, but I am proud to be his daughter. The principles he instilled in me shape my path and inspire me in every challenge I face and every success I achieve.

- Chaymae.

Acknowledgments

In these important moments, I think of my family, starting with my mother as I will forever be grateful for her support, love, and the values she has installed in me, and my amazing siblings for always being there for me, as well as the rest of my family and friends.

To all the professors who have made considerable efforts to import their knowledge under the best possible conditions at the School of Collective Intelligence, thank you for your dedication to our education.

Professor Mark Klein and **Mr. Moad Hicham SAFHI** assisted me with their support, insights, guidance, suggestions, and patience, I wish to express my sincere appreciation and thanks to you. I also want to thank the jury members for evaluating my work.

Mr. Oussama Ghandari, my supervisor at Nokia. His priceless knowledge has given me much knowledge and learning possibilities in addition to his support and assistance have helped me to achieve my goals.

Mrs. LASRY Lamia, **Mrs. Selma EL Fassi**, and **Mr. Lazzouzi Abdellatif** the NI team at Nokia, were always available, and attentive, and their valuable advice was immeasurable for all the interns.

Lastly, my dearest friend, **Ghizlane Rehioui**, who was always there for me along the way when I needed her. Thank you for your help, support, and love.

Abstract

This thesis investigates the enhancement of image classification models within the framework of AI and ML digital deployments, focusing on a case study conducted at Nokia Morocco. The research addresses critical challenges in industrial image classification, such as overfitting, data quality, class imbalance, and the need for scalable, high-accuracy solutions.

The study evaluates the performance of several advanced convolutional neural network (CNN) architectures, including NASNetLarge, InceptionV3, DenseNet169, and EfficientNetB7, alongside ensemble learning techniques. The results demonstrate that while individual models like DenseNet169 achieved moderate success with a maximum accuracy of 62.37 percent, ensemble learning significantly improved overall performance, resulting in an accuracy of 69 percent.

The findings emphasize the importance of data augmentation and hyperparameter tuning in reducing misclassification rates and improving model generalization. Despite the advances, the study highlights the ongoing challenge of balancing model accuracy with computational efficiency, particularly in complex industrial environments. In conclusion, the study suggests a new direction for future research that will focus on fixing class imbalances in industrial datasets. This will make AI-driven image classification systems more reliable and useful.

Key Words : Image classification, AI, ML, overfitting, CNN, ensemble learning, model precision, misclassification, feature extraction, accuracy.

Contents

List of Figures

Chapter 1

General overview

As part of my master's training at the School of Collective Intelligence, I interned at Nokia Morocco. During this internship, I had the opportunity to participate in various missions that allowed me to learn the job of AI/ML in the field of digital deployments and to grow professionally and personally.

AI/ML Digital deployments involve the practical use and integration of artificial intelligence and machine learning technologies in digital settings. These deployments are complex, involving stages such as data collection, model training, testing, and real-world application. This thesis explores the ways of improving image classification models and uses them effectively in these types of settings. It emphasizes the importance of having flexible, scalable solutions that can handle data and achieve high accuracy.

In the field of artificial intelligence and machine learning, image recognition is a critical application that uses the ability of deep learning to recognize and classify objects within images with high accuracy. Image recognition, a key component of computer vision, has evolved into a critical technology that supports a wide range of applications, including autonomous vehicles, facial recognition systems, medical diagnostics, and industrial automation. This thesis examines the advances in image recognition technology, particularly in the context of implementing these methods in Nokia Morocco's operations.

This research not only improves the understanding of how image recognition within AI/ML digital deployments works but also underscores the strategic importance of this technology in driving innovation and efficiency in industrial applications. The purpose of this study is to provide valuable knowledge that can be used in the broader field of AI/ML. It also provides practical recommendations and advanced methods that can be implemented in many industries where image recognition is critical.

In addition to this general overview, the thesis consists of four chapters:

- The general scope of this thesis that has been discovered in Chapter 1.

- The Project context will be described in detail in Chapter 2. which also identifies the issues and emphasizes the potential solutions.

- The literature review on image classification is the focus of Chapter 3, which also examines previous works in this field and advanced technologies in image classification.

- The methodology is the primary focus of Chapter 4, which outlines the steps and processes that were employed to resolve the issue.

- The results of the applied methodologies are presented in Chapter 5, which is followed by a discussion that evaluates the work conducted within Nokia.

The thesis concludes with a comprehensive conclusion that summarizes the work that has been completed, emphasizing the proposals and perspectives.

Chapter 2

Project context

Introduction

This chapter provides an overview of the project's historical context, main goals, and essential elements. It sets the context by presenting the general structure of the project. The chapter ends by providing a concise overview of the project's objectives, essential activities, and the incorporation of a Gantt chart to aid in project planning and administration.

2.1 Project framework

This project is an essential part of Mohammed VI Polytechnic University's School of Collective Intelligence final studies internship. The internship attempts to bridge the gap between academic knowledge and practical application by allowing students to work on industry-specific projects. The project at Nokia Morocco aims to use AI/ML technologies to enhance image classification, which is critical to the company's deployment operations.

During this internship, we began working toward certification as a Nokia Certified in AI/ML. This certification, along with others, served as a first step toward understanding digital deployment and its importance in our work. My work with the NI Deployment Process and Tools Manager team was centered on developing a model accuracy and providing updates on the trends to improve quality.

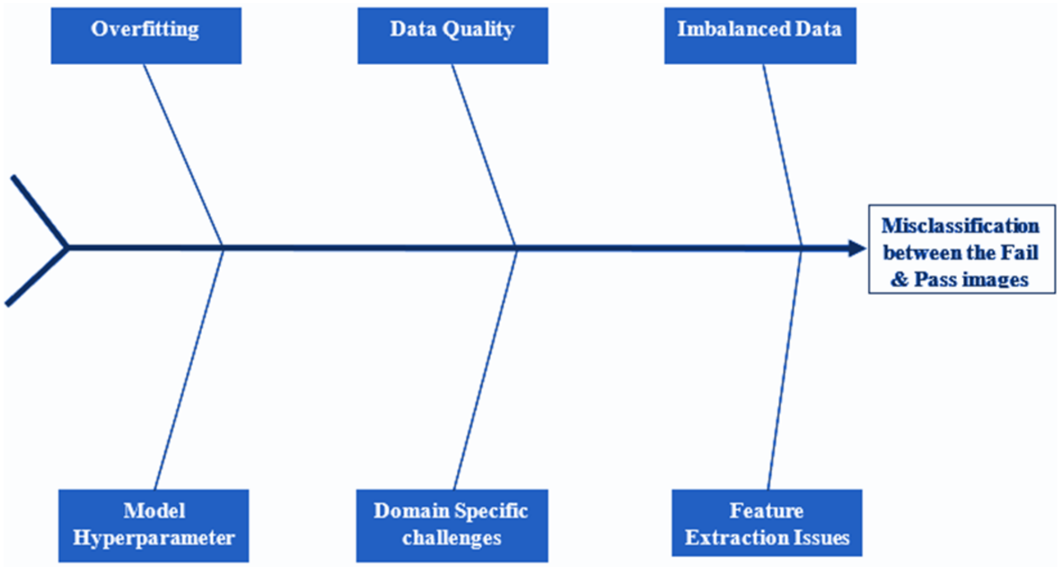

Figure 2.1: Ishikawa diagram.

2.2 Project objectives

2.2.1 Context of the project

The NI Deployment Process and Tools Manager is critical to the successful deployment of Nokia products. It creates and manages processes, tools, and strategies to provide effective and high-quality deployment services. In this regard, incorporating AI/ML for image classification is a critical step toward improving deployment capabilities. The project's goal is to use AI/ML technologies to improve the accuracy and confusion matrix efficiency of image classifications while deployed. Overall, AI/ML will be integrated to enhance the quality, efficiency, and dependability of Nokia's deployment operations, leading to increased customer satisfaction.

2.2.2 Problem statement

Although the company Nokia uses a powerful model (ResNet50) with a high accuracy level to classify images, this model suffers from classifying unseen images, and it predicts 'Fail' images as 'Pass' and predicts 'Pass' images as 'Fail'. This error originates from several issues such as overfitting, data quality, imbalanced data, ...etc (Figure 2.1).

This thesis will address the following research questions:

1. How can advanced deep learning techniques and ensemble learning be optimized to improve the

accuracy and robustness of image classification models in industrial AI/ML deployments?

2. What role does ensemble learning play in enhancing the predictive performance and reducing the error rates of image classification models in industrial environments?

2.2.3 Tasks to accomplish

- **Improve confusion Matrices.**
- **Collaborate with cross-functional teams**.
- **Evaluate and Test Model Performance.**
- **Stay current on trends and breakthroughs**.
- **Manage deployment and performance.**
- **Report and document progress.**

This section provides a detailed summary of the critical tasks required for the project's success, focusing on improving the precision and model accuracy of image categorization systems throughout the organization.

2.2.4 Gantt Chart

The period of this project is from the 15th of March to the 15th of September. Figure 2.2 stands for the Gantt chart of this project.

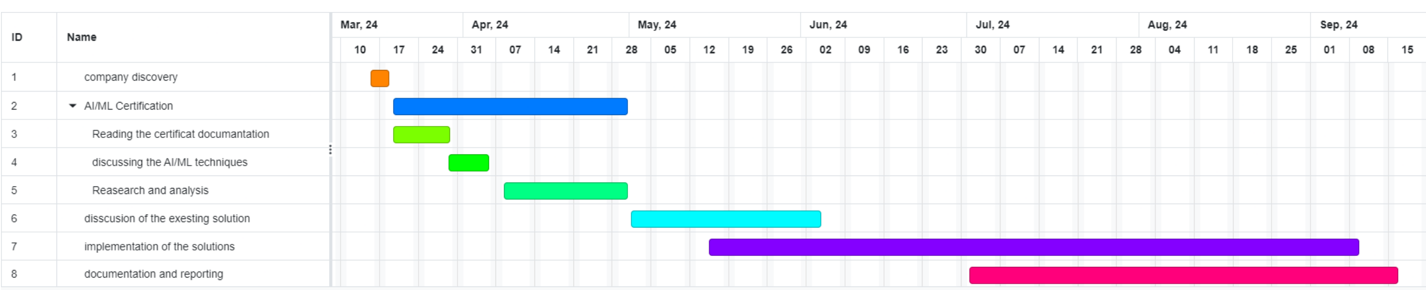

Figure 2.2: Gantt chat

2.3 Conclusion

This chapter presented the project titled "Deploy Services of Network Infrastructure: AI/ML Digital Deploy." We discussed the project's context, which encompassed the project structure. In addition, we discussed the project goals, placing particular emphasis on the context and problem statements that need attention. In addition, we delineated the tasks that need to be accomplished. Lastly, we deliberated on the Gantt chart, which will offer a graphical depiction of the project's timetable and the interdependence of tasks. In the upcoming chapter, we will provide a comprehensive assessment of the existing literature and previous research that has been undertaken to address the problem.

Chapter 3

Literature Review

Introduction

This chapter discusses a comprehensive review of the existing research in image classification, focusing on previous works and recent advancements. This section will provide key concepts and technologies that have marked the AI/ML world; those techniques addressed will be used in later chapters. This section offers the basis for comprehending the approaches and implementations that will be utilized to enhance image classification models in this thesis.

3.1 Previous work

Image classification is a fundamental component of computer vision but has historically faced significant difficulties due to the complex features embedded in visual inputs. Traditional image classification methods, which rely primarily on human-generated features, often face limitations when used on diverse datasets. These algorithms struggled, especially when dealing with data with a large number of dimensions. They were often unable to capture the nuances required for proper classification.

Deep learning has changed the image classification process by allowing models to construct hierarchical representations from raw data. This paradigm change has successfully surmounted various constraints of old approaches, resulting in major increases in precision and durability. (Bhavya et al., 2023) [8] note that the use of deep learning approaches, notably Convolutional Neural Networks (CNNs), has proved critical in overcoming the challenges posed by complex visual patterns and enormous datasets.

CNNs have been the favored approach for image classification because they possess the capability to automatically extract and acquire significant information from images without requiring operator involvement (Vocaturo, 2021). [10] CNNs are highly suitable for jobs that involve complicated and high-dimensional visual input due to their capacity to adapt and improve features across several layers of abstraction (Lorente et al., 2021) [13].

In addition, the use of data augmentation methods, as investigated by (Zhang et al.,2022) [12], has bolstered the resilience of deep learning models by producing varied training data that reduces the likelihood of overfitting. By using edge enhancement techniques as described by (Bu et al., 2024) [1], image classification systems have experienced substantial enhancements in both accuracy and computational efficiency, resulting in increased reliability for real-world applications.

3.2 Advanced CNN architecture

CNNs have formed the foundation of modern image classification systems due to their capacity to capture spatial hierarchies of data. Due to the emergence of many CNN designs, the task of image classification has become more achievable, as each structure provides distinct benefits.

InceptionV3, a well-known CNN architecture, is renowned for its pioneering use of factorized convolutions. This architectural form is extremely economical in situations when computational resources are limited, making it a popular choice for both academic research and industrial applications. (Setyo et al., 2023) [5] emphasizes the effectiveness of InceptionV3, particularly in achieving a good balance between performance and resource utilization. When developing models in resource-constrained environments, it is critical to strike a compromise between performance and resource utilization. This demonstrates the usefulness of InceptionV3 for a variety of picture classification problems.

EfficientNetB7 is an advanced CNN model that presents a novel scaling strategy. EfficientNetB7 differs from traditional models in that it scales the network's depth, width, and resolution proportionally. This strategy produces a model that is both balanced and efficient. (Farizi et al., 2024) [3] emphasize the model's resilience and capacity to handle large and complex image collections. This feature is crucial in contemporary image classification problems, since datasets are not only extensive but also intricate, frequently posing difficulties that other CNN designs find difficult to surmount.

DenseNet topologies, such as DenseNet169 and DenseNet201, use dense connectivity to effectively address issues such as fading gradients and improve information flow throughout the network. The extensive interconnection allows for the most efficient use of characteristics, which is especially useful in complex neural networks. Ritter et al., 2023) [7] examine the advantages of DenseNet architectures, specifically in producing more accurate and dependable classification results, particularly for jobs that necessitate complex layers of features.

NASNETLarge is a leading CNN design that uses neural architecture search (NAS) to automatically determine the most appropriate network topology for a specific task. This methodology has resulted in the creation of exceptionally efficient and precise models that outperform a wide range of image classification benchmarks. (Dewan et al., 2023) [2] examine the significance of NASNETLarge in advancing the limits of CNN performance, specifically in attaining state-of-the-art precision while reducing the computational resources needed.

3.3 Feature extraction

Feature extraction is an important part of image classification since the quality of the extracted features directly determines the model's performance. Traditionally, feature extraction relied significantly on manual human engineering, which frequently resulted in features that did not capture all of the data's fine characteristics. Nevertheless, with the development of deep learning, the procedures for extracting features have become increasingly automated and effective.

(Keying et al., 2024) [11] emphasize the enhanced capability of the modified YOLOv5s algorithm to extract high-quality information from photographs while simultaneously minimizing computational effort. This development is especially important for tasks that demand quick responses, such as surveillance and autonomous driving. YOLOv5's ability to identify significant features quickly and reliably allows these systems to operate effectively in real time.

Furthermore, the incorporation of deep learning models into feature extraction methods has permitted the creation of more powerful algorithms capable of efficiently processing large datasets. (Jian et al., 2023) [4] discusses how deep learning-based feature extraction has improved the precision of optical remote sensing in ship classification and detection, especially in difficult settings where standard methods fail. The paper shows that automated feature extraction improves accuracy while also reducing data processing time, making it more suitable for real-time applications.

3.4 Ensemble learning

Ensemble learning is the process of merging many models or classifiers to enhance the overall performance and accuracy of a prediction or classification job. While specific CNN architectures like InceptionV3, EfficientNetB7, DenseNet169, DenseNet201, and NASNETLarge have their unique benefits, the classification performance can be improved even more by employing ensemble learning approaches that merge different models.

Ensemble learning combines the advantages of different models to generate a final prediction that is more robust and accurate. Ensemble approaches can enhance model generalization and minimize variance by combining predictions from several models. This makes them highly effective in complex classification tasks. Rane et al. (2023) [6] highlight the potential of combining ensemble deep learning and machine learning approaches to develop strong models that surpass individual models, especially in situations involving high-dimensional data. Their research showcases the efficacy of ensemble approaches in enhancing the precision and dependability of classifications, especially in settings where data intricacy presents substantial obstacles.

In addition, Truong and Tran (2023) [9] investigate ensemble learning methods that are specifically designed for classifying high-dimensional data. They emphasize that in such situations, individual models frequently have difficulties in comprehending the complete intricacy of the data. Ensemble approaches, on the other hand, can combine different viewpoints from numerous models, which leads to an improvement in the overall classification performance. This method not only enhances accuracy but also offers a more thorough comprehension of the data, which is essential in domains where precision is important.

This method is especially useful in situations where no single model can perform optimally in all elements of the classification challenge. (Rane et al., 2024) [6] demonstrate how combining predictions from multiple models can overcome the constraints of individual models, resulting in improved overall performance. This capacity makes ensemble learning an effective tool for dealing with complex classification tasks, where the variety and complexity of the data necessitate a more detailed approach.

Conclusion

The selection of YOLOv5, InceptionV3, EfficientNetB7, DenseNet169, DenseNet201, NASNETLarge, and ensemble learning techniques is motivated by their capacity to overcome the constraints of existing image classification algorithms. These advanced methods, supported by a large range of academic literature, provide a solid foundation for obtaining high accuracy, efficiency, and robustness in image classification, which is consistent with the goals of this study, and this is what we will discovering in the methodology section.

Chapter 4

Methodology

Introduction

This chapter provides a comprehensive explanation of the approach used in the project to accomplish precise and effective image classification. The approach consists of several stages, namely data cleansing, data augmentation, feature extraction, model training, and evaluation. Two strategies are suggested to improve the accuracy, each utilizing distinct methodologies and workflows as seen in Figure 4.1.

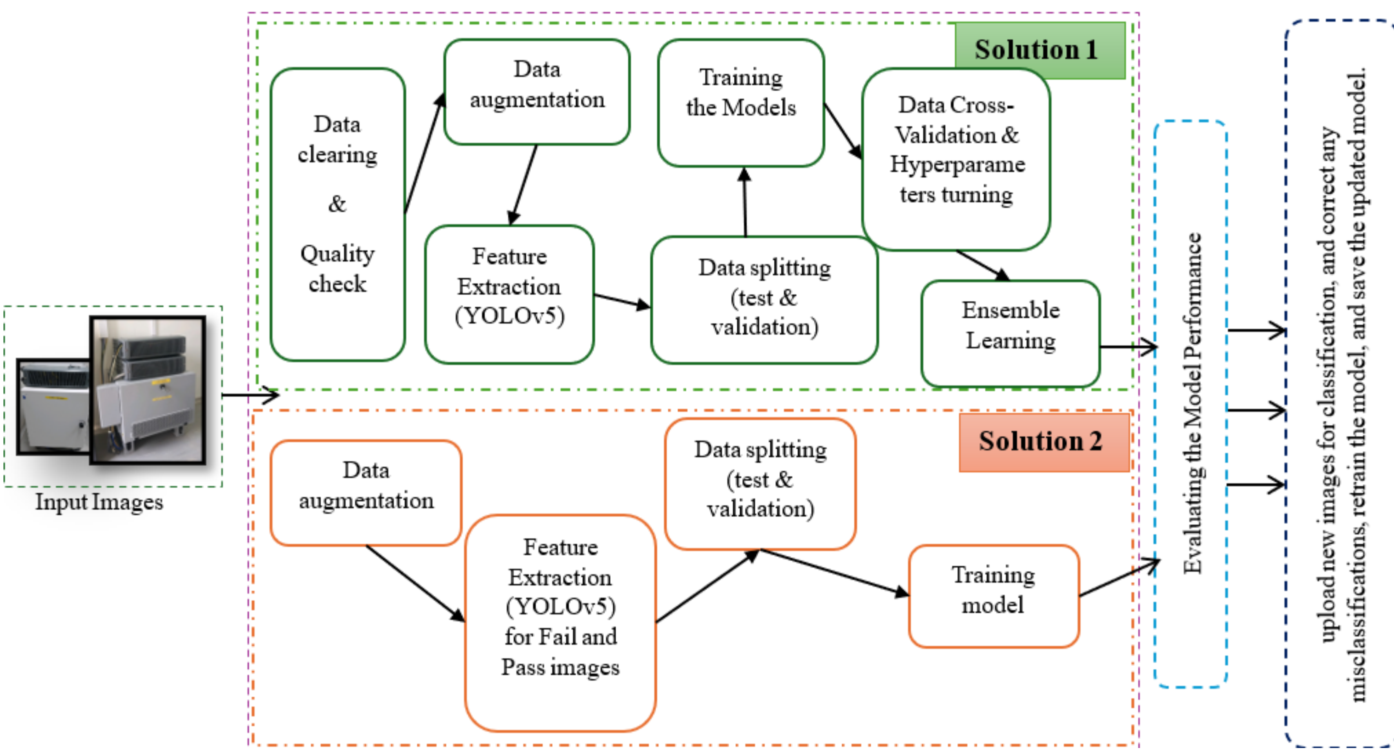

Figure 4.1: Workflow Diagram for Image Classification Methodology.

4.1 Dataset

The dataset employed in this study is organized in a way that enables precise image classification for quality assessments in different domains. The pictures are categorized into two folders according to various domains and their respective labels, ”Pass” and ”Fail. Two distinct approaches are proposed, each of which trains and validates using a certain dataset format.

Note: Please be aware that because of company proprietary rights, the pictures used in this project will be displayed in a blurred state.

4.1.1 Solution 1: Three-Domain Structure

The dataset is divided into three key domains in the first solution:

- Global Quality Checkpoints, 1407 pictures (Pass: 740, Fail: 667).
- Local Quality Checkpoints, (Customer Specific) 342 pictures (Pass: 172, Fail: 170).
- NI Checkpoints, 1295 pictures (Pass: 706, Fail: 589).

Each domain has ”Pass” and ”Fail” folders, as seen in Figure 4.2. This architecture ensures that each area’s data contains a diverse set of relevant scenarios and configurations. The goal is to develop models that have high cross-domain generalization capabilities, increasing the classification system’s resilience and dependability.

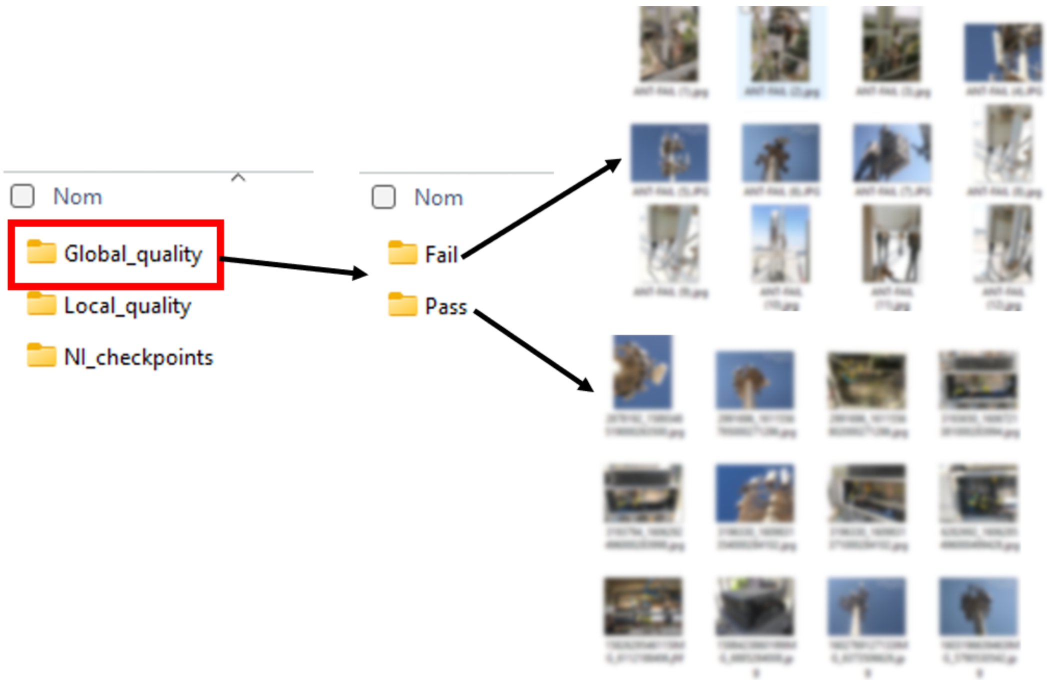

Figure 4.2: Dataset of the solution 1

4.1.2 Solution 2: Detailed Folder Structure

The second solution uses an additional comprehensive dataset that is organized into 27 distinct folders. Each folder contains photographs that are classified as either ”Pass” or ”Fail” (Figure 4.3). These folders serve as checkpoints and criteria for installation, guaranteeing thorough coverage of setups and quality requirements.

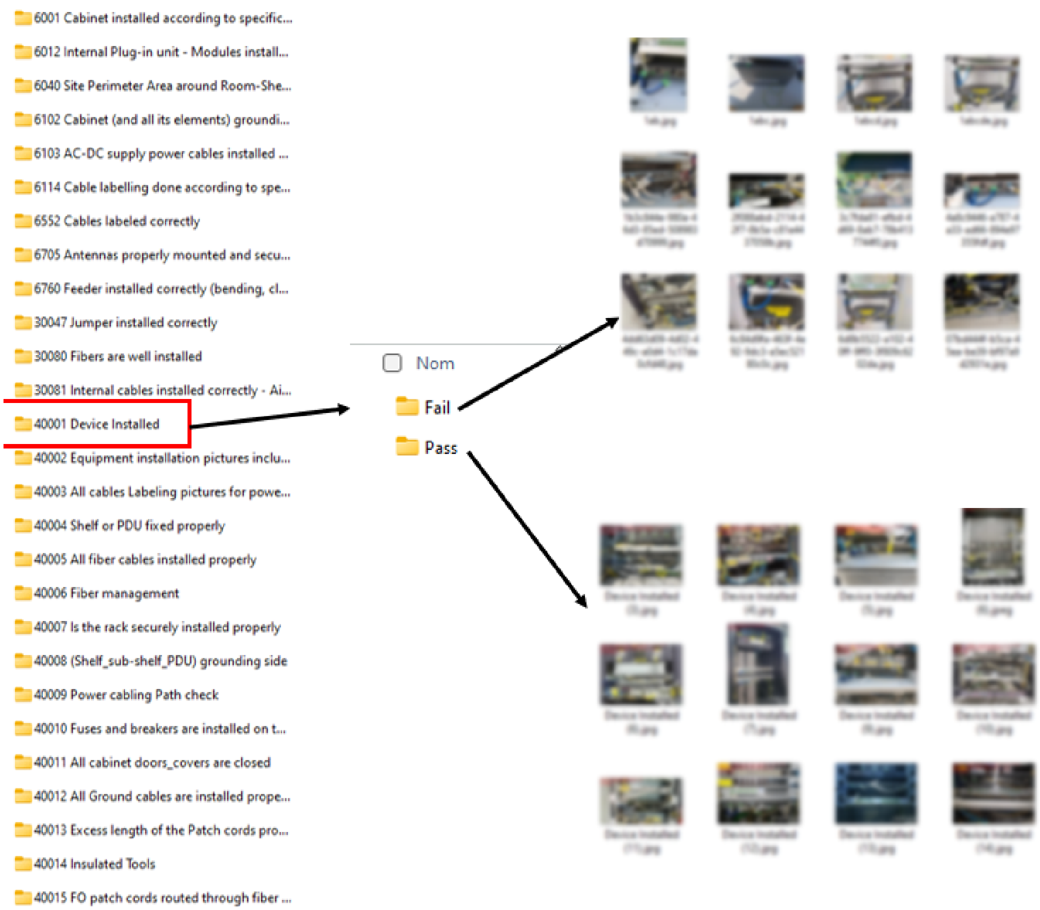

Figure 4.3: Dataset of the solution2

This thorough categorization allows for the focused training of models, highlighting certain aspects of the setup and installation process. Using this template allows the model to identify quality concerns with more detail and accuracy.

Instead of focusing on the domains themselves, the classification task for both solutions makes use of the "Pass" and "Fail" images. This method guarantees that the models are trained to identify quality indicators from the photos themselves, independent of the particular domain to which they belong.

4.1.3 Categorized Data Table

The dataset is labeled in Figure 4.4 below, which provides domains with their codes and checkpoint descriptions. The labeled data table provides an organized and easy-to-understand summary of the image domain classification, making it easier to organize and understand the data for both solutions. By maintaining similar images across both solutions, one can accurately evaluate the effectiveness of different classification algorithms and ensure consistent results.

Categorized Data Table

Domain	Code	Description
Crew Checkin-Checkout	8002,8001	Crew Auth,Crew check
Domaine A	18,19016,19015,19012,19005,125,19020,1	from Ground,Complex checkpoint associate
Domaine B	,214,208,210,201,272,268,263,239B,211C	ver - Hybrid - Hybrid cable grounding at Lo
Domaine E	51537,5005	ector mechanical Tilt (spirit level),Safety sig
H&S checkpoints	9016,9114,9033,9018,9006,9118,9104,90	rest Lanyard - Back,Barricade Taping to cre
NI checkpoints	05,40012,40011,40015,40007,40006,4001	bending radius (closed angles are not allow
Other	,D_9005,D_9004,D_9003,D_9002,D_9001	g lanyard) replicate - for MWC Demo,Check
TI-Non Standard AI Codes	0,024,023,021,009A,007,004,009B,001A,0	bel showing Site ID,Tower - Antenna Mount
global quality checkpoints	6648,6103,6758,6651,6127,6614,6113,60	en door,OVP boxes (both ends) for remote
local quality checkpoints (customer specific)	,30369,30160,30035,31767,31490,31389,	s,Generator main switch, if not part of the r

Figure 4.4: Categorized data table

4.2 Data Preprocessing

The preparation of data is essential for ensuring the dataset's quality. This section describes the procedures for cleaning and validating the dataset before using it for training.

4.2.1 Data Clearing and Quality Check

The data cleaning and quality check process involves scanning through the dataset to identify and manage any corrupted or irrelevant files. This is essential for training dependable and precise machine learning models. This automated procedure enhances efficiency and mitigates the potential for errors that may result from humans.

4.2.2 Data Augmentation

The augmentation process involves applying random transformations to the images and creating new, varied versions of the original images.

These adjustments include :

- scaling and normalizing pixel values to a range of 0-1.
- Rotation: Randomly rotates photos by 40 degrees.
- Randomly shift photos horizontally and vertically up to 30 percent.

- Shear: Performs random shearing alterations within a 30 percent range.
- Zoom: Change the zoom level by up to 30 percent at random.
- Randomly flip photos horizontally.
- Fill Mode: Replaces freshly introduced pixels with their nearest values.
- Validation Split: Set aside 20 percent of data for validation reasons.

4.2.3 Visualizing the Preprocessed Data

The dataset used in this project covers three primary domains in Solution 1:

- Global Quality Checkpoints: Contains 1,398 images across "Pass" and "Fail" categories.
- Local Quality Checkpoints (Customer Specific): Contains 335 photos in the "Pass" and "Fail" categories.
- NI Checkpoints: Includes 1,401 photos in the "Pass" and "Fail" categories.

The figure below(Figure 4.5) shows various preprocessed photos classified as "Pass" or "Fail". This image aids in understanding how the data is structured and the diversity that remains in the dataset following the preparation stages.

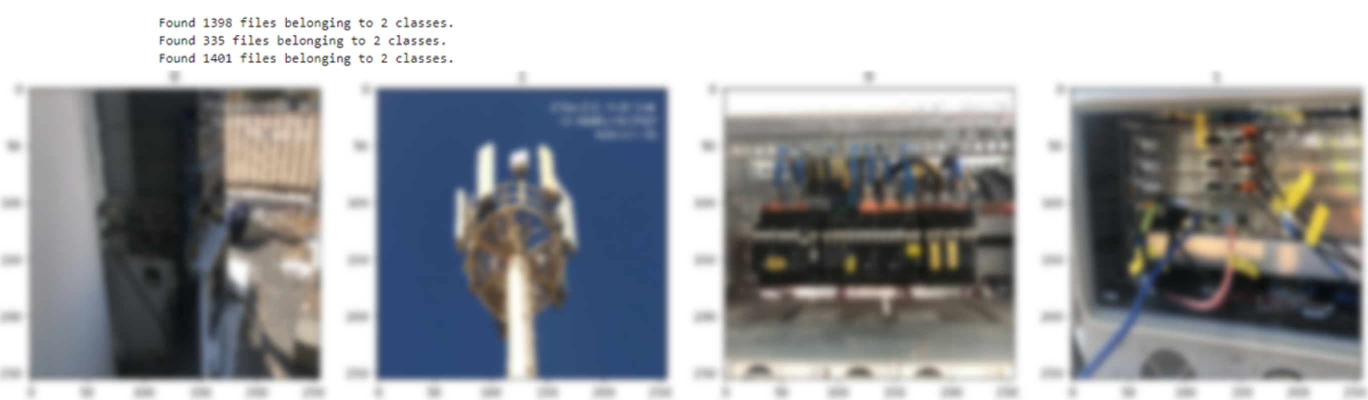

Figure 4.5: Figure 6: Examples of Preprocessed Images with Class Labels "0" (Fail) and "1" (Pass).

4.3 Feature Extraction

The YOLOv5 model is utilized for extracting features to improve image classification tasks in quality control. The YOLOv5 model, which has been trained in advance, is utilized using the torch.hub library to analyze photos and identify objects. The model then converts the collected bounding box data into feature vectors. If no objects are identified, a vector with all elements set to zero is returned. This feature extraction method is used on both the training and validation datasets, and the photos are preprocessed to suit the required format.

The recovered features are then utilized to train classifiers, which take advantage of YOLOv5's strong object identification capabilities to increase classification models' precision and efficacy in discriminating between "Pass" and "Fail" pictures.

4.4 Data splitting

Data splitting is the step before going to the training and validation. The dataset needs to be divided into two subsets in order to train the model and assess its performance. This stage guarantees that the model is assessed on data that it has not been exposed to during training to provide a more precise evaluation of the model's ability to generalize. The data utilized in this study is split into training and validation sets, with a designated amount reserved for validation to monitor the model's performance during training. By employing this methodology, the issue of overfitting can be identified, hence guaranteeing the model's capacity to generalize to novel, unobserved data.

4.5 Data Cross-Validation and Hyperparameters turning

Two essential techniques for improving the efficiency of machine learning models are used; cross-validation and hyperparameter tinkering. By creating diverse groups from the training set, cross-validation enables the model to be trained on some and validated on others. A more accurate way to measure the model's performance is to take the average of the findings. In contrast, hyperparameter tuning adjusts the model's parameters to determine the optimal combination of learning rate, batch size, and number of layers for maximum performance. The best configurations to maximize the accuracy and resilience of the model can be found by carefully experimenting with various setups.

4.6 Model Training and Validation

4.6.1 Training Process

The machine learning models are provided with preprocessed and enhanced data during the training phase. NASNetLarge, InceptionV3, EfficientNetB7, DenseNet169 (Solution 1), and DenseNet201(Solution 2) are examples of models that have undergone training to accurately identify and classify images. To have good performance, hyperparameter adjustment, and cross-validation are used during the model's training process. The results of each model will be detailed in the next chapter.

4.6.2 Ensemble Learning

The process of evaluating ensemble models involves combining predictions from different models in the ensemble and then averaging these predictions to generate a final output (bagging). The method begins by processing sequential batches of validation data. For each batch, each model in the ensemble generates predictions. The predictions across models are then combined to generate a conclusive diagnosis for each sample. The values are compared to the actual labels to calculate several performance metrics, including confusion matrix, classification report, F1 score, precision, and recall. These metrics provide insights into how well the ensemble model is performing at distinguishing between "pass" and "fail" classes.

4.6.3 Validation and Evaluation

A distinct subset of the dataset is used for data validation, which is retained during the data augmentation procedure. The effectiveness of the models is evaluated using various metrics, including:

Accuracy

Accuracy is the total number of precise predictions acquired overall is defined as accuracy, which is expressed in Equation (1).

$$Accuracy = (TP + TN)/(TP + FP + TN + FN)$$

where TP and TN are the true positive and true negative; FP and FN are the false positive and false negative.

Precision

Precision is the ratio between the true positives and the sum of true positives and false positives and is expressed in Equation (2).

$$Precision = TP/TP + FP$$

Recall

Recall is the ratio between the true positives and true positives and false negatives, which is shown in Equation (3).

$$Recall = TP/TP + FN.$$

F1-score

F1-score combines measures of precision and recall, based on the rates of true positives, false positives, and false negatives.

$$F1 - score = 2 * Precision * Recall/Precision + Recall$$

4.7 Evaluating the Model Performance

Evaluating model performance is an ongoing process that involves retraining models, finding misclassifications, evaluating models on unseen images, and updating models as necessary (Figure 4.6). This iterative process ensures that the model is continually improved and modified according to new data and needs. Models are retrained using updated data to address and examine any misclassifications that are found. This ensures that even as new data is released, the models will always be accurate and reliable.

Figure 4.6: Example of model evaluation

Conclusion

A comprehensive overview of the project's methodology is given in this chapter. Creating a solid and reliable image classification system requires completing each stage, which includes data preprocessing, feature extraction, model training, validation, and evaluation. The thorough procedures and explanations guarantee understanding, make replication easier, and make the methods used easier to replicate. This methodology lays the foundation for the results and analysis presented in the next chapter, where the effectiveness of the proposed solutions will be thoroughly evaluated and discussed.

Chapter 5

Results and Discussion

Introduction

The results of the approaches used in the previous chapter are presented and examined in this chapter. A detailed discussion is held regarding the outcomes of the several phases of the image classification process, which include feature extraction, data preprocessing, model training, and evaluation. The performance of the several models and methods used, including the advantages of ensemble learning, the influence of feature extraction using YOLOv5, and the efficacy of data augmentation schemes, are also compared. The consequences of these findings are further explored in this chapter, which also emphasizes their importance for the project's goals and offers suggestions for future development areas.

5.1 Results

5.1.1 Results of Solution 1: Models Performance Analysis

This section summarizes the findings of Solution 1's implementation, with a focus on evaluating the performance of various deep-learning models and ensemble learning approaches.

5.1.1.1 Individual Model Performance

```
NASNetLarge Model Confusion Matrix:
 [[99 53]
 [55 72]]
InceptionV3 Model Confusion Matrix:
 [[68 75]
 [37 99]]
DenseNet169 Model Confusion Matrix:
 [[88 61]
 [44 86]]
EfficientNetB7 Model Confusion Matrix:
 [[157   0]
 [122   0]]
            Model  Accuracy  F1-Score  Precision    Recall
0     NASNetLarge  0.612903  0.571429   0.576000  0.566929
1     InceptionV3  0.598566  0.638710   0.568966  0.727941
2     DenseNet169  0.623656  0.620939   0.585034  0.661538
3  EfficientNetB7  0.562724  0.000000   0.000000  0.000000
```

Figure 5.1: Confusion Matrix of 4 Models.

The **NASNetLarge** model's confusion matrix (Figure 5.1) shows that the model's overall accuracy was 61.29 percent because it misclassified 53 instances of "Fail" as "Pass" and 55 examples of "Pass" as "Fail." The model had superior performance with an F1-score of 0.5714, recall of 0.5669, and precision of 0.5760.

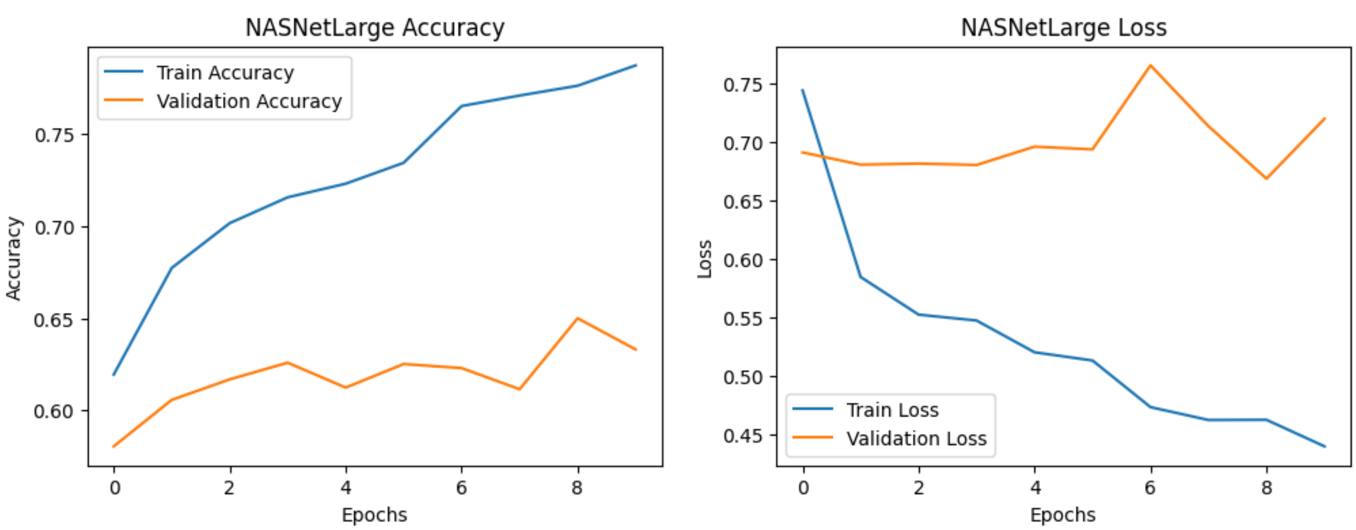

Figure 5.2: Accuracy and Loss plot of NASNetLarge Model.

The **NASNetLarge** model exhibited a steady improvement in training accuracy, reaching approximately 0.75 by the final epoch (Figure 5.2). However, there was a variation in the validation accuracy, which suggested overfitting. The model was able to reduce errors in the training data but had difficulty maintaining consistent performance on the unknown validation data, as evidenced by a steadily decreasing training loss while the validation loss fluctuated.

With a slightly lower accuracy of 59.86 percent, the **InceptionV3** model scored a higher F1-score of 0.6387, indicating a better-balanced trade-off between recall and precision. According to the confusion matrix (Figure 5.1), **InceptionV3** accurately classified 99 ”Pass” and 68 ”Fail” pictures, with a few misclassifications in each category. Though at a lower precision, the precision and recall measures show that InceptionV3 was very good at recalling ”Pass” instances.

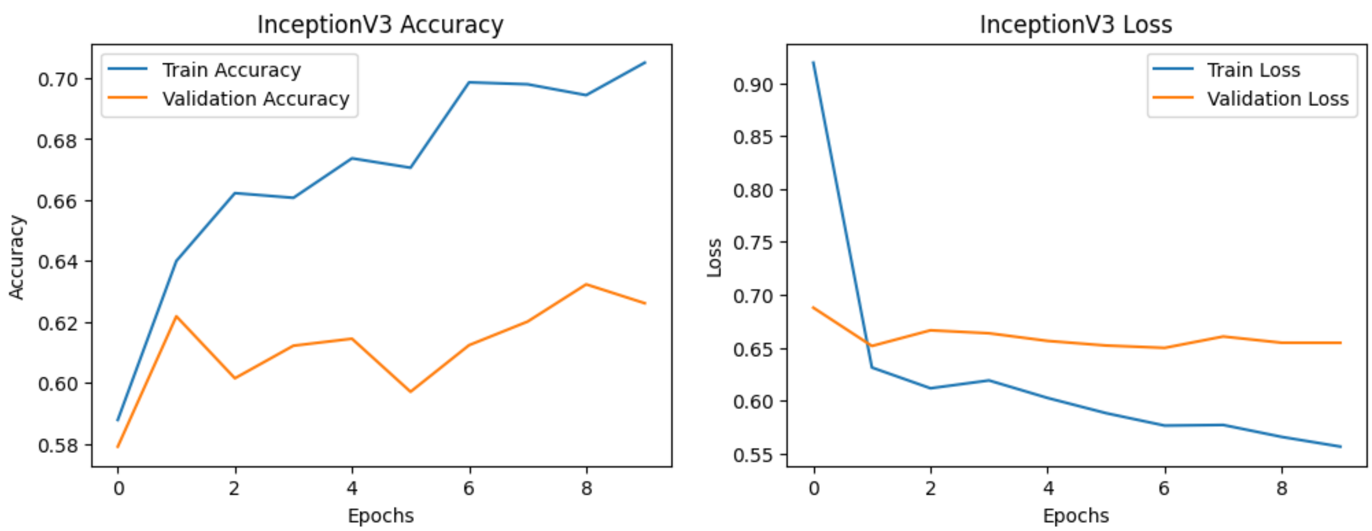

Figure 5.3: Accuracy and Loss plot of InceptionV3 Model.

The performance of **InceptionV3**, as depicted in Figure 5.3, demonstrated a significant increase in training accuracy, with validation accuracy showing some fluctuations. This shows that, while the model trained well, it struggled to generalize to new data. The training loss decreased suddenly at first, then gradually, showing effective learning. However, the stable validation loss indicates that there may be overfitting.

The **DenseNet169** model has the highest accuracy of all the models, measuring 62.37 percent. According to the confusion matrix (Figure 5.1), it correctly classified 88 instances of the ”Fail” class and 86 instances of the ”Pass” class. However, 44 cases of ”Pass” were incorrectly classed as ”Fail,” while 61 cases of ”Fail” were incorrectly tagged as ”Pass.” An F1-score of 0.6209, a recall of 0.6615, and an accuracy of 0.5850 were generated by the model.

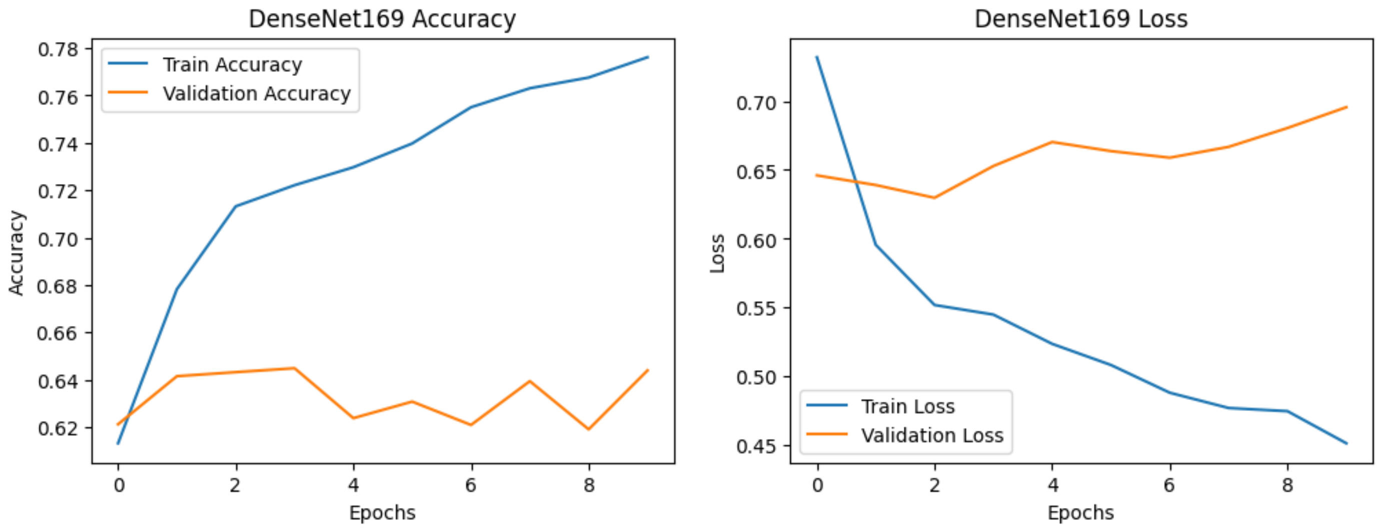

Figure 5.4: Accuracy and Loss plot of DenseNet169 Model

As shown in Figure 5.4, **DenseNet169** exhibited a consistent increase in training accuracy, reaching approximately 0.78 by the end of the training epochs. The validation accuracy also improved steadily, but at a slower rate, and with less variation than the other models. This shows that **DenseNet169** found a better balance between learning from training data and generalizing to new validation data. The training loss reduced with time, whereas the validation loss declined more gradually, confirming the conclusion that **DenseNet169** could learn successfully while preserving a reasonable level of generalization.

Compared to the other models, the **EfficientNetB7** model performed significantly worse, with an overall accuracy of only 56.27 percent. The confusion matrix (Figure 5.1) reveals a severe issue: all 122 ”Pass” samples are mistakenly categorized as ”Fail” by the model, which failed to correctly classify any instances of the ”Pass” class. This yielded an F1-score, precision, and recall of 0.000, indicating that **EfficientNetB7** confronted significant difficulty and was unable to generate meaningful generalizations with this dataset. Overfitting or insufficient feature extraction could be the reason for the model's poor performance, indicating that its design is inappropriate for this specific classification task.

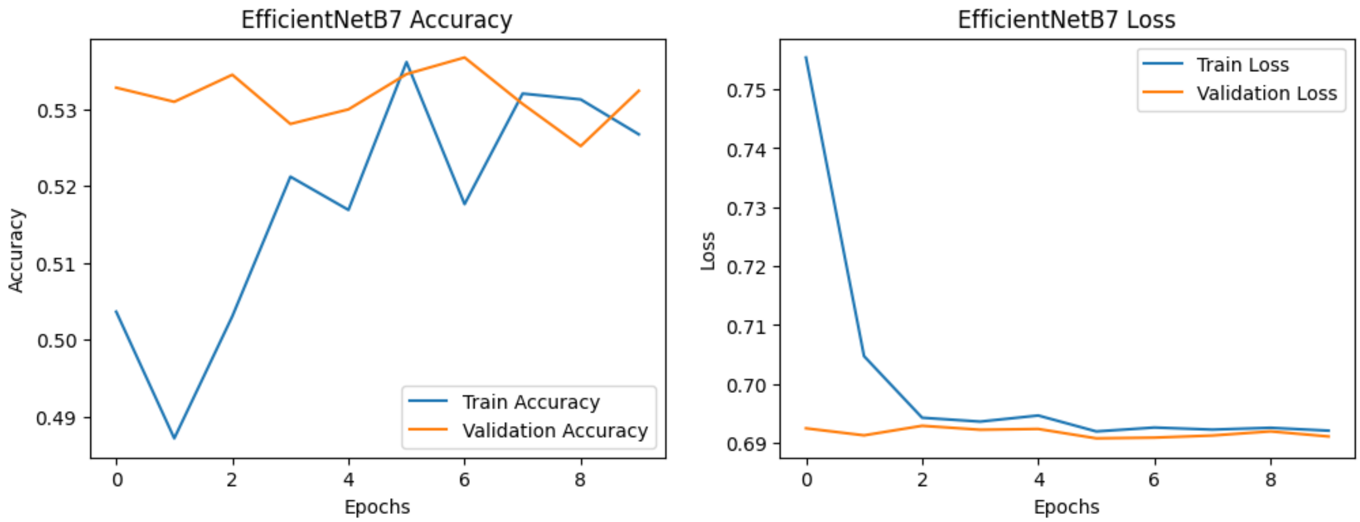

Figure 5.5: Accuracy and Loss plot of EfficientB7 Model.

The training and validation dynamics of **EfficientNetB7**, depicted in Figure 5.5, indicate that the model had a little increase in training accuracy but remained at a low level. The validation accuracy exhibited instability and even demonstrated a drop towards the conclusion of the training epochs, so further suggesting inadequate generalization. While the validation loss showed notable variations instead, the training loss dropped quickly and followed no pattern. These dynamics indicate that, either overfitting or a mismatch between the model architecture and the dataset.

5.1.1.2 Confusion matrix and classification metrics for ensemble model

The confusion matrix and classification report offer comprehensive information on the accuracy and ability of the Ensemble Model to generalize across different classes, enabling a further evaluation of its performance.

```
Confusion Matrix for Ensemble Model:
[[237  73]
 [110 180]]

Classification Report for Ensemble Model:
              precision    recall  f1-score   support

        Fail       0.68      0.76      0.72       310
        Pass       0.71      0.62      0.66       290

    accuracy                           0.69       600
   macro avg       0.70      0.69      0.69       600
weighted avg       0.70      0.69      0.69       600

F1-Score for Ensemble Model: 0.6629834254143646
Precision for Ensemble Model: 0.7114624505928854
Recall for Ensemble Model: 0.6206896551724138
```

Figure 5.6: Confusion Matrix of Ensemble Model.

The confusion matrix of the figure 5.6 indicates that the Ensemble Model accurately classified 237 cases as belonging to the ”Fail” class and 180 instances as belonging to the ”Pass” class. Unfortunately, 73 cases were incorrectly labeled as ”Pass” and 110 cases were incorrectly labeled as ”Fail”. These misclassifications point to potential improvements in the model’s forecast accuracy.

With a precision of 0.68 for the ”Fail” class, were found to be appropriate. The precision for the ”Pass” class was higher at 0.71, suggesting that it performed better in terms of forecasting instances of ”Pass.” The recall for the ”Fail” class was higher at 0.76, indicating that the model was more successful at recognizing them. On the other hand, the recall for the ”Pass” class was lower at 0.62, meaning that the model missed a sizable percentage of actual «Pass» cases.

The model got an overall F1-score of 0.69, based on precision and recall, of 0.72 for the ”Fail” class and 0.66 for the ”Pass” class. This suggests that both courses have a moderate balance between recall and precision. With an overall accuracy of 69 percent. Both the weighted average and macro-average F1 scores were 0.69, indicating a balanced performance in both classes. To enhance its capacity to accurately identify instances of this class, the model may require additional optimization, as indicated by the reduced recall for the ”Pass” class.

5.1.1.3 Summary of Findings

The evaluation of Solution 1 across various deep learning models reveals a range of performances, with some models like **DenseNet169** and **NASNetLarge** showing robust results, while others like

EfficientNetB7 struggled significantly with this dataset. The ensemble approach provided a moderate improvement in overall precision, though it also highlighted the trade-offs between precision and recall.

5.1.2 Results of Solution 2 Model (DenseNet201) Performance Analysis

This section provides a comprehensive analysis of the performance results of Solution 2, where it employs a more detailed approach by separating the dataset into particular checkpoints for categorization. The results are evaluated for each checkpoint, emphasizing the models' performance in terms of accuracy, precision, recall, F1-score, and overall learning curves.

Checkpoint 30047: Jumper Installed Correctly

```
Classification Report for AIML CHECKPOINTS\30047 Jumper installed correctly:
              precision    recall  f1-score   support

           0       1.00      0.23      0.38        13
           1       0.52      1.00      0.69        11

    accuracy                           0.58        24
   macro avg       0.76      0.62      0.53        24
weighted avg       0.78      0.58      0.52        24

Confusion Matrix for AIML CHECKPOINTS\30047 Jumper installed correctly:
 [[ 3 10]
 [ 0 11]]
```

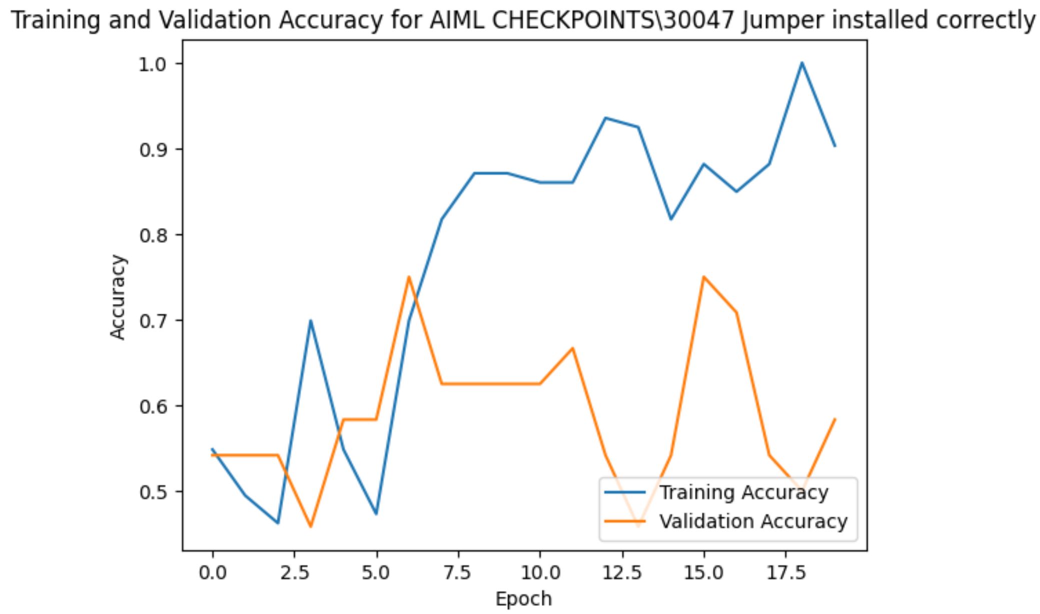

Figure 5.7: AI/ML checkpoints 30047.

This checkpoint demonstrates a modest level of achievement since the model accurately identifies both classes to a certain extent, as evidenced by the F1-score of 0.69 for the ”Pass” class. The confusion matrix demonstrates a more equitable distribution of accurate ”Pass” and accurate ”Fail” predictions. However, the training and validation accuracy plot exhibits indications of instability and overfitting, characterized by substantial oscillations in validation accuracy. (Figure 5.7)

Checkpoint 30080: Fibers Are Well Installed

```
Classification Report for AIML CHECKPOINTS\30080 Fibers are well installed:
              precision    recall  f1-score   support

           0       0.45      1.00      0.62        10
           1       0.00      0.00      0.00        12

    accuracy                           0.45        22
   macro avg       0.23      0.50      0.31        22
weighted avg       0.21      0.45      0.28        22

Confusion Matrix for AIML CHECKPOINTS\30080 Fibers are well installed:
 [[10  0]
 [12  0]]
```

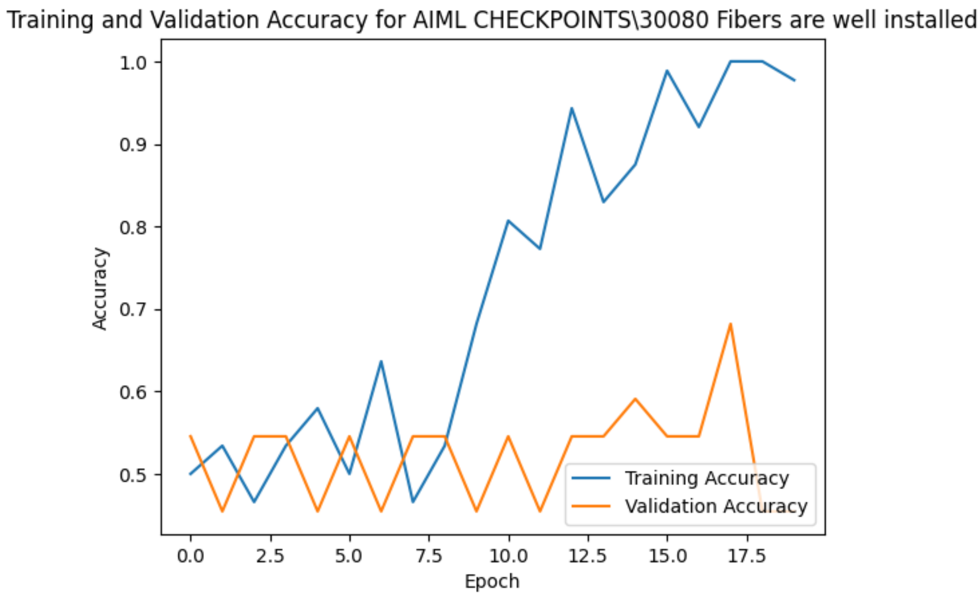

Figure 5.8: AI/ML checkpoints30080.

The model's performance in accurately identifying installed fibers reveals significant challenges. A 0 precision, recall, and F1-score in the classification report indicates that the model is unable to correctly identify any instances of properly installed fibers in this dataset(Pass images). The confusion matrix substantiates this finding, revealing that every "Pass" instance was classified as a "Fail". The training and validation accuracy plot shows more evidence of the model's weak capacity to generalize, as the validation accuracy does not demonstrate a consistent improvement, but instead varies significantly and regularly declines towards the end of the training period. This indicates that the model faces difficulties in acquiring valuable characteristics for this categorization assignment, either because it is overfitting or because it lacks sufficient training data. (Figure 5.8)

Checkpoint 30081: Internal Cables Installed Correctly - Airscale BBU (AMIA)

```
Classification Report for AIML CHECKPOINTS\30081 Internal cables installed correctly - Airscale BBU (AMIA):
              precision    recall  f1-score   support

           0       0.47      0.90      0.62        10
           1       0.67      0.17      0.27        12

    accuracy                           0.50        22
   macro avg       0.57      0.53      0.44        22
weighted avg       0.58      0.50      0.43        22

Confusion Matrix for AIML CHECKPOINTS\30081 Internal cables installed correctly - Airscale BBU (AMIA):
 [[ 9  1]
 [10  2]]
```

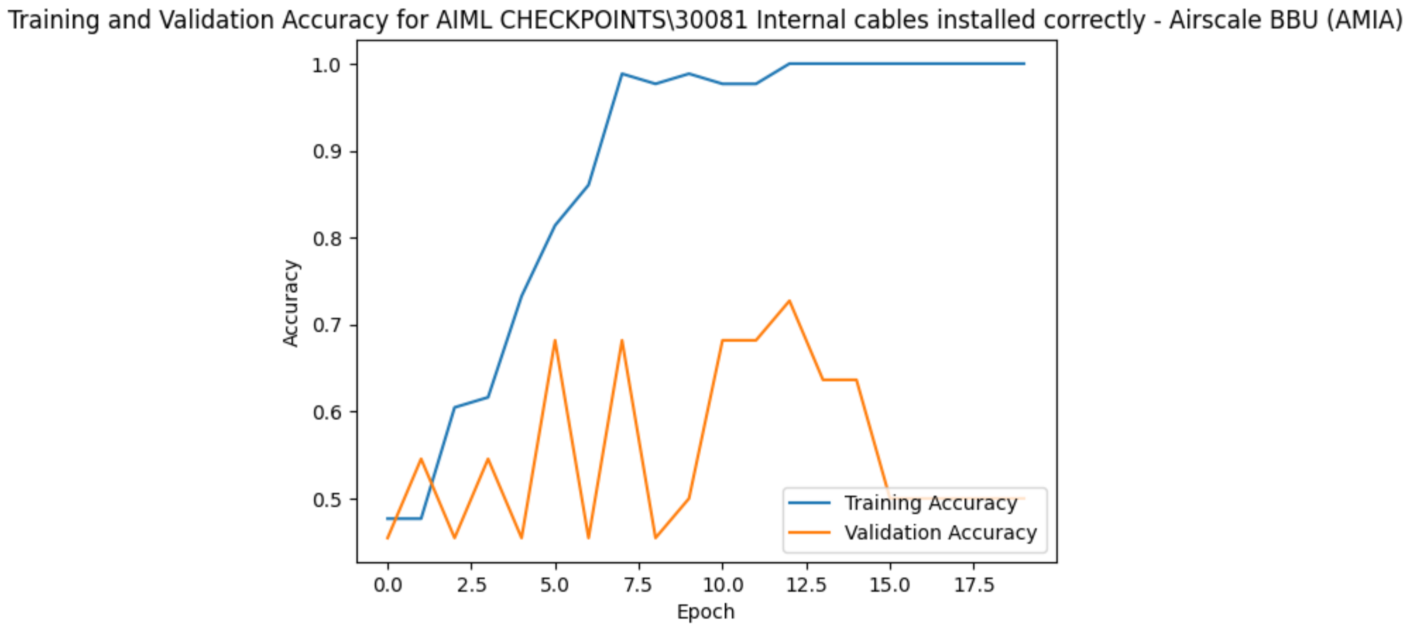

Figure 5.9: AI/ML checkpoints30081.

The model demonstrated moderate performance during the internal cable installation, the confusion matrix suggests that the model accurately classified 9 out of 10 ”Fail” instances; however, it encountered difficulty with ”Pass” instances, correctly identifying only 2 out of 12. The training and validation accuracy plots exhibit substantial fluctuations, which implies that the model is not stable and may be overfitting. The training accuracy has been gradually increasing, while the validation accuracy has fluctuated, indicating difficulties with generalization. (Figure 5.9)

Checkpoint 40001: Device Installed

```
Classification Report for AIML CHECKPOINTS\40001 Device Installed:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        12
           1       0.43      1.00      0.60         9

    accuracy                           0.43        21
   macro avg       0.21      0.50      0.30        21
weighted avg       0.18      0.43      0.26        21

Confusion Matrix for AIML CHECKPOINTS\40001 Device Installed:
 [[ 0 12]
 [ 0  9]]
```

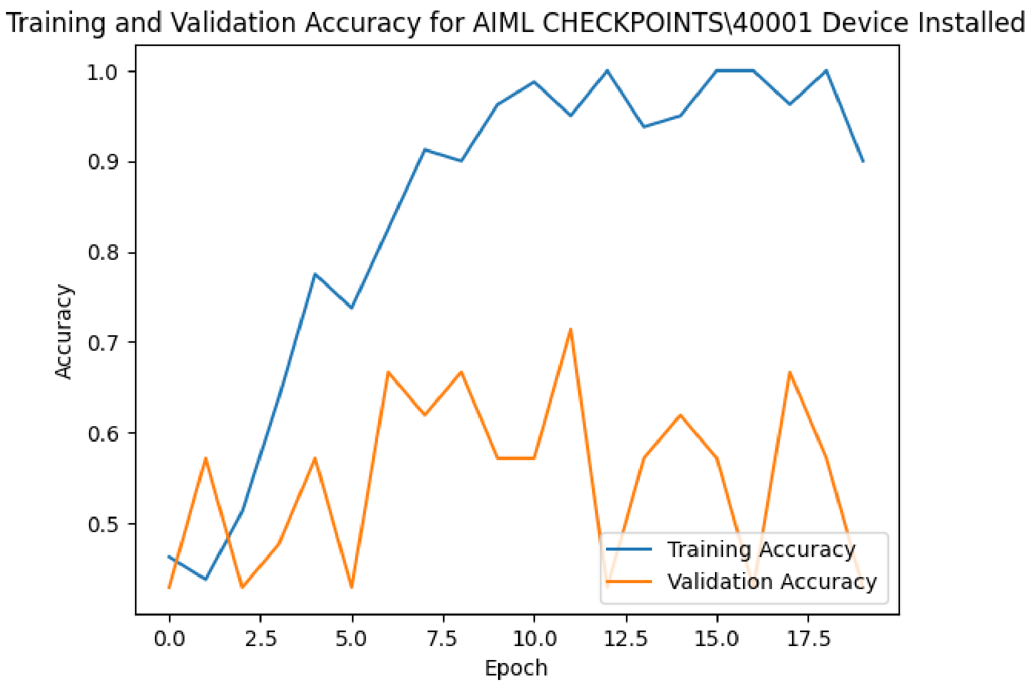

Figure 5.10: AI/ML checkpoints 40001

This model shows substantial difficulty in correctly classifying the installation of devices. The 0 precision, recall, and F1-score for class 0 indicate that the model was unable to correctly identify any instances of ”Pass” labels, according to the confusion matrix. The accuracy plot reflects the difficulties encountered since the validation accuracy does not show constant improvement. This suggests that there may be overfitting or a lack of ability to distinguish features in the training phase. (Figure 5.10)

Checkpoint 40002: Equipment Installation Pictures Including All Cables

```
Classification Report for AIML CHECKPOINTS\40002 Equipment installation pictures including all cables:
              precision    recall  f1-score   support

           0       0.62      1.00      0.76        13
           1       0.00      0.00      0.00         8

    accuracy                           0.62        21
   macro avg       0.31      0.50      0.38        21
weighted avg       0.38      0.62      0.47        21

Confusion Matrix for AIML CHECKPOINTS\40002 Equipment installation pictures including all cables:
 [[13  0]
 [ 8  0]]
```

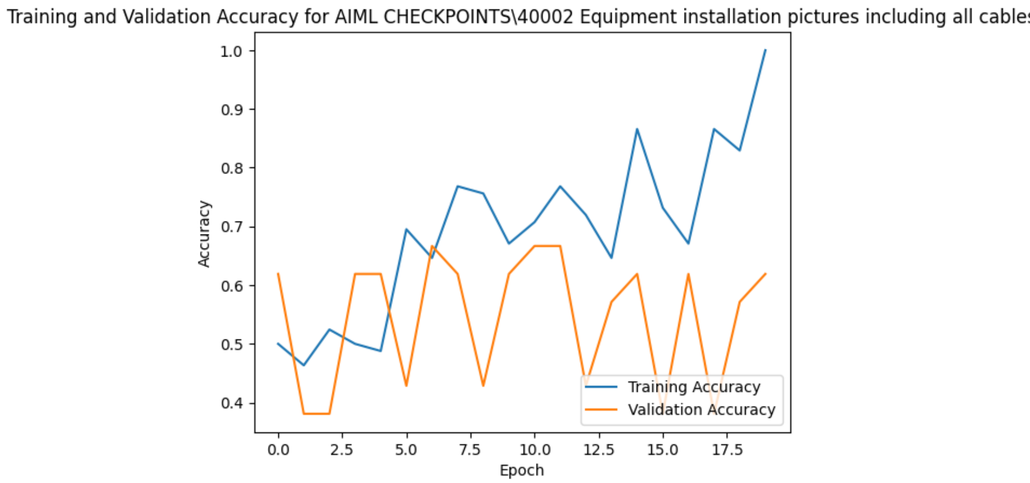

Figure 5.11: AI/ML checkpoints 40002

The categorization accuracy for equipment installation presents comparable difficulties. The precision and recall scores of 0 show that the identification of ”Pass” cases is inaccurate, according to the confusion matrix. The plot of training and validation accuracy indicates that the model faces difficulties in generalizing, as evidenced by the inconsistent validation accuracy observed during the training process. (Figure 5.11).

Checkpoint 40003: All Cables Labeling Pictures for Power Fiber Grounding

```
Classification Report for AIML CHECKPOINTS\40003 All cables Labeling pictures for power_fiber_grounding:
              precision    recall  f1-score   support

           0       1.00      0.67      0.80        18
           1       0.54      1.00      0.70         7

    accuracy                           0.76        25
   macro avg       0.77      0.83      0.75        25
weighted avg       0.87      0.76      0.77        25

Confusion Matrix for AIML CHECKPOINTS\40003 All cables Labeling pictures for power_fiber_grounding:
 [[12  6]
 [ 0  7]]
```

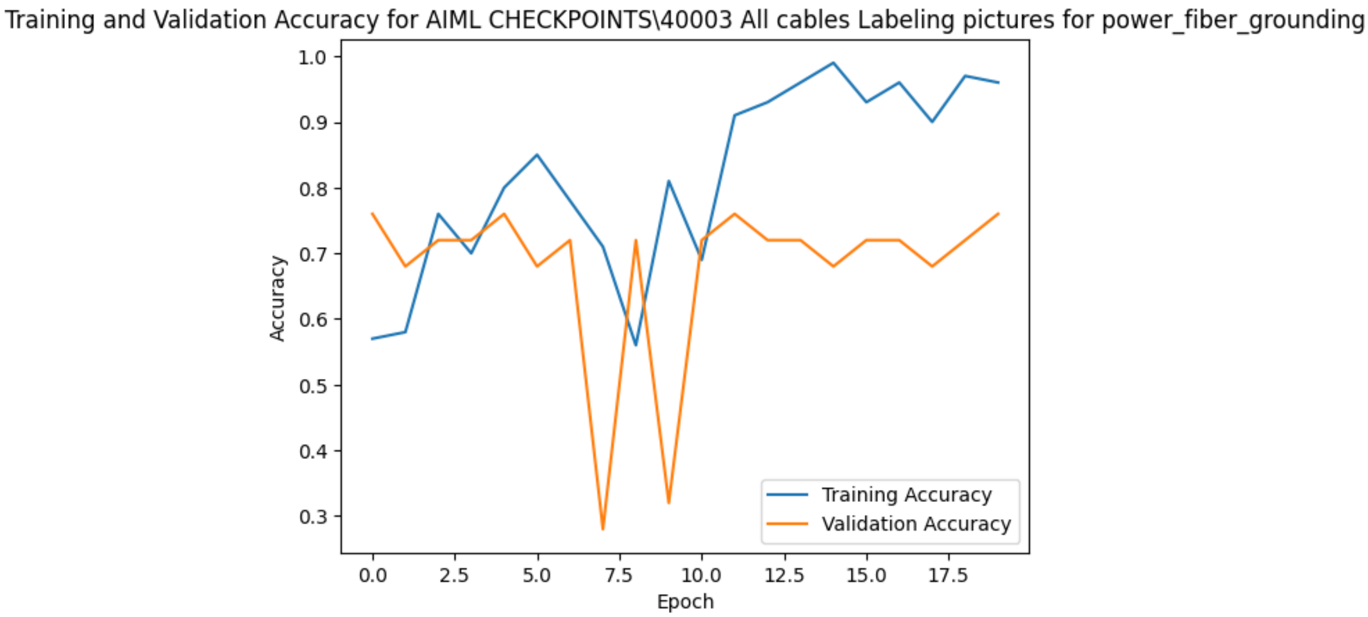

Figure 5.12: AI/ML checkpoints 40003.

The model performed moderately well with an overall accuracy of 0.76. The precision for the ”Fail” class was 1.00, indicating that when the model predicted ”Fail,” it was always correct. Nevertheless, the recall rate for this class was just 0.67, indicating that the model failed to identify certain occurrences of ”Fail.” The ”Pass” class demonstrated a precision of 0.54, signifying that the model accurately classified 54 of the instances labeled as ”Pass”. Nevertheless, it attained a flawless recall rate of 1.00, indicating that it accurately detected all cases classified as ”Pass” but also generated some incorrect ”Pass” results. The confusion matrix showed that the model struggled more with correctly identifying ”Fail” cases, despite high accuracy in other areas. (Figure 5.12)

Checkpoint 40004: Shelf or PDU Fixed Properly

```
Classification Report for AIML CHECKPOINTS\40004 Shelf or PDU fixed properly:
              precision    recall  f1-score   support

           0       0.67      0.84      0.74        19
           1       0.00      0.00      0.00         8

    accuracy                           0.59        27
   macro avg       0.33      0.42      0.37        27
weighted avg       0.47      0.59      0.52        27

Confusion Matrix for AIML CHECKPOINTS\40004 Shelf or PDU fixed properly:
 [[16  3]
 [ 8  0]]
```

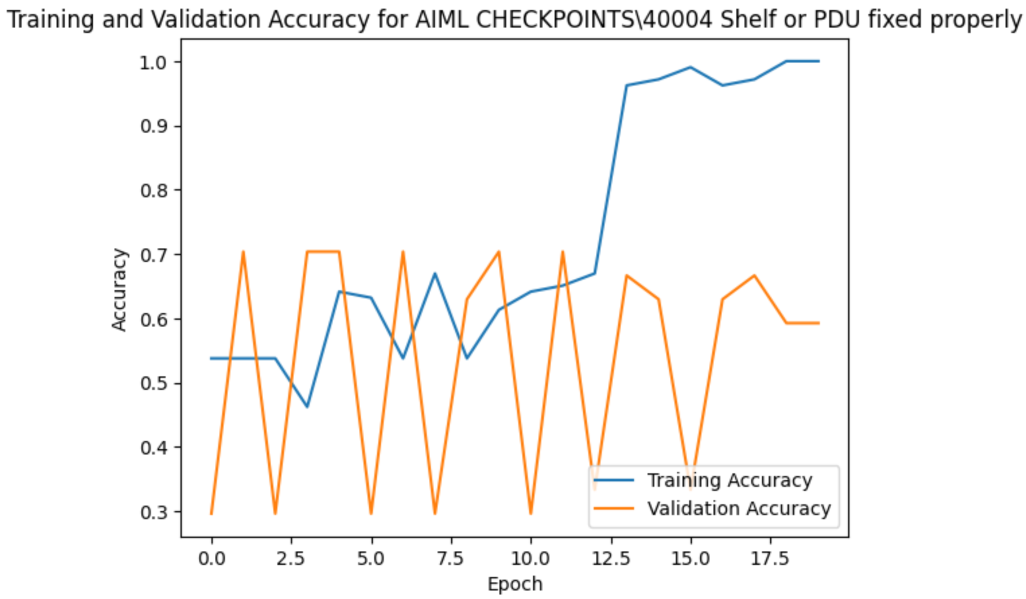

Figure 5.13: AI/ML checkpoints40004.

The model achieved an accuracy of 59 with a slight imbalance in precision and recall, as indicated by the macro average F1-score of 0.37. The training and validation accuracy plot shows improvement in training accuracy, but validation accuracy does not follow a similar trend, highlighting potential overfitting and difficulty in generalization. (Figure 5.13).

Checkpoint 40005: All Fiber Cables Installed Properly

```
Classification Report for AIML CHECKPOINTS\40005 All fiber cables installed properly:
              precision    recall  f1-score   support

           0       0.41      0.55      0.47        22
           1       0.29      0.19      0.23        21

    accuracy                           0.37        43
   macro avg       0.35      0.37      0.35        43
weighted avg       0.35      0.37      0.35        43

Confusion Matrix for AIML CHECKPOINTS\40005 All fiber cables installed properly:
 [[12 10]
 [17  4]]
```

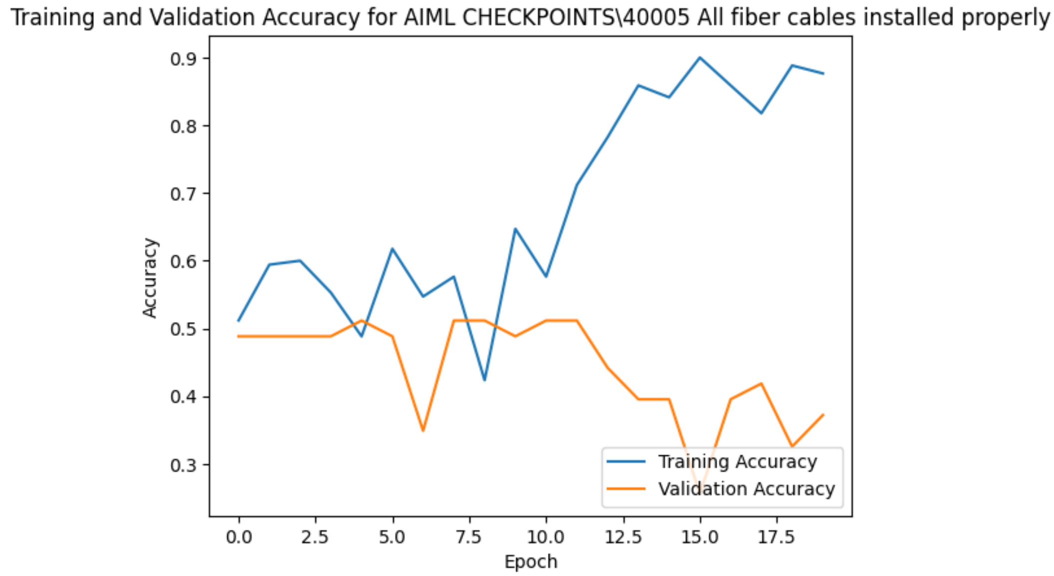

Figure 5.14: AI/ML checkpoints 40005.

For this checkpoint, the model had a lower overall accuracy of 0.37. Precision and recall were significantly lower for both groups, particularly for the ”Pass” class, which had a precision of 0.29 and a recall of 0.19. This highlights a substantial obstacle in accurately detecting properly installed cables, with a considerable amount of misclassifications. The training and validation accuracy curves exhibited significant fluctuations, suggesting that the model was overfitting to the training data and failing to generalize effectively to the validation set. (Figure 5.14).

Checkpoint 40006: Fiber Management

```
Classification Report for AIML CHECKPOINTS\40006 Fiber management:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        14
           1       0.56      1.00      0.72        18

    accuracy                           0.56        32
   macro avg       0.28      0.50      0.36        32
weighted avg       0.32      0.56      0.40        32

Confusion Matrix for AIML CHECKPOINTS\40006 Fiber management:
 [[ 0 14]
 [ 0 18]]
```

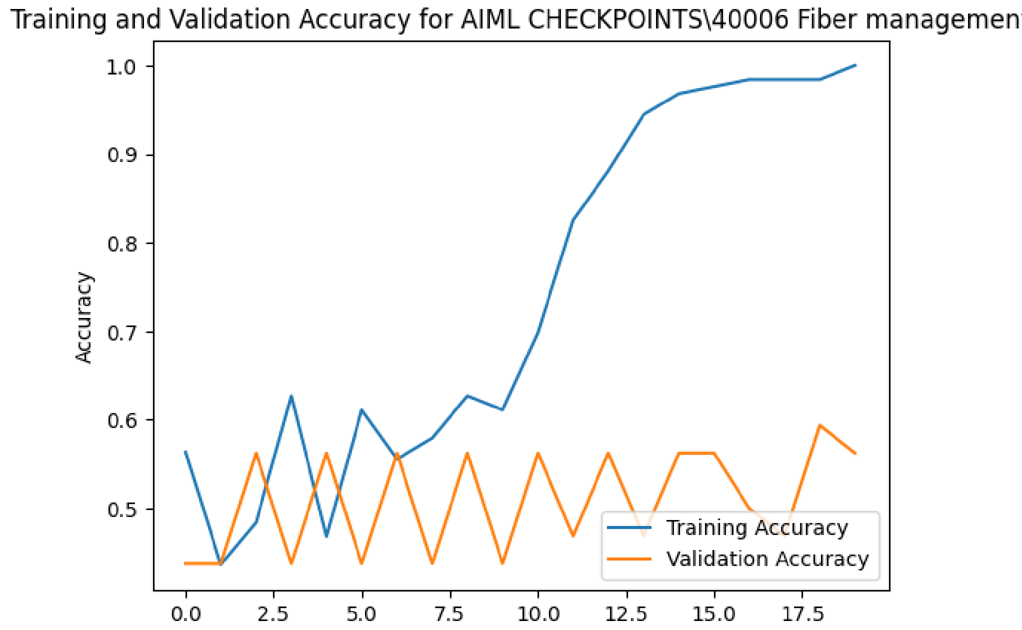

Figure 5.15: AI/ML checkpoints 40006.

The performance at this checkpoint was slightly better, with an accuracy of 0.56. The ”Pass” class showed powerful performance with a recall of 1.00, but the ”Fail” class had no correct identifications (precision and recall of 0.00). This result indicates that the model was strongly biased toward predicting ”Pass” and failed to detect ”Fail” events. The mismatch in training and validation accuracy revealed overfitting as well, according to the accuracy curve. (Figure 5.15).

Checkpoint 40007: Is the Rack Securely Installed Properly

```
Classification Report for AIML CHECKPOINTS\40007 Is the rack securely installed properly:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        11
           1       0.48      1.00      0.65        10

    accuracy                           0.48        21
   macro avg       0.24      0.50      0.32        21
weighted avg       0.23      0.48      0.31        21

Confusion Matrix for AIML CHECKPOINTS\40007 Is the rack securely installed properly:
 [[ 0 11]
 [ 0 10]]
```

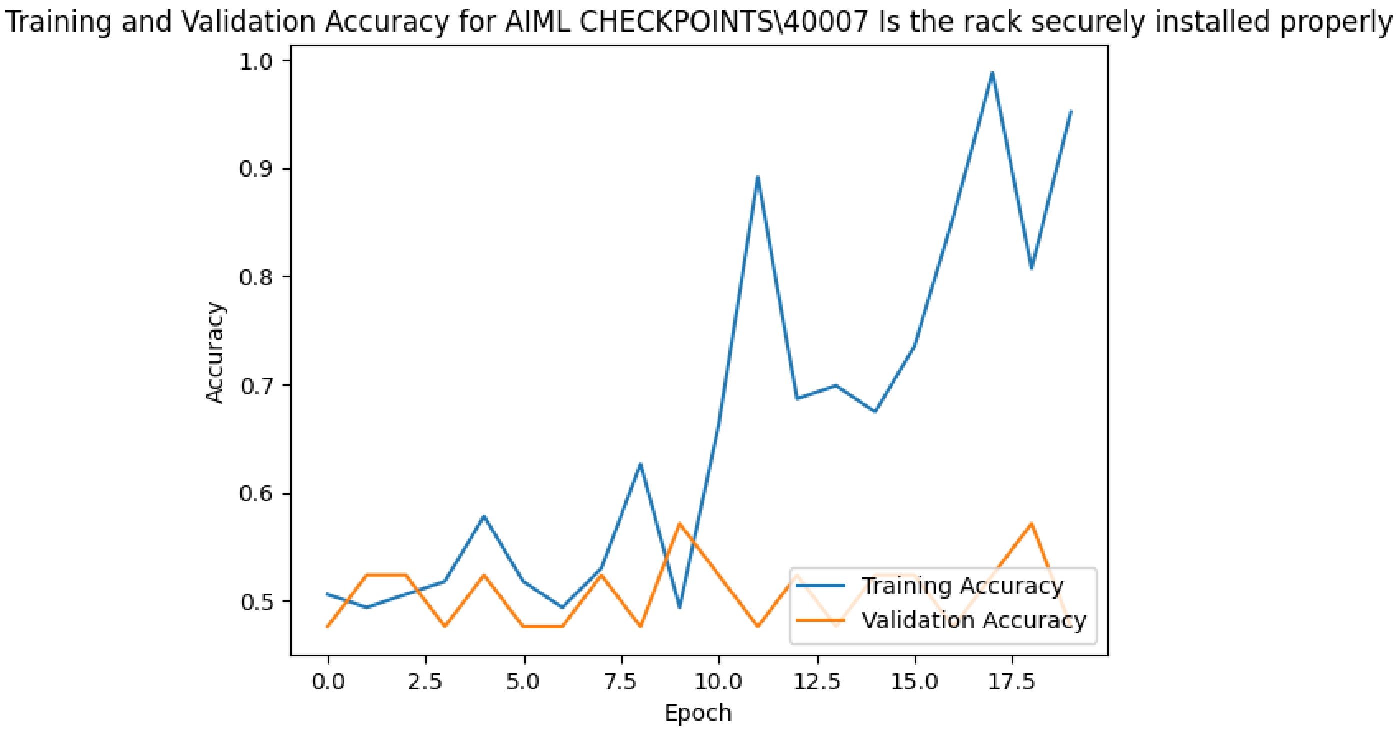

Figure 5.16: AI/ML checkpoints 40007.

The model's accuracy here was 0.48, with both precision and recall for the "Fail" class being 0.00, meaning the model failed to identify any "Fail" cases correctly. The "Pass" class, on the other hand, had a recall of 1.00, but its precision was only 0.48. This imbalance demonstrates that the model was once again biased toward the majority class. The considerable variance in the validation accuracy curve indicates overfitting and a lack of generalization capacity. (Figure 5.16).

Checkpoint 40008: Shelf (sub shelf PDU) Grounding Side

```
Classification Report for AIML CHECKPOINTS\40008 (Shelf_sub-shelf_PDU) grounding side:
              precision    recall  f1-score   support

           0       1.00      0.08      0.15        12
           1       0.42      1.00      0.59         8

    accuracy                           0.45        20
   macro avg       0.71      0.54      0.37        20
weighted avg       0.77      0.45      0.33        20

Confusion Matrix for AIML CHECKPOINTS\40008 (Shelf_sub-shelf_PDU) grounding side:
 [[ 1 11]
 [ 0  8]]
```

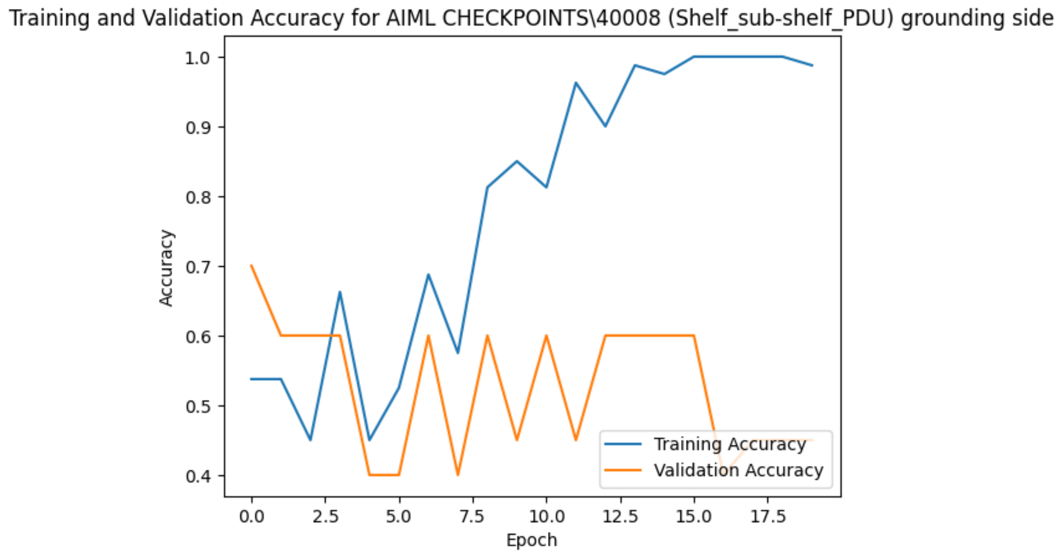

Figure 5.17: AI/ML checkpoints 40008.

The findings show that the model performs badly, with significant misclassification. The confusion matrix demonstrates that the model fails to accurately identify the majority of ”Pass” examples, which is reflected in the precision and recall ratings. The accuracy plot shows that, while the model fits the training data, it struggles to generalize to the validation set, indicating overfitting and insufficient feature learning. (Figure 5.17).

Checkpoint 40009: Power Cabling Path Check

```
Classification Report for AIML CHECKPOINTS\40009 Power cabling Path check:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00        11
           1       0.52      1.00      0.69        12

    accuracy                           0.52        23
   macro avg       0.26      0.50      0.34        23
weighted avg       0.27      0.52      0.36        23

Confusion Matrix for AIML CHECKPOINTS\40009 Power cabling Path check:
 [[ 0 11]
 [ 0 12]]
```

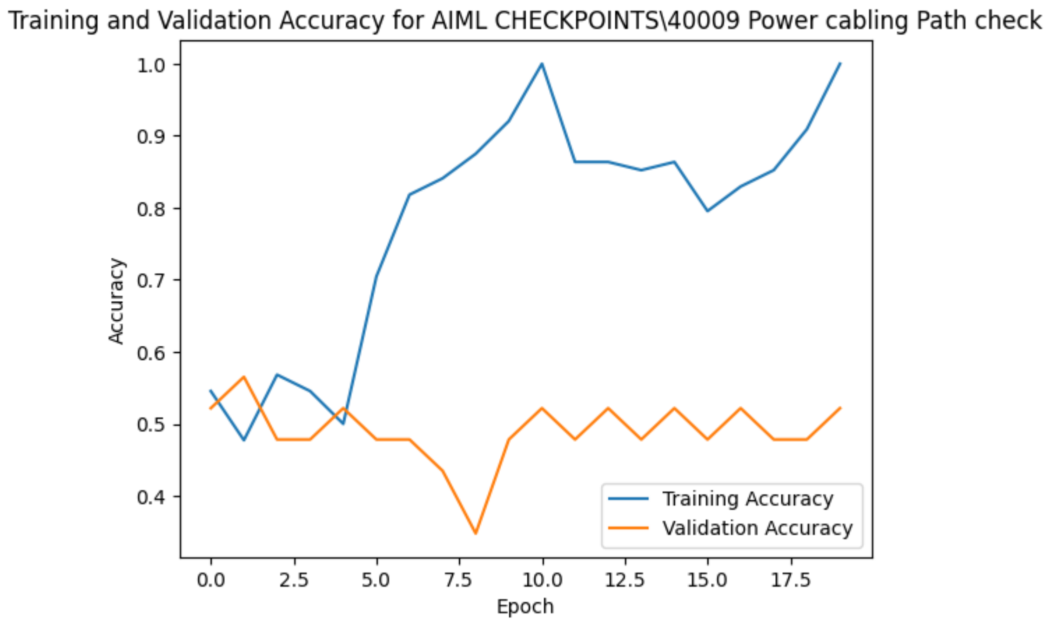

Figure 5.18: AI/ML checkpoints 40009.

The performance of the model in this checkpoint is below the ideal level, as evidenced by the classification report and confusion matrix. The precision and recall values for the ”Fail” class are both 0, indicating that the model is unable to accurately classify any instances of the ”Fail” class. The overall accuracy is low, at 0.52. The model's inability to generalize and potential overfitting to the training data are indicated by the substantial fluctuations in the validation accuracy plot between the training and validation data. The model's architecture or the features being utilized may not be optimally suited for this particular purpose. (Figure 5.18).

Checkpoint 40010: Fuses and Breakers Installed on the TRU PDU Side

```
Classification Report for AIML CHECKPOINTS\40010 Fuses and breakers are installed on the TRU_PDU side Power cables connection check (PDU_TRU side):
              precision    recall  f1-score   support

           0       0.33      0.50      0.40         2
           1       0.75      0.60      0.67         5

    accuracy                           0.57         7
   macro avg       0.54      0.55      0.53         7
weighted avg       0.63      0.57      0.59         7

Confusion Matrix for AIML CHECKPOINTS\40010 Fuses and breakers are installed on the TRU_PDU side Power cables connection check (PDU_TRU side):
 [[1 1]
 [2 3]]
```

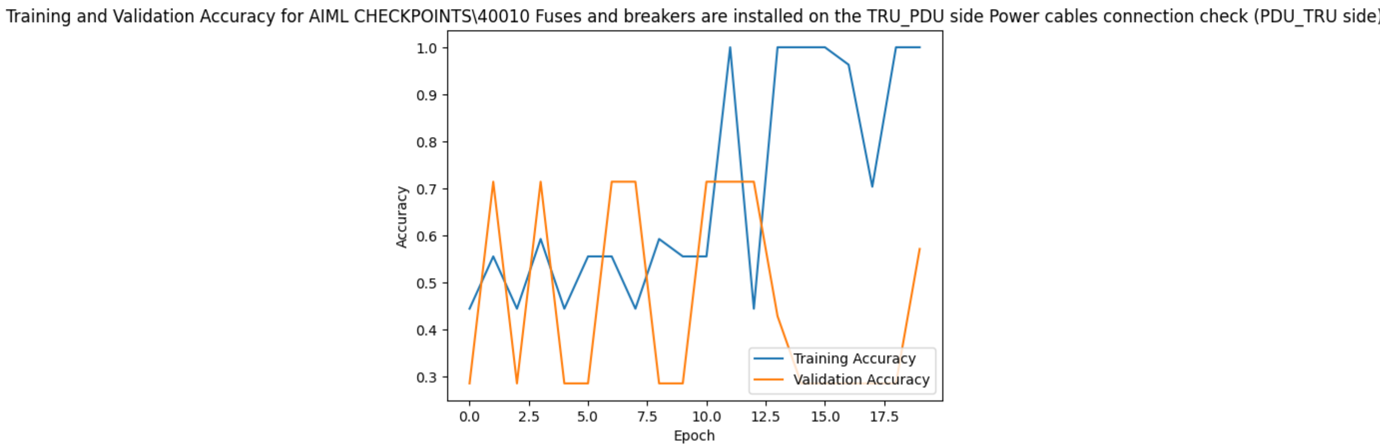

Figure 5.19: AI/ML checkpoints 40010.

In these checkpoints, the confusion matrix shows that the model correctly identified most instances but still had some misclassifications. The plot displaying the training and validation accuracy reveals that the training accuracy exhibits a consistent upward trend, whereas the validation accuracy displays notable fluctuations. This implies the possibility of overfitting or challenges in effectively generalizing to unfamiliar data. Further refinement or augmentation of data could enhance the stability and accuracy of this model. (Figure 5.19).

Checkpoint 40011: All Cabinet Doors/Covers are Closed

```
Classification Report for AIML CHECKPOINTS\40011 All cabinet doors_covers are closed:
              precision    recall  f1-score   support

           0       0.59      1.00      0.74        13
           1       1.00      0.10      0.18        10

    accuracy                           0.61        23
   macro avg       0.80      0.55      0.46        23
weighted avg       0.77      0.61      0.50        23

Confusion Matrix for AIML CHECKPOINTS\40011 All cabinet doors_covers are closed:
 [[13  0]
 [ 9  1]]
```

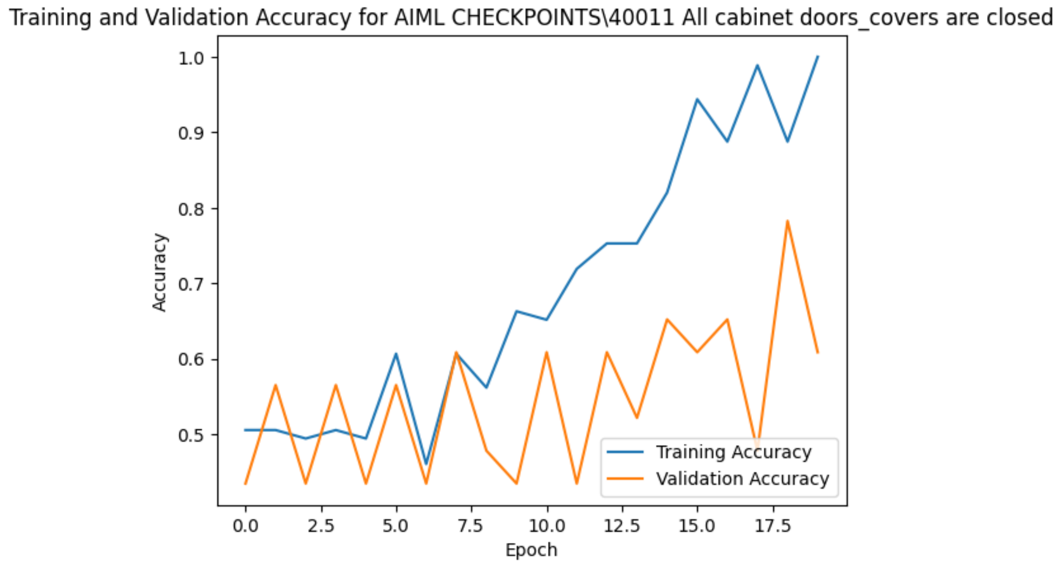

Figure 5.20: AI/ML checkpoints 40011.

The model shows consistent and dependable performance in this checkpoint, with an accuracy of 0.61. The confusion matrix demonstrates an equitable classification of both ”Pass” and ”Fail” categories. Nevertheless, the accuracy plot demonstrates a degree of instability, as the validation accuracy does not exhibit continuous improvement throughout the epochs. This implies that although the model can acquire knowledge from the training data, it faces difficulties in applying that knowledge to new situations, which suggests a possible requirement for stronger regularization methods or a wider range of training data. (Figure 5.20).

Checkpoint 40012: All Ground Cables are Installed Properly

```
Classification Report for AIML CHECKPOINTS\40012 All Ground cables are installed properly:
              precision    recall  f1-score   support

           0       1.00      0.09      0.17        11
           1       0.47      1.00      0.64         9

    accuracy                           0.50        20
   macro avg       0.74      0.55      0.40        20
weighted avg       0.76      0.50      0.38        20

Confusion Matrix for AIML CHECKPOINTS\40012 All Ground cables are installed properly:
 [[ 1 10]
 [ 0  9]]
```

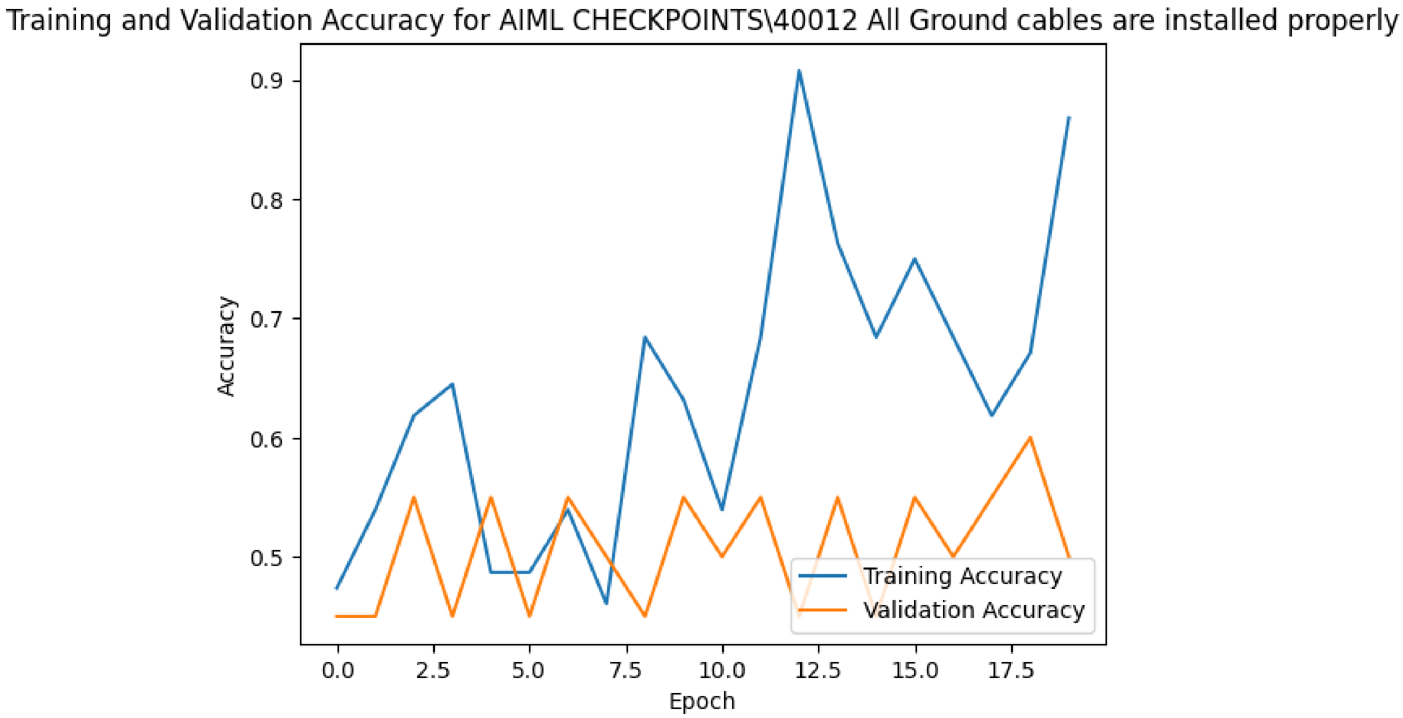

Figure 5.21: AI/ML checkpoints 40012.

The model's ability to accurately categorize correctly installed ground wires is restricted, as seen by the overall accuracy of 50 and an F1-score of 0.64 for the "Pass" class. The confusion matrix reveals that the model frequently struggles to accurately categorize "Fail" situations, as indicated by the low recall score. The curve depicting the accuracy of the training and validation data exhibits notable instability since the validation accuracy fluctuates during different epochs. This instability indicates overfitting and implies that the model's capacity to generalize is weakened, maybe due to a scarcity of diverse training data or inadequate model complexity. (Figure 5.21).

Checkpoint 40013: Excess Length of Patch Cords Properly Stored

```
Classification Report for AIML CHECKPOINTS\40013 Excess length of the Patch cords properly stored in the FO cable organizer_fiber storage tray:
              precision    recall  f1-score   support

           0       0.44      1.00      0.62         4
           1       0.00      0.00      0.00         5

    accuracy                           0.44         9
   macro avg       0.22      0.50      0.31         9
weighted avg       0.20      0.44      0.27         9

Confusion Matrix for AIML CHECKPOINTS\40013 Excess length of the Patch cords properly stored in the FO cable organizer_fiber storage tray:
 [[4 0]
 [5 0]]
```

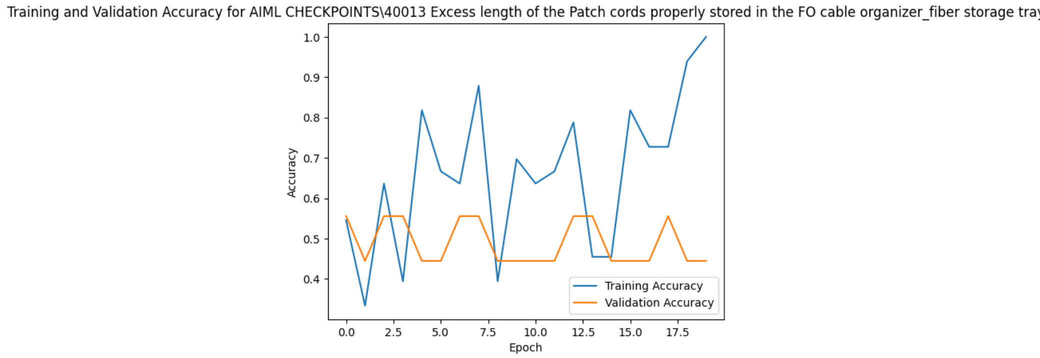

Figure 5.22: AI/ML checkpoints 40013.

The model has substantial difficulties with this checkpoint, as evidenced by the classification report, which reveals a 0 precision, recall, and F1-score for the ”Pass” class. The confusion matrix provides additional confirmation by indicating that there were no instances of ”Pass” cases that were accurately identified. The training and validation accuracy plot reveals an obvious lack of improvement in validation accuracy, with significant fluctuations, indicating that the model is overfitting the training data and failing to generalize. This may require rethinking the feature selection or model architecture. (Figure 5.22).

Checkpoint 40014: Insulated Tools

```
Classification Report for AIML CHECKPOINTS\40014 Insulated Tools:
              precision    recall  f1-score   support

           0       0.75      1.00      0.86        12
           1       1.00      0.56      0.71         9

    accuracy                           0.81        21
   macro avg       0.88      0.78      0.79        21
weighted avg       0.86      0.81      0.80        21

Confusion Matrix for AIML CHECKPOINTS\40014 Insulated Tools:
 [[12  0]
 [ 4  5]]
```

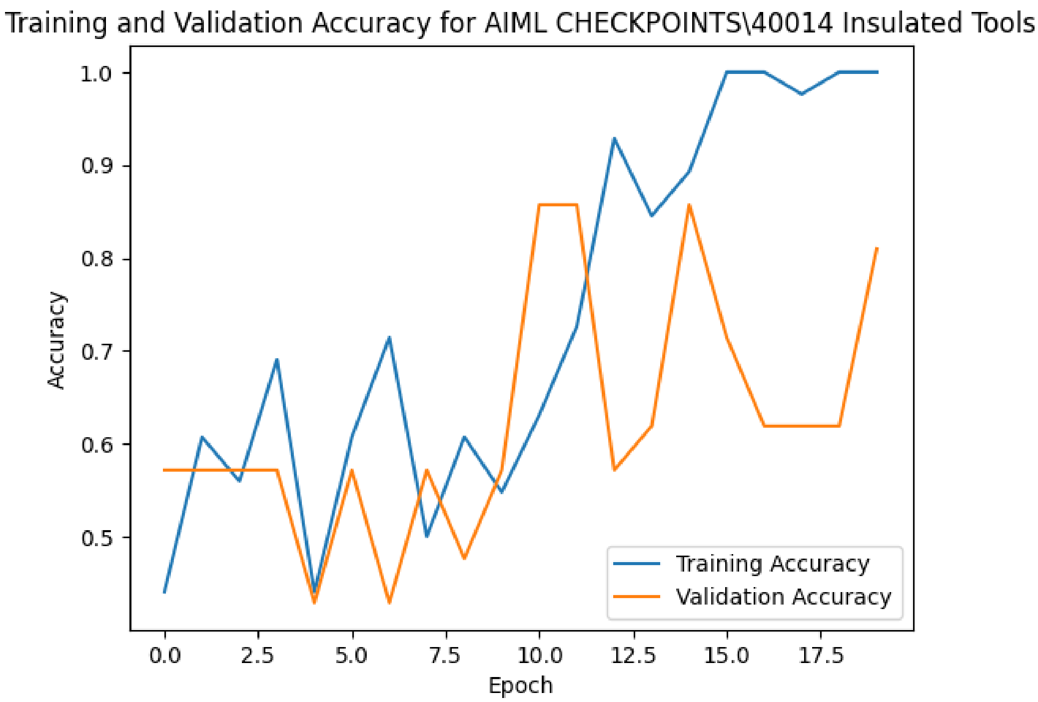

Figure 5.23: AI/ML checkpoints 40014.

The model's performance in the insulated tools checkpoint is commendable, with an overall accuracy of 0.81. The confusion matrix demonstrates a more equitable distribution of accurately categorized events across both classes. The plot depicting the training and validation accuracy demonstrates a consistent upward trend in both metrics, indicating that the model is effectively acquiring knowledge and possesses the ability to generalize effectively. Nevertheless, there are still variations in the validation accuracy, suggesting the possibility of enhancing it further, perhaps by fine-tuning the model or increasing the size of the dataset. (Figure 5.23).

Checkpoint 40015: FO Patch Cords Routed Through Fiber Trays

```
Classification Report for AIML CHECKPOINTS\40015 FO patch cords routed through fiber trays_flixble pipes:
              precision    recall  f1-score   support

           0       0.83      0.42      0.56        12
           1       0.53      0.89      0.67         9

    accuracy                           0.62        21
   macro avg       0.68      0.65      0.61        21
weighted avg       0.70      0.62      0.60        21

Confusion Matrix for AIML CHECKPOINTS\40015 FO patch cords routed through fiber trays_flixble pipes:
 [[5 7]
 [1 8]]
```

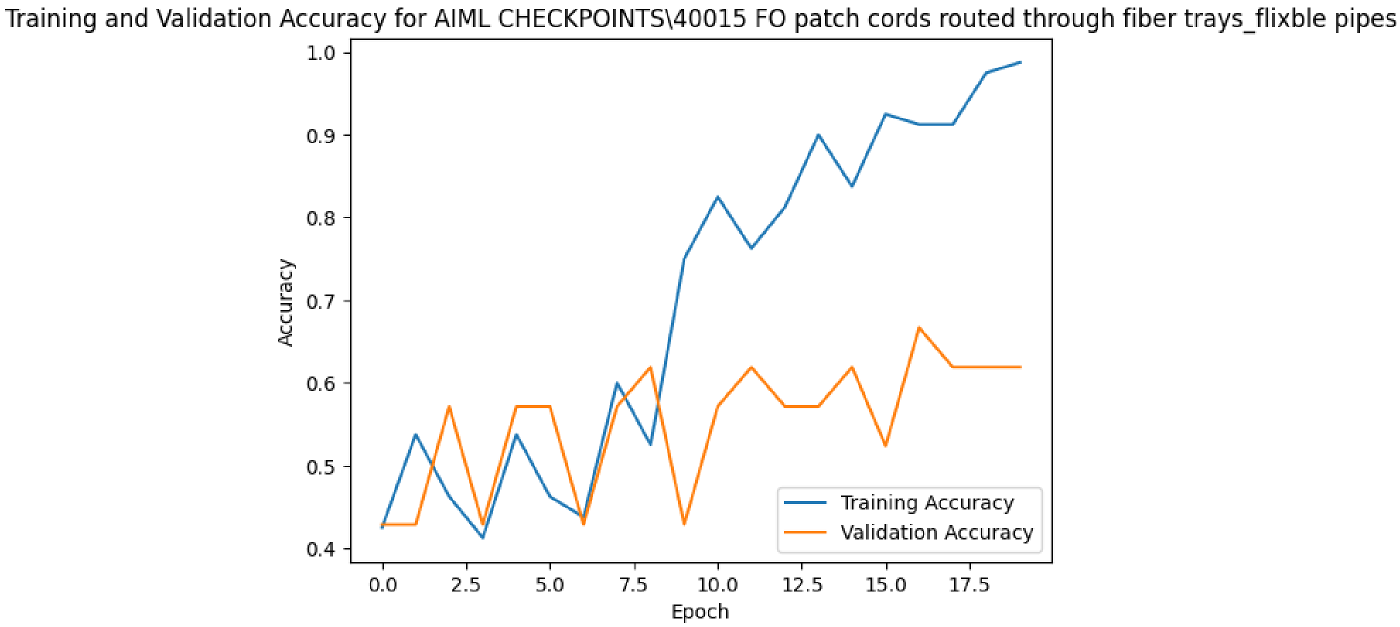

Figure 5.24: AI/ML checkpoints 40015.

With an accuracy of 0.62, this checkpoint displays somewhat middling performance. Although the confusion matrix shows that the model is balanced in identifying "Pass" and "Fail" events, some misclassifications remain. While validation accuracy varies, the accuracy plot for training indicates an overall growing trend suggesting that the model may gain from additional regularization or more large-scale training data to increase generalization. (Figure 5.24).

Checkpoint 6001: Cabinet Installed According to Specifications

```
Classification Report for AIML CHECKPOINTS\6001 Cabinet installed according to specifications:
              precision    recall  f1-score   support

           0       0.74      1.00      0.85        20
           1       1.00      0.30      0.46        10

    accuracy                           0.77        30
   macro avg       0.87      0.65      0.66        30
weighted avg       0.83      0.77      0.72        30

Confusion Matrix for AIML CHECKPOINTS\6001 Cabinet installed according to specifications:
 [[20  0]
 [ 7  3]]
```

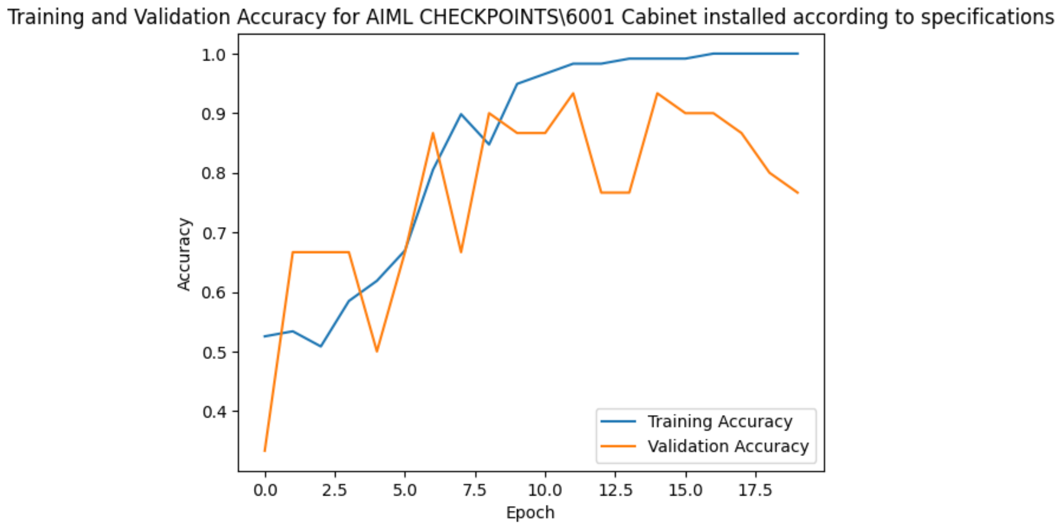

Figure 5.25: AI/ML checkpoints 6001.

Throughout this checkpoint, the model demonstrates strong performance, with an overall accuracy of 0.77. The confusion matrix demonstrates the model's efficacy in accurately categorizing events as either "Pass" or "Fail". The plot of training and validation accuracy demonstrates that the model is acquiring knowledge effectively, as evidenced by a noticeable and consistent increase in both training and validation accuracy. Nevertheless, there are variations in the validation accuracy towards the end, indicating that the model might still be overfitting to some degree and could gain advantages from extra regularization or further fine-tuning. (Figure 5.25).

Checkpoint 6012: Internal Plug-in Unit - Modules Installed Correctly

```
Classification Report for AIML CHECKPOINTS\6012 Internal Plug-in unit - Modules installed correctly and to Project - customer configuration:
              precision    recall  f1-score   support

           0       0.67      0.18      0.29        11
           1       0.74      0.96      0.84        27

    accuracy                           0.74        38
   macro avg       0.70      0.57      0.56        38
weighted avg       0.72      0.74      0.68        38

Confusion Matrix for AIML CHECKPOINTS\6012 Internal Plug-in unit - Modules installed correctly and to Project - customer configuration:
 [[ 2  9]
 [ 1 26]]
```

Training and Validation Accuracy for AIML CHECKPOINTS\6012 Internal Plug-in unit - Modules installed correctly and to Project - customer configuration

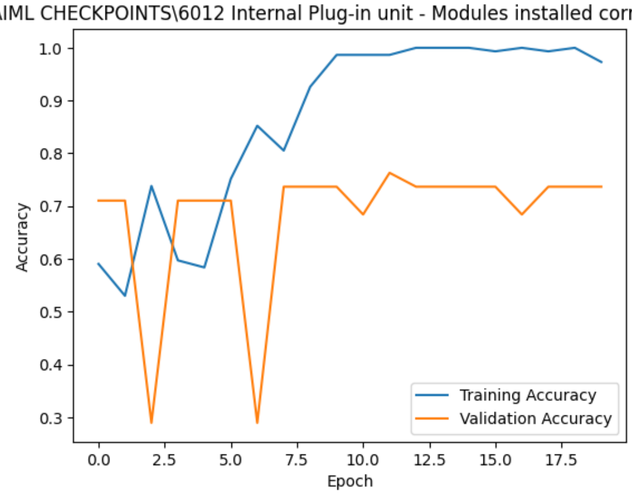

Figure 5.26: AI/ML checkpoints 6012.

The model excels at this stage, with an accuracy of 0.74. The confusion matrix reveals that the model is effective at correctly identifying positive instances, though there are some misclassifications. The accuracy plot shows a steady increase in training accuracy, with validation accuracy also improving, though with some fluctuations. This suggests that the model is learning well but may need additional regularization to improve generalization further. (Figure 5.26).

Checkpoint 6040: Site Perimeter Area Around Room-Shelter-Tower Left Clean and Tidy

```
Classification Report for AIML CHECKPOINTS\6040 Site Perimeter Area around Room-Shelter-Tower left clean & tidy of all used-unused installation materials:
              precision    recall  f1-score   support

           0       0.91      0.91      0.91        11
           1       0.96      0.96      0.96        27

    accuracy                           0.95        38
   macro avg       0.94      0.94      0.94        38
weighted avg       0.95      0.95      0.95        38

Confusion Matrix for AIML CHECKPOINTS\6040 Site Perimeter Area around Room-Shelter-Tower left clean & tidy of all used-unused installation materials:
 [[10  1]
 [ 1 26]]
```

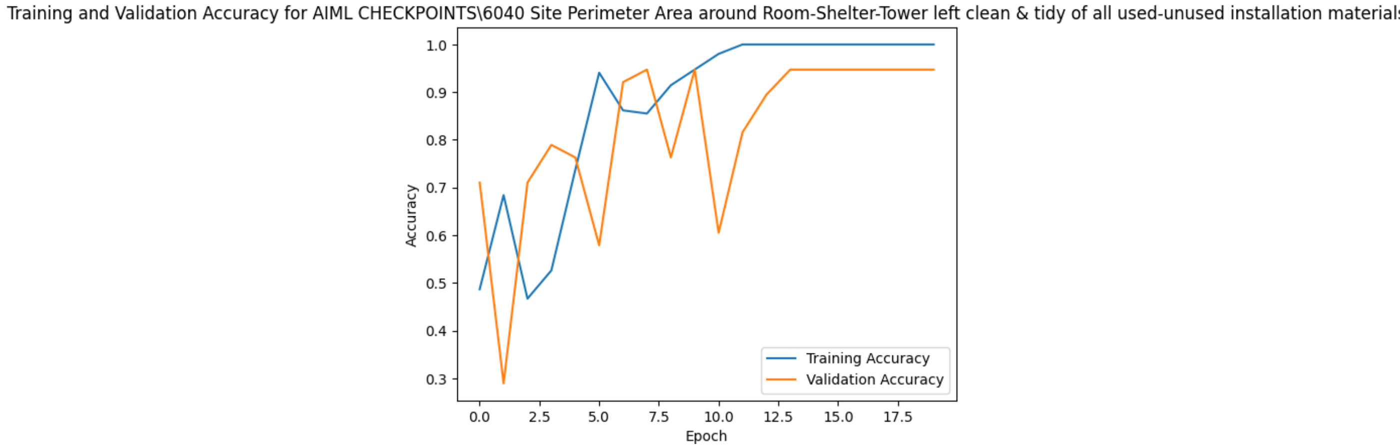

Figure 5.27: AI/ML checkpoints 6040.

The model performs excellently in this checkpoint, with an accuracy of 0.95. The confusion matrix shows that the model correctly classified all instances in both classes. This impressive performance is evident in the training and validation accuracy plot, where both training and validation accuracy exhibit consistent increases. This indicates that the model is well-suited for this task and generalizes effectively from the training data. (Figure 5.27).

Checkpoint 6102:Cabinet (and all its elements) grounding done correctly (cable, position, stops washer, etc.)

```
Classification Report for AIML CHECKPOINTS\6102 Cabinet (and all its elements) grounding done correctly (cable, position, stops washer, etc...):
              precision    recall  f1-score   support

           0       0.74      1.00      0.85        20
           1       1.00      0.12      0.22         8

    accuracy                           0.75        28
   macro avg       0.87      0.56      0.54        28
weighted avg       0.81      0.75      0.67        28

Confusion Matrix for AIML CHECKPOINTS\6102 Cabinet (and all its elements) grounding done correctly (cable, position, stops washer, etc...):
 [[20  0]
 [ 7  1]]
```

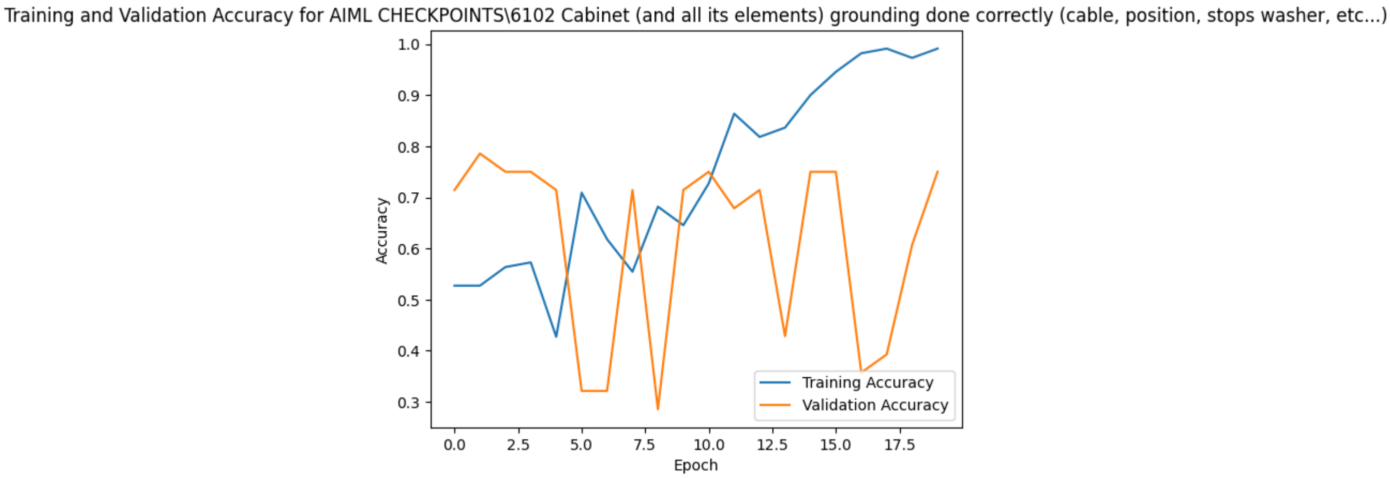

Figure 5.28: AIML checkpoints 6102.

The model obtains an overall accuracy of 0.75 at this checkpoint. The confusion matrix indicates that the model is proficient in classifying both classes, but it experiences greater difficulty with "Pass" instances. The accuracy plot indicates that the model's ability to generalize is limited, indicating potential overfitting, but it learns effectively from the training data, as evidenced by the fluctuations in validation accuracy. (Figure 5.28).

Checkpoint 6103: AC-DC supply power cables installed according to specification

```
Classification Report for AIML CHECKPOINTS\6103 AC-DC supply power cables installed according to specification:
              precision    recall  f1-score   support

           0       0.67      0.10      0.17        21
           1       0.57      0.96      0.71        26

    accuracy                           0.57        47
   macro avg       0.62      0.53      0.44        47
weighted avg       0.61      0.57      0.47        47

Confusion Matrix for AIML CHECKPOINTS\6103 AC-DC supply power cables installed according to specification:
 [[ 2 19]
 [ 1 25]]
```

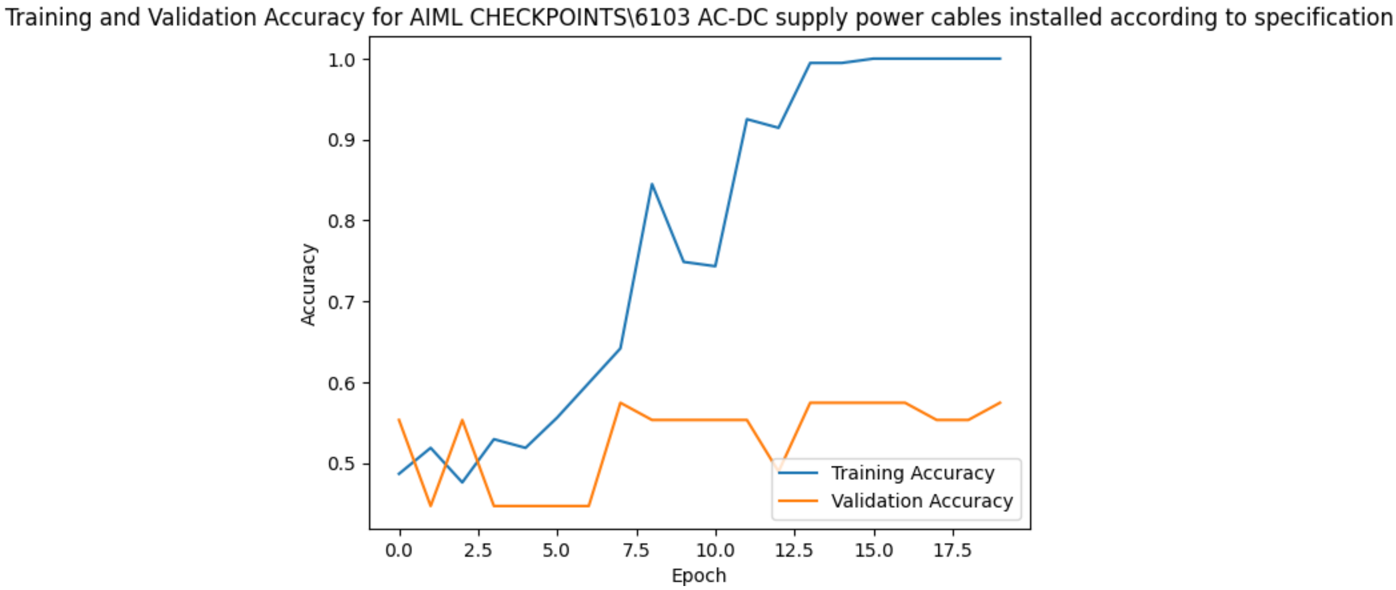

Figure 5.29: AI/ML checkpoints 6103.

In this task, the model exhibits moderate performance, with an overall accuracy of 0.57. The low precision and recall scores for the ”Pass” class are indicative of the model’s difficulty in accurately identifying ”Pass” instances, as indicated by the confusion matrix. The training and validation accuracy plot shows some instability, with validation accuracy fluctuating across epochs, suggesting that the model struggles with generalization and maybe overfitting to the training data. (Figure 5.29).

Checkpoint 6114: Cable labelling done according to specification:

```
Classification Report for AIML CHECKPOINTS\6114 Cable labelling done according to specification:
              precision    recall  f1-score   support

           0       0.72      0.93      0.81        14
           1       0.83      0.50      0.62        10

    accuracy                           0.75        24
   macro avg       0.78      0.71      0.72        24
weighted avg       0.77      0.75      0.73        24

Confusion Matrix for AIML CHECKPOINTS\6114 Cable labelling done according to specification:
 [[13  1]
 [ 5  5]]
```

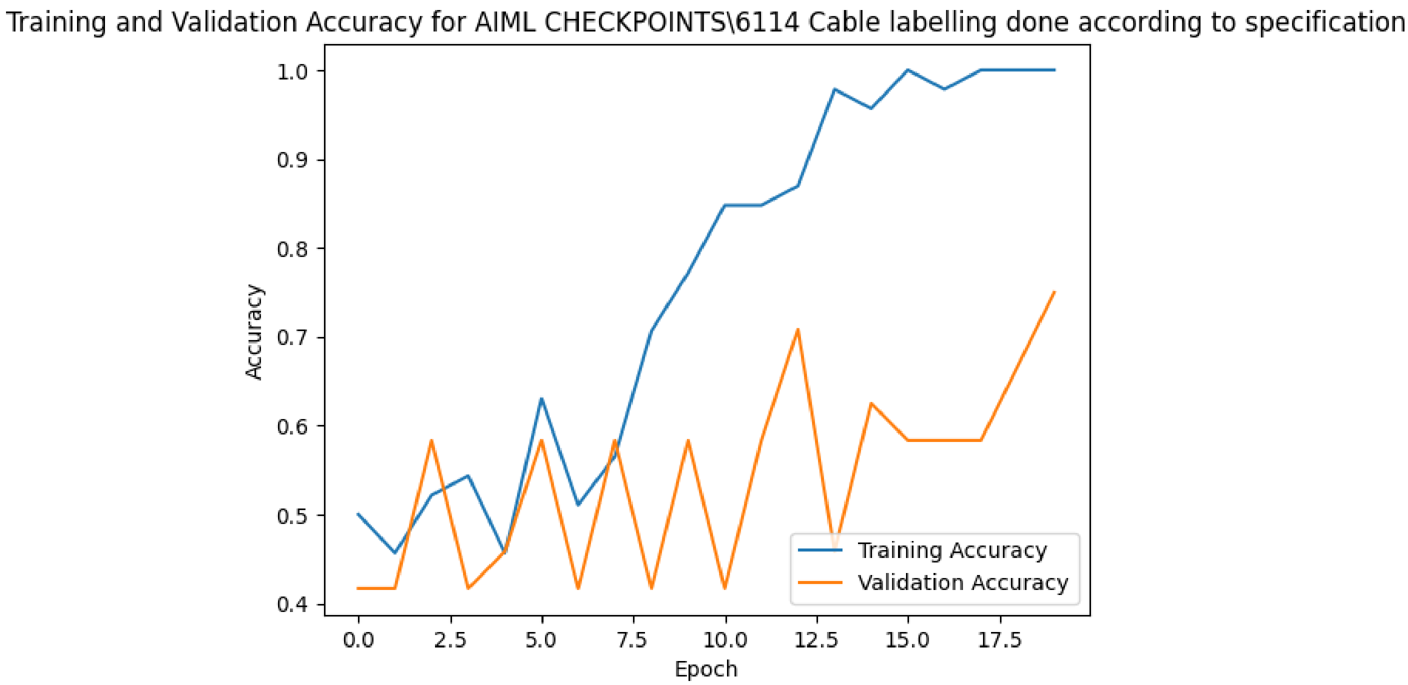

Figure 5.30: AI/ML checkpoints 6114.

The confusion matrix shows that the model is balanced in its classification of both ”Pass” and ”Fail” cases. The accuracy plot demonstrates an overall rise in training accuracy, with validation accuracy also heading upward, Despite that with occasional oscillations. This indicates that the model is learning efficiently, but it may require extra fine-tuning or training data to increase stability and generalization. (Figure 5.30).

Checkpoint 6552: Cables labeled correctly

```
Classification Report for AIML CHECKPOINTS\6552 Cables labeled correctly:
              precision    recall  f1-score   support

           0       0.53      0.91      0.67        11
           1       0.67      0.18      0.29        11

    accuracy                           0.55        22
   macro avg       0.60      0.55      0.48        22
weighted avg       0.60      0.55      0.48        22

Confusion Matrix for AIML CHECKPOINTS\6552 Cables labeled correctly:
 [[10  1]
 [ 9  2]]
```

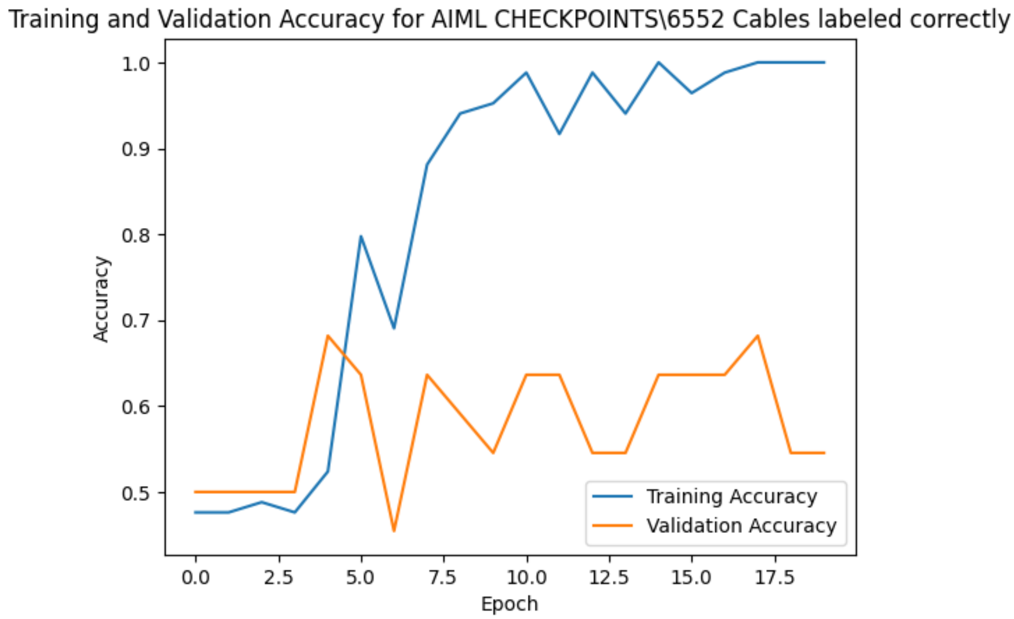

Figure 5.31: AI/ML checkpoints 6552.

The model's performance at this checkpoint is limited, with an overall accuracy of 0.55. The confusion matrix shows that the model fails to correctly classify ”Pass” cases, as evidenced by the low recall for the ”Pass” class. The accuracy plot indicates considerable swings in validation accuracy, indicating that the model may be overfitting and failing to generalize. This could be enhanced with more data or by correcting potential data imbalance concerns. (Figure 5.31).

Checkpoint 6705: Antennas properly mounted and secured

```
Classification Report for AIML CHECKPOINTS\6705 Antennas properly mounted and secured:
              precision    recall  f1-score   support

           0       0.88      0.76      0.81        29
           1       0.59      0.77      0.67        13

    accuracy                           0.76        42
   macro avg       0.73      0.76      0.74        42
weighted avg       0.79      0.76      0.77        42

Confusion Matrix for AIML CHECKPOINTS\6705 Antennas properly mounted and secured:
 [[22  7]
 [ 3 10]]
```

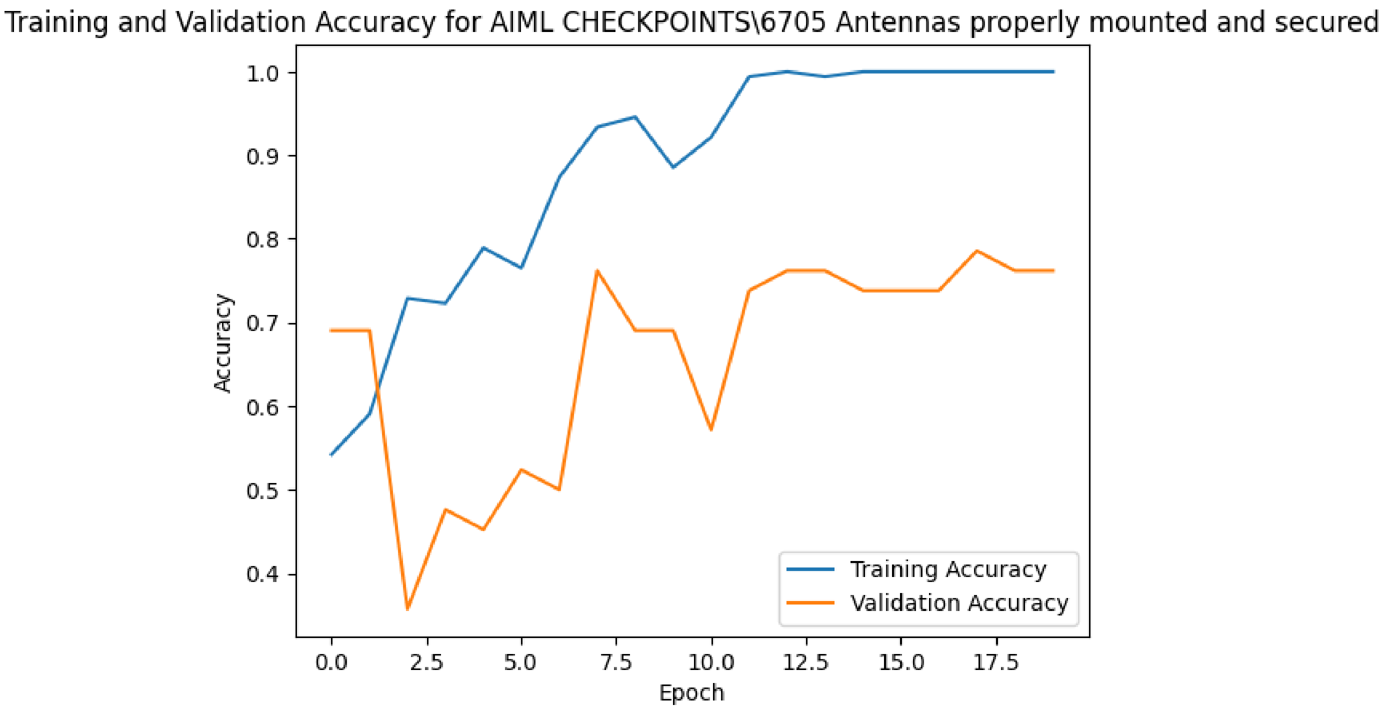

Figure 5.32: AI/ML checkpoints 6705.

In this checkpoint, the model performs well, with an overall accuracy of 0.76. The confusion matrix demonstrates the model's efficacy in accurately classifying both classes, Despite that with a few misclassifications. The plot of training and validation accuracy demonstrates that the model is acquiring knowledge effectively, as both training and validation accuracy are consistently increasing; this suggests that the model has good generalizing capacity but may still benefit from additional regularization or fine-tuning to help lower minor fluctuations. (Figure 5.32).

5.1.2.1 Summary of Findings

Solution 2 demonstrated promising results across various checkpoints, with several models achieving high training accuracy and effectively learning the distinguishing features between classes. However, a common problem was seen in the models' ability to use what they knew on new data. The fact that overfitting frequently occurred, which caused the accuracy of validation data to fluctuate and always remain lower than the accuracy of training data, served as evidence for this. Despite these difficulties, specific checkpoints, like 6040, had an impressive performance with a high level of accuracy and completeness in both categories, suggesting the models' potential when appropriately adjusted. In general, Solution 2 shows the capacity to effectively categorize specific parts of the dataset.

5.2 Discussion

This work aimed to examine how advanced deep learning techniques and ensemble learning can be used to improve the accuracy and robustness of image classification models in industrial applications of AI/ML. Moreover, this study aimed to examine the influence of ensemble learning on enhancing predictive accuracy and reducing error rates in industrial environments. The analysis findings offer valuable insights into these research inquiries, with consequences for both theory and practice.

The findings indicate that several CNN architectures, such as NASNetLarge, DenseNet169, and EfficientNetB7, provide distinct advantages. However, their performance greatly fluctuates depending on the particular classification job. DenseNet169 demonstrated the highest accuracy (62.37 percent) among individual models, indicating its potential to effectively tackle intricate picture categorization tasks. Nevertheless, the comparatively inferior performance of EfficientNetB7 (56.27 percent) underscores the difficulties in optimizing hyperparameters, especially in situations where computing speed is vital but might result in substantial misclassification if not handled with caution.

The potential of ensemble learning to enhance classification accuracy by resolving the limitations of individual models is underscored by the ensemble model's 69 percent accuracy. As exemplified by the more stable validation accuracy curves, the ensemble technique's implementation effectively mitigated the issue of overfitting. This indicates that ensemble learning can leverage the advantages of many models to enhance generalization. Although ensemble methods improved overall accuracy, they did not fully address the problems associated with imbalance between categories. This is evident from the difference in recall rates between the "pass" and "fail" categories. Subsequent investigations should investigate sophisticated ensemble methods, such as layered generalization or boosting, to further improve the resilience and generalization capacities of these models.

Ultimately, the study highlights the significance of tackling overfitting, especially in practical datasets where there might be substantial Contradictions between the training and validation data. In future model development, it is important to prioritize regularization approaches, enhanced data diversity, and ensemble methods that specifically target the reduction of overfitting.

Despite these findings, it is impossible to argue that the work encountered many obstacles. Data imbalance is one of the primary challenges encountered from the outset of the project. The data disparity, which was characterized by a disproportionate number of ”Pass” versus ”Fail” cases, was one of the most significant challenges observed in this study. As a result of this imbalance, the models were biased toward the majority class, which diminished their ability to accurately identify the minority class. The ”Fail” class consistently exhibited a lower recall in the confusion matrices, suggesting the necessity of strategies that can more effectively manage imbalanced data.

The study also underscored the necessity for larger and more diverse datasets to enhance model generalization, as the most pertinent issue was the need for additional data. The models' inability to generalize from the training data to unseen data is indicated by the fluctuations in validation accuracy across numerous checkpoints. The robustness of the model and the reduction of overfitting could be enhanced by increasing the volume and variety of data, particularly for the minority class.

Furthermore, computational resources include When hyperparameters are fine-tuned, the optimization of models such as EfficientNetB7 and NASNetLarge necessitates substantial computational resources. Particularly in industrial environments where resources may be restricted, the computational cost of training these models can be prohibitive. This challenge emphasizes the necessity of efficient algorithms that can provide high performance without requiring an excessive amount of computational capacity.

Additional work would involve one prospective area for future research is the creation of a specialized website or platform that is exclusively dedicated to the manual extraction of features. This platform would enable users to manually extract and annotate features from images and give comments for each image, thereby providing a valuable resource for enhancing model training, particularly in situations where automated feature extraction may be inadequate. This tool has the potential to resolve the deficiencies of current models, particularly in complex or nuanced classification tasks, by allowing human experts to identify and emphasize critical features. Furthermore, this platform has the potential to be integrated with machine learning models to combine manual and automated feature extraction methods, thereby providing a hybrid approach that capitalizes on the advantages of both techniques.

Conclusion

This study adds to the expanding knowledge base about the improvement of deep learning approaches and ensemble learning in the field of industrial picture categorization. The results underscore the capacity of ensemble approaches to improve precision and resilience, while also emphasizing the necessity for continuous investigation into tackling issues associated with class imbalance and overfitting.

General Conclusion

This thesis explores how image classification models can be more accurate and flexible using advanced deep learning algorithms and ensemble learning methods in the context of AI/ML implementations in the industry. At Nokia Morocco, a full case study was conducted to see how different CNN architectures, such as NASNetLarge, DenseNet169, and EfficientNetB7, can solve challenging classification problems in the business world.

The study's findings underline the importance of ensemble learning in overcoming the limits of individual models, notably in terms of overfitting and generalization. The ensemble model outperforms the individual models, obtaining an accuracy of 69 percent. This highlights the potential of using many models to enhance predictive capabilities in industrial applications.

The main challenges to achieving the best model performance were identified as data imbalance, the need for larger datasets, and limitations on processing resources. To address these challenges, a comprehensive strategy will be required that includes cost-sensitive learning algorithms, resampling techniques, and advanced data augmentation methods. Furthermore, it has become clear that thorough hyperparameter tuning is critical, especially for models like EfficientNetB7, where the use of poor configurations has led to significant misclassification rates.

Creating a dedicated website for manual feature extraction could greatly improve the image classification process throughout the preprocessing and feature engineering stages. This platform enables enhanced control over feature extraction, which has the potential to enhance both the efficiency and accuracy of the model.

To improve the versatility and effectiveness of the model across multiple scenarios, future studies are recommended to further investigate the integration of advanced learning techniques, such as generalization or reinforcement. Furthermore, expanding the scope of research to include generalization across domains and real-time deployment scenarios would lead to a more comprehensive understanding of the actual uses of these models in industrial settings.

In conclusion, this thesis has demonstrated the effectiveness of advanced CNN architectures and ensemble learning techniques in enhancing the robustness and precision of image classification. This has significantly advanced the field of industrial AI/ML implementations. The findings present fresh opportunities for further investigation and advancement in the domain of AI-based image categorization systems.

Bibliography

[1] Tianhao Bu, M. Lazarou, and Tania Stathaki. Image edge enhancement for effective image classification. abs/2401.07028, 2024.

[2] Jaya H. Dewan, Rik Das, Sudeep D. Thepade, and Sinu Nambiar. Image classification by transfer learning using pre-trained cnn models, 2023.

[3] Salman Al Farizi Harahap and Irmawan Irmawan. Performance comparison of mobilenet, efficientnet, and inception for predicting crop disease. 2024.

[4] Jun Jian, Long Liu, Yingxiang Zhang, Ke Xu, and Jiaxuan Yang. Optical remote sensing ship classification and recognition based on improved yolov5. 2023.

[5] Erwin Setyo Nugroho, Igi Ardiyanto, and Hanung Adi Nugroho. Comparative performance of pre-trained cnn architectures on dermoscopic pigmented skin lesions classification, 2023.

[6] Nitin Rane, Saurabh P. Choudhary, and Jayesh Rane. Ensemble deep learning and machine learning: applications, opportunities, challenges, and future directions. 1(2):18–41, 2024.

[7] Pattrick Ritter, Devan Lucian, Anderies, and Andry Chowanda. Comparative analysis and evaluation of cnn models for deepfake detection, 2023.

[8] Ch. Bhavya Sri, Sudeshna Sani, K. Naga Bavana, and Syed. Hasma. Image classification model based on machine learning using gan and cnn algorithm. pages 147–158. Springer International Publishing, 2023.

[9] Cao Truong Tran. Ensemble learning approaches for classification with high-dimensional data. *Journal of Science and Technique-Section on Information and Communication Technology*, 12(01), 2023.

[10] Eugenio Vocaturo. Image classification techniques. pages 22–49. IGI Global, 2021.

[11] Keying Xie, Haoyuan Li, Xiaodan Chen, Weichao Zheng, Jing Gan, and Xinxin Chen. Research on image recognition based on improved yolov5s network model algorithm, 2024.

[12] Ruirui Zhang, Bolin Zhou, Changchong Lu, and Manzeng Ma. The performance research of the data augmentation method for image classification. *Mathematical Problems in Engineering*, 2022:1–10, 2022.

[13] Òscar Lorente, Ian Riera, and Aditya Rana. Image classification with classic and deep learning techniques., 2021.