The European Master in Building Information Modelling is a joint initiative of:

Universidade do Minho

UNIVERZA
V LJUBLJANI

AI integrated in Construction BIM 4D Planning, Progress Tracking and Reporting

Daria Udalova

BIM A+

UMinho | 2025

Universidade do Minho
Escola de Engenharia

Daria Udalova

AI integrated in Construction BIM 4D Planning, Progress Tracking and Reporting

BIM A+
European Master in
Building Information Modelling

September 2025

Universidade do Minho
Escola de Engenharia

Daria Udalova

AI integrated in Construction BIM 4D Planning, Progress Tracking and Reporting

BIM A+ European Master in Building Information Modelling

Master Dissertation
European Master in Building Information Modelling

Work conducted under supervision of:
José Carlos Basto Lino
Manuel Afonso Parente

September, 2025

AUTHORSHIP RIGHTS AND CONDITIONS OF USE OF THE WORK BY THIRD PARTIES

ACKNOWLEDGEMENTS

I would like to begin this acknowledgment by expressing my gratitude to all the professors and partners of the master's, who, with their knowledge, dedication, and guidance, made the development of this work possible. Each class, discussion, and advice received throughout this course was fundamental to my academic and professional growth.

I would especially like to thank José Carlos Lino, who allowed me to do this research, offering valuable insights and dedicated support. His guidance was essential for this work to materialize. Additionally, I would like to thank Manuel Parente for his understanding, support, and professional and academic assistance. Namely, I would like to thank: Miguel Azenha, José Granja, Tomo Cerovšek, Vlatko Bosiljkov, Žiga Turk, Maria Laura Leonardi.

I would like to express my gratitude to BIMMS for its contribution to the digital development of the construction industry, for the topic of this dissertation, the data provided for research, advice and assistance in writing it, where the collaboration of Bruno Caires and Luis Lima was vital for ensuring the practical relevance of this research.

I would like to thank my colleagues Faruk Gokce Basaran, Rustam Doschanov, Dmitriy Bitkin for showing me the importance of Building Information Modeling and 4D modeling, for giving me the opportunity to grow as a professional and to be involved in large-scale projects.

I would like to thank my family, particularly my parents - Mikhail Udalov and Tatiana Udalova, who always supported me in all my intensions. Without your support and encouragement, I would not be the one who I am now.

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration.

I further declare that I have fully acknowledged the Code of Ethical Conduct of the University of Minho.

RESUMO

A indústria de Arquitetura, Engenharia e Construção (AEC) tem passado por uma fase ativa de transformação digital nos últimos anos, após o avanço do Building Information Modeling (BIM). A integração da dimensão tempo, o 4D, no BIM expande o seu potencial, permitindo não apenas a visualização de sequências de construção, mas também um planeamento mais eficiente, alocação de recursos, soluções de engenharia e gestão de logística. A aplicação do BIM 4D na análise de projetos em tempo real e na monitorização automatizada ainda é limitada, apesar do seu impacto comprovado na melhoria da execução e coordenação de projetos. A integração de Inteligência Artificial (IA), Aprendizagem de Máquina (ML) e Visão Computacional (CV) pode fundamentalmente remodelar e desenvolver a forma como o progresso da construção é rastreado e reportado na indústria AEC.

A dissertação aborda esta lacuna de investigação desenvolvendo uma metodologia para incorporar abordagens baseadas em IA em fluxos de trabalho BIM 4D com o objetivo de automatizar a monitorização do progresso, melhorar a precisão do cronograma e produzir mecanismos de relatórios objetivos. A base empírica do estudo baseia-se em dados fornecidos por BIMMS, recolhidos de quatro projetos de construção na Europa, onde foram obtidas imagens panorâmicas de 360 graus. Estas imagens documentaram várias fases da obra e serviram de fonte para a construção de um conjunto de dados dedicado de 600 fotografias de instalações de divisórias em gesso cartonado. O conjunto de dados foi classificado em três categorias distintas, correspondentes a diferentes fases de progresso, desde a instalação da estrutura até à conclusão do acabamento da superfície.

O trabalho experimental principal envolveu o desenvolvimento e teste de uma Rede Neural Convolucional (CNN), projetada para classificar as fases de instalação de paredes diretamente a partir de imagens de construção do mundo real. O modelo foi treinado em condições desafiadoras do mundo real, incluindo iluminação variável, oclusões, ruído de construção e ângulos e pontos de vista instáveis da câmara. Através de ciclos iterativos de treino, análise e avaliação, a CNN alcançou uma precisão robusta na classificação das fases de construção de uma parede de gesso cartonado e demonstrou resiliência às complexidades típicas dos dados de um estaleiro de construção.

A contribuição desta dissertação reside na integração de aprendizagem de máquina e BIM 4D para a monitorização do progresso em ambientes de construção do mundo real. Os resultados do estudo comprovam o conceito de que a classificação baseada em CNN pode fornecer informações automatizadas e fiáveis sobre o estado do progresso da construção, reduzindo significativamente a dependência de inspeções e avaliações manuais. Com a integração dos resultados da classificação automatizada no projeto, o estudo introduz um enquadramento para permitir a comparação sistemática entre os prazos planeados e os prazos reais dos cronogramas. Isso, por sua vez, apoia a formação de relatórios baseados em evidências, o que aumenta a transparência, fortalece a comunicação entre as partes interessadas e facilita a tomada de decisões proativa. A relevância prática dessa integração é evidente no potencial de adoção do enquadramento proposto nas plataformas digitais existentes para melhorar a produtividade, simplificar os relatórios e minimizar atrasos.

Keywords: 4D, Inteligência Artificial (IA), Machine Learning, BIM

ABSTRACT

Architecture, Engineering, and Construction (AEC) industry has been undergoing an active phase of digital transformation, for the past years after advancement of Building Information Modeling (BIM). The integration of the time dimension 4D into BIM extends its potential, enabling not only visualization of construction sequences but also more efficient scheduling, resource allocation, engineering solutions and logistics management. Application of 4D BIM in real-time project analysis and automated monitoring are still limited in use regardless of its proven impact in improving project delivery and coordination. The integration of Artificial Intelligence (AI), Machine Learning (ML) and Computer Vision (CV) can move the AEC industry to fundamentally reshape and develop the way construction progress is tracked and reported.

The dissertation addresses this research gap by developing a methodology for embedding AI-driven approaches into 4D BIM workflows with the goal of automating progress monitoring, enhancing scheduling accuracy and producing objective reporting mechanisms. The empirical foundation of the study is based on data provided by BIMMS, collected from four construction projects in Europe, where 360-degree panoramic imagery was obtained. These images documented various phases of construction work and served as a source for building a dedicated dataset of 600 photographs of drywall partition installations. The dataset was classified into three distinct categories corresponding to different progress stages, ranging from the installation of framing to the completion of surface finishing.

The core experimental work involved the development and testing of a Convolutional Neural Network (CNN) designed to classify wall installation stages directly from real-world construction imagery. The model was trained under challenging real-world conditions, including varying lighting, occlusions, construction visual noise and unstable camera view angles and viewpoints. Through iterative cycles of training, analysis and evaluation, CNN achieved robust accuracy in classifying construction stages of a drywall and demonstrated resilience to the typical complexities of construction site data.

The contribution of this dissertation lies in the integration of machine learning and 4D BIM for progress monitoring in real-world construction environments. The results of the study prove the concept that CNN-based classification can provide reliable, automated insights into the state of construction progress, significantly reducing dependence on manual inspections and assessments. With integrating automated classification outputs into project, the study introduces a framework for enabling systematic comparison between planned and actual timelines of schedules. This in turn supports the evidence-based reports formation which enhances transparency, strengthens stakeholder communication and facilitates proactive decision-making. The practical relevance of that integration is evident in the potential for adopting the proposed framework within existing digital platforms to improve productivity, streamline reporting, and minimize delays.

Keywords: 4D, Artificial Intelligence (AI), Machine Learning (ML), BIM

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

2/3/4/5D	2/3/4/5 Dimension
AEC	Architectural, Engineering and Construction
AECO	Architectural, Engineering, Construction and Operations
AI	Artificial Intelligence
AR	Augmented Reality
BAM	Building Automation and Management System
BCPD	Bayesian Coherent Point Drift
BIM	Building Information Modelling
BOMs	Bill of Materials
CAD	Computer-Aided Design
CNNs	Convolutional Neural Networks
CRF	Conditional Random Field
CV	Computer Vision
DL	Deep Learning
DPT	Dynamic Process Template
FFNN	Feedforward Neural Network
GA	Genetic Algorithm
GAN	Generative Adversarial Networks
GF	Guided Filtering
GPU	Graphics Processing Unit
HVAC	Heating, Ventilation, and Air Conditioning
HSV	Hue, Saturation, Value
IEQ	Indoor Environmental Quality
IoT	Internet of Things
IoU	Insertion over Union
LSTM	Long Short-Term Memory
MEP	Mechanical, Electrical and Plumbing
ML	Machine Learning
mAP	mean Average Precision
MILP	Mixed-Integer Linear Programming
NLP	Natural Language Processing
O&M	Operation and Maintenance
PRG	Precedence Relationship Graph
SfM	Structure from Motion
SLAM	Simultaneous Localization and Mapping
UAS	Unmanned Aerial System
UAV	Unmanned Aerial Vehicle
VDC	Virtual Data Center
VR	Virtual Reality

1. INTRODUCTION

1.1. Framing and Motivation

The digital transformation of the Architecture, Engineering and Construction (AEC) industry is constantly developing and transforming, with innovative technologies appearing upon the foundations of Building Information Modeling (BIM). While the 3D BIM models aim to provide digital and information representation of a project, the 4D BIM model takes a step further by integrating time dimension into it. By connecting digital model to a timeline 4D BIM model becomes a base for deeper analysis and allows stakeholders to visualize construction process and understand clearly how the project will be built over time step by step. This represents several benefits: improved scheduling and planning, helps with resource and logistics management by ensuring materials, assembled structures, vehicles arrive precisely when they are supposed to be used.

Research shows that using 4D BIM can reduce construction delays, increase labor productivity, and optimize resource planning. In a case study in central Jakarta, 4D BIM helped reduce the construction duration from 511 to 429 days (Moeisra et al., 2023). It is also an indispensable tool for managing logistics and reducing rework. The article by Fazeli et al. (2022) describes a 4D BIM prototype that uses optimization algorithms for automated time calculation and improved planning accuracy. Other studies use BIM-based methods for detecting and resolving spatial-temporal clashes in complex projects (Yin et al., 2024), which is novel approach to use 4D modeling.

One of the biggest aims of 4D modeling is spatial-temporal and logical clash detection. While 4D BIM model enables visualizing construction sequence, the ability to predict, find out and mitigate conflicts remains limited without automation or coding. The integration of Artificial Intelligence (AI) into 4D BIM modeling may appear a promising solution providing enabling real-time clash detection, smart scheduling and reporting, and data-driven analytics.

AI can process and work with big data, which is human unreachable or time-consuming, to automate tasks, predict outcomes, and optimize the workflow. For example, applying AI in BIM technology helps to achieve multi-stage classification of objects on a construction site, which results in improvement of project management, procurement and safety (Dolhopolov et al., 2024). Furthermore, AI is useful in predictive analytics to recognize and assess potential risks and delays, as well as to optimize complex tasks, which contribute to sustainable development and decision-making (Ajirotutu et al., 2024). At the same time, Glinka S. (2024) proposed using AI to combine satellite remote sensing data with 4D BIM models to monitor construction progress which author considers useful for largescale infrastructure projects. Synergistic use of 4D BIM and AI potentially may create a comprehensive and proactive approach to improve project management and tracking of construction that can overcome and reduce challenges.

This dissertation aims to explore how AI-based methodologies can be embedded into 4D BIM modeling to create a more adaptive, efficient, and automated construction planning system. By implementing machine learning algorithms, this study seeks to recognize stages of wall installation, ultimately contributing to optimizing scheduling, cost efficiency, better decision-making and reduced project delays.

1.2. Objectives

This dissertation aims to explore and demonstrate the capabilities of machine learning (ML) in automating construction progress monitoring. Specifically, it seeks to develop a methodology for identifying the construction phase of structural elements directly from real-world images (screenshots). This research is motivated by the need to bridge the gap between theoretical benefits of digital construction and practical, automated site monitoring, thereby enhancing project management and cost efficiency and accuracy.

In pursuit of this goal, the dissertation has several specific objectives. Initially, this research will conduct a comprehensive literature review to systematically organize existing knowledge pertaining to image-based construction progress monitoring, AI implementation in construction and scheduling, ML applications in the AEC sector, and the integration of visual data with BIM and 4D scheduling.

This review will serve to identify innovative approaches in applying ML, Internet of Things (IoT) and visual technologies in construction, and using 4D modeling in project management and design relevant to the automated detection and classification of construction phases from visual inputs.

The subsequent objective involves the development and validation of a robust ML model, specifically designed to accurately classify the current installation phase of architectural component (plasterboard wall) as depicted in real-world 360-degree footage. The steps include collection, classification, editing and annotation dataset of screenshots from real-world projects, access to which is provided by the dissertation partner company - BIMMS, and after training of computer vision algorithm. BIMMS is an international company specializing in developing and managing digital models and solutions for construction industry. Large-scale projects and automatization of the workflows are the specialization of the company, so provided guidance and data was crucial to understand what the current practices are and how they can be improved.

Finally, this study aims to establish a practical framework for integrating the automated stage classification results with existing project schedules. This integration will enable the generation of automated progress or completeness of work reports, attaching with real-captured and systematic evidence of installation or other process stage for stakeholders or clients. Furthermore, the framework will facilitate a direct comparison between captured progress and planned timelines, allowing the identification of delays or accelerations and supporting proactive decision-making in project management, logistic and procurement.

1.3. Methodology

The methodology to achieve these objectives is following with a comprehensive review of existing literature resources. Review incorporates various source types, including books, to establish a broad understanding of the subject; and scholarly articles, to ensure the information is current and clearly supported by empirical findings. Flexible and inclusive approaches were chosen for sources searching, in line with the exploratory nature of the dissertation. Databases, in particular ScienceDirect, Google Scholar, and ResearchGate were used to search for peer-reviewed papers and articles, through specific terms and key words as "Artificial Intelligence", “AI”, “BIM”, "4D", "planning", "Machine Learning", "ML", "construction", "scheduling" with a focus on the date of publication not older than 2015, and published in journals of high reputation.

The second part of the data collection involved working in close collaboration with BIMMS company, through the openspace.ai platform. 360-degree footage was obtained from four real construction sites in Europe at various stages of completion. By examining all available paths and dates of capture, screenshots of the architectural elements of interest were extracted. The screenshots were taken from different angles, locations, and in varying lighting and noise conditions to ensure data diversity. The photos were manually classified by their construction phase and cropped to isolate the elements where necessary. This data collection phase was designed to create a representative dataset for training and testing the ML model.

Finally, to develop and implement the ML model, the existing code from the publicly available source tensorflow.org was utilized. This code was adjusted to achieve the target accuracy and ensure the model's reliability under various conditions. After each adjustment, the model was tested, and, if necessary, the testing data was refined. The methodology described here is illustrated below (Figure 1).

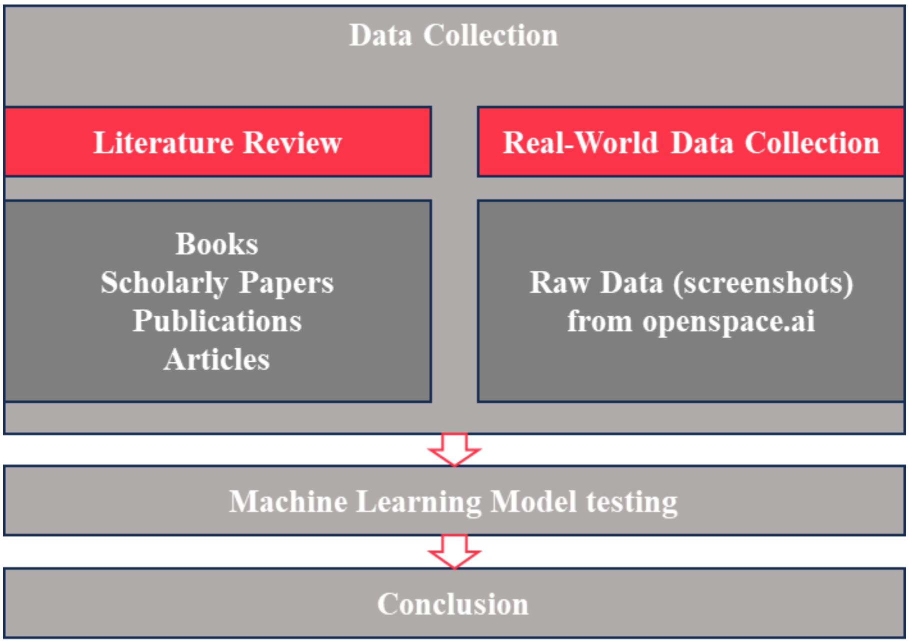

Figure 1 – Framework of the Dissertation

1.4. Guide Structure

Current study explores the key components of creating a ML model for monitoring construction progress is organized into five chapters. The first chapter serves as an overview of the dissertation central theme. In the beginning the obstacle of reconnecting theoretical project schedules with actual on-site progress is outlined, and it highlights how artificial intelligence can be utilized to address this discrepancy. Expected aims and general principles of the research, defining its objectives and overall direction are also lays out this section. The chapter concludes by introducing the methodology that was followed and providing a concise summary of the study overall structure.

Chapter two reviews the theoretical foundations and key concepts that frame the integration of Artificial Intelligence into 4D BIM for construction progress monitoring. The chapter provides a structured overview of previous research on BIM-based planning, AI and computer vision applications, as well as the role of neural networks in automating construction tasks. It further discusses existing approaches to visual monitoring and highlights the opportunities and limitations of current methods.

Chapter three describes the methodology used for collecting empirical data from real-world construction projects. The chapter outlines the process of navigating and extracting 360-degree imagery, explains the rationale for selecting drywall partitions as the primary object of study, and details the classification of construction progress. Furthermore, it discusses development of the dataset, including the limitations.

Chapter four presents the testing of machine learning CNN model developed for the classification of construction progress stages based on real-world imagery. The chapter highlights the outcomes of several training attempts with progressively expanded datasets, augmentation techniques and code architecture, showing how model accuracy and robustness improved over attempt. Discussion part reflects upon the challenges that had been faced during the modelling process, and provides critical insights into the experimental findings, setting the foundation for the conclusions, study limitations and recommendations for further research.

Chapter five summarizes the main conclusions of the study, reflecting on the outcomes of machine learning model testing for drywall installation progress stages classification. Chapter critically reviews the limitations of the research, in particular issues related to dataset size, variability of real-world conditions and generalizability of results. Finally, the chapter provides recommendations for future research, suggesting directions for improving model performance, expanding and improving datasets size and quality, improving and further integrating AI into 4D BIM-based workflows.

2. LITERATURE REVIEW

2.1. 4D BIM as a Foundation for Construction Planning, Optimization and Visualization

Building Information Modeling (BIM) is not only about digital representation of projects but a complex multidimensional management tool. With the extension of BIM from third dimension into fourth, with adding time, results in 4D BIM, which connects and enriches digital models to scheduling and enables stakeholders to simulate and visualize construction sequences step by step, and analyze construction sequences virtually, enabling better decision-making and coordination throughout the project lifecycle. As highlighted in the book by Daniotti et al. (2020), "by connecting 3D geometric models to the schedule data it is possible to obtain a 4D aspect that is able to provide several efficient advantages in order to facilitate site planning and management." Thus, according to BIM dictionary 4D BIM model determined as - '… a model or a modelling workflow is considered to be 4D when the time is added to model objects to allow Construction Scheduling'. This integration transforms static model representation into dynamic simulations that support planning, communication and optimization.

4D BIM can be implemented using different tools and depending on the goals and capabilities of the company. The research by Stéphania Sanon and Conrad Boton (2024) provides a detailed evaluation of three leading 4D BIM simulation tools: Navisworks Manage, Synchro 4D Pro, and Fuzor. The study aims to assess their effectiveness, usability, and feature sets across four categories: 4D features, collaboration, 3D capabilities, and planning functionalities. Evaluation criteria were drawn from literature and practical needs. In terms of 4D features, Fuzor leads with advanced animation capabilities, including realistic construction growth simulations and equipment movement. Navisworks and Synchro also support animations, but with more manual processes and limitations in equipment handling. Fuzor and Navisworks offer both manual and automatic linking of 3D elements to schedules, while Synchro only supports manual linking, which can be time-consuming for large projects. Collaboration features are strong across all tools. Regarding 3D features, Navisworks and Synchro are limited to model import and basic manipulation, whereas Fuzor allows full model creation, material editing, and realistic rendering. Fuzor also supports bidirectional updates with Revit, enhancing model synchronization. Planning functionalities show Synchro and Fuzor as more capable than Navisworks. Synchro allows dynamic schedule editing with immediate visual feedback, while Fuzor supports detailed task types and simulation control. Navisworks lacks network analysis tools and relies on pre-established schedules for effective use. The final comparison shows Fuzor as the most innovative and capable tool for realistic construction simulations, offering superior features across all categories. Despite benefits of 4D BIM regardless of a tool adoption of it remains limited due to technical challenges, user experience gaps and low client demand. As Daniotti et al. (2020) examine in their book the challenges and opportunities of implementing modern technologies in the construction industry, especially in the context of skill development and training, which is a critical factor for successful digital transformation.

The development of 4D models supports continuous progress tracking, as noted by Daniotti et al. (2020): "The 4D and 5D dimensions model development allows to update and monitor the construction phase progress." The argument that 4D BIM is not just a visualization, but a management tool to improve planning, constructability analysis and stakeholder communication, is proven by several studies, confirming that 4D visualization approach leads to mitigating delays and optimizing construction workflows. For example, a case study by Moeisra et al. (2023) in a real-world project in central Jakarta showed that the use of 4D BIM helped reduce construction duration from 511 to 429 days demonstrates 4D application effectiveness in solving planning problems and substantial execution delays caused by various factors such as labor inefficiencies, adverse weather conditions, project revisions, and the COVID-19 pandemic. The same result - two-month reduction in the overall project timeline, which helped to win the contract, improvement of communication across teams through daily progress monitoring with automated production control charts, because of replacing traditional Gantt charts was shown in the article by Iordanova et al. (2020). Authors present a case study of a mega-hospital project in a dense urban environment, where 4D BIM was integrated to optimize the project schedule. They describe the digital artefact that combines location-based planning, takt-time scheduling (which "is a powerful Lean strategy for flow stabilization") and 4D simulation, which they develop and implement to manage the complex demolition and construction phases in that project. To short the timeline, design and site team used BIM to model both existing and new structures with location breakdown structure (LBS), enabling precise coordination, additionally, a takt-time plan adapted to the constraints of subcontractors and site logistics was linked with detailed 4D simulation to visualize spatial-temporal relationships between demolition, excavation and other construction activities, and, in the end, developed Unity-based platform that connected BIM model, master schedule, takt-time plan and project documentation, including 2D drawings and virtual reality environments, to enable real-time updates and made scheduling information accessible to all stakeholders in formats suited to their roles.

4D BIM with integrated BIM/VDC/Lean platform provides a dynamic, visual and intuitive representation of the construction sequence results in real-time progress tracking and enhancing planning accuracy, early detection of scheduling conflicts and improved communication and fostered collaboration among stakeholders, which enhancing decision-making and operational efficiency. The findings of the case studies underscore the potential of BIM and 4D BIM to transform planning development by making it more efficient, predictable and resilient to disruptions. In turn, it contributes to a Lean culture on-site, and future adaptations could tailor the system to different project types and stakeholder maturity levels, reinforcing its value as both practical and theoretical contribution to construction management.

4D BIM helps not only in meeting, optimizing and monitoring deadlines, but also it is indispensable tool in logistics, resource management and procurement. In their work, Khondoker et al. (2024) illustrate how integrating 4D BIM with optimization methods helps improve the procurement of steel bars, for the site with limited storage space, which helps reduce costs, over stockpiling and prevent delays in deliveries. BIM was used to extract material quantities and link them to the construction schedule to visualize procurement needs and anticipate when and how much steel to order with the help of programming. Created procurement plan resulted in minimizing costs, orders excess prevention, and efficient vertical space storage using a double-sided cantilever rack system for different bar sizes. Authors note that enables precise scheduling based on crew productivity and activity durations “the unit price of rebar throughout the construction period greatly influences its procurement procedure,” and forecasting allows for more informed decisions, ensures uninterrupted construction while minimizing rental costs. It “partially satisfies a long-sought research need for establishing a comprehensive construction steel bar procurement system” and offers practical benefits for both construction and production managers. Authors highlight that supplier integration and digital planning tools can enhance competitiveness in the construction industry.

Spatial clashes are often mentioned in the context of BIM, for example, when a pipe crosses a beam. This is visually visible and easy to check already at the 3D modeling stage. Beyond visualization, 4D BIM introduces a deeper layer of construction logic by enabling detection of time-based clashes, referred to in the book by Danotti et al. (2020) as “workflow clash”. This concept identifies instances where two components might be scheduled to overlap during assembly, potentially causing logistical conflicts even when spatial arrangements seem valid. Workflow clash is a time conflict, when two elements or processes are scheduled for the same time, although logically they should be sequential, which may go unnoticed at the planning stage without using 4D BIM. For example, plastering a wall is scheduled before installing drywall - this is not a geometric error, but an error in the logic of the process. Traditional manual methods to find spatiotemporal clashes are inefficient and prone to oversight 4D BIM allows for the prevention of conflicts before work even begins playing a key role in ensuring safety and planning efficiency, especially through automated detection, time calculations and tasks. In the paper, Yin et al. (2024) proposed a 4D BIM-based method for detecting and optimizing spatiotemporal clashes in underground pipeline construction because of the complexity of underground pipeline relocation, which often involves multiple disciplines and phases. The 4D detection and an optimization model based on mixed-integer linear programming (MILP) approach significantly improved planning efficiency and reduced risks and spatial-temporal clashes from 45.5% to 0% in a real-world project. Developed 4D BIM model enabled dynamic simulation and identification of temporal overlaps between pipeline activities. Optimization model assigned tasks to workers, sequenced pipeline sections, and applied constraints to avoid scheduling conflicts. In turn, the article by Fazeli et al. (2022) describes a 4D BIM prototype that uses optimization algorithms to automatically calculate time and improve planning accuracy. Their proposed solution leverages resource specification techniques and ML to streamline the scheduling process. The prototype was validated using two real-life residential projects in Tehran, and the Genetic Algorithm (GA)-based model successfully calculated the project completion date.

There is space for future research in both cases with bigger datasets and more complex and broader projects. The studies contribute significantly to the construction field by offering standardized, automated and scalable solutions for time estimation, reducing human error, enhancing constructability, and improving project planning and safety efficiency. Such application extends the role of BIM from visual planning to process validation, reinforcing its value as a decision support system. According to Danotti et al. (2020), BIM has “innovative potential allows stakeholders to exploit different project levels”, where 4D is positioned as a tool that not only enhances site management but aligns with emerging digital standards in construction. This reinforces its growing importance in both operational practice and academic discourse around transformation of the built environment.

While 4D BIM establishes a robust base for planning and optimization, its potential multiplies when combined with artificial intelligence. The next section examines how AI and CV enhance construction monitoring and automation.

2.2. Integrating Artificial Intelligence into Construction

Modern construction field is increasingly moving towards digital, automated and sustainable practices. In this context, the integration of AI and BIM is becoming a key development area. Artificial Intelligence (AI) transforms construction by automating tasks, processing vast datasets, and enabling predictive analytics. Zhang et al. (2024) define it 'intelligent construction' - 'as a system integrating software, hardware, seamlessly networked sensors, cloud computing platforms and computational elements, which is geared towards automation, data-driven technologies, and advanced digital applications', what implies the shift from traditional labor-intensive and manual methods to smart, resilient and sustainable practices. In their study they systematically reviewed the current state, implementation strategies, key technologies and future directions. Authors went through over 8900 publications with the aim to bridge the digital gap in the construction industry, system knowledge, identifying trends, challenges and opportunities for innovation in the field. Through bibliometric analysis, the study reveals a huge increase in research interest since 2015 on eight dominant themes: machine learning, AI and data analytics, blockchain and BIM, data models and monitoring, IoT and sensor networks, smart buildings and homes, cloud computing and data security, and smart cities and big data, which inform six core research topics: digital design and automation, prefabrication and modular construction, smart construction sites, intelligent operation and maintenance (O&M), intelligent buildings with smart services, and smart city development (Figure 2). For example, smart construction sites integrate robotics, computer vision and IoT for real-time safety monitoring and automation.

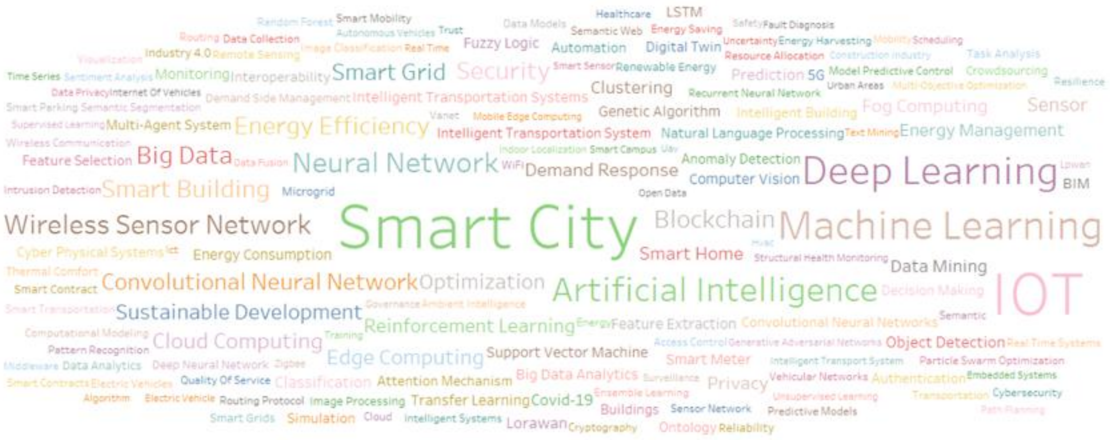

Figure 2 – Word cloud of keywords in the selected papers (Zhang et al. 2024)

Another paradigm of construction - Construction 4.0 is explored by Baduge et al. (2022) in their study in particular the integration of AI, ML and DL into building and construction industry. These technologies transform every phase of the building lifecycle - from design and material selection to construction, operation, maintenance and end-of-life recycling, and authors explore influence of them on each lifecycle. The paper is structured around seven key domains: architectural design and visualization; material design and optimization; structural analysis; offsite manufacturing and automation; construction management, progress monitoring and safety; smart building operation and health monitoring; and sustainability, life cycle analysis and circular economy. Key findings include the use of ML for predicting material properties and optimizing mix designs and the application of DL for structural damage detection and performance forecasting. AI is also shown to improve construction safety via CV and wearable sensors, and support smart building operations through energy optimization and fault detection in HVAC systems.

Despite all the challenges including data scarcity, model generalizability, training complexity and integration with existing workflows, authors emphasize the growing role of smart vision, robotics, IoT, sensor technologies and cloud computing platforms, and forecast a shift from the current era of Construction 4.0 focused on digitizing and automating processes towards to Construction 5.0 with intelligent systems created by AI, ML, DL is deployed from pure productivity in a human-centric, sustainable and collaborative manner (Figure 3). Those systems work in tandem with Digital Twins to manage project full lifecycle, with Additive Manufacturing to reduce waste, with Generative Design, which uses AI to optimize projects, with Augmented and Virtual Reality (AR/VR) enhance safety and precision on-site, with Blockchain ensures transparency in the supply chain. Nowadays there are three major challenges with the transition: limited practical application of digital twins, insufficient human-machine-environment interaction and the need for sustainable practices under carbon neutrality goals. To address these, future directions including the development of digital twin-enabled engineering cloud platforms, integration of human factors into intelligent systems and promotion of net-zero carbon buildings using renewable energy and smart controls are proposed.

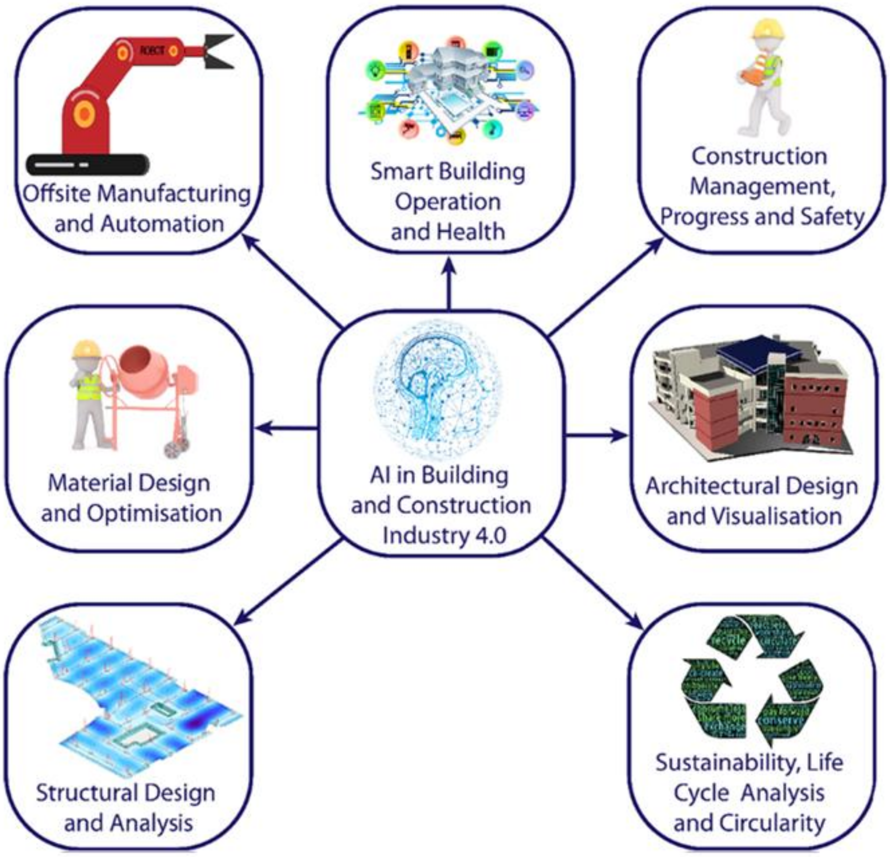

Figure 3 – Application AREAS of AI in building and construction industry 4.0 (Baduge et al. (2022)

The article Pan Y. and Zhang L. (2023) makes a comprehensive review of how BIM and AI can be applied synergistically to boost project management across entire lifecycle of construction. Authors emphasize that BIM and AI integration, in particular machine learning, DL and cognitive computing, can unlock new levels of automation, information processing and reliability in construction processes. This integration is especially valuable for complex and uncertain construction environments and projects, enabling smarter decision-making and reducing reliance on humans. The study identifies and examples six advanced interests in BIM-AI integration for research: automated design and rule checking, 3D as-built reconstruction, event log mining, building performance analysis, virtual/augmented reality (VR/AR) and digital twin technology. For instance, DL models like PointNet++ are used to convert point cloud data into semantically rich BIM models, while Natural Language Processing (NLP) and semantic web technologies facilitate automated compliance checking of design rules. The paper also highlights the potential of AI in the design phase, as a support of generative design and automated rule checking, in construction, as predictive safety monitoring, logistics optimization and smart robotics, in operation and maintenance phase, as preventive maintenance, energy efficiency and facility management through data-driven insights (Figure 4). Looking ahead, the authors propose three future research directions to overcome current limitations: synthesis of human-machine intelligence to incorporate human factors into AI-driven decision-making; development of city-level digital twins for smart urban management; integration of blockchain with BIM to enhance data security, transparency and collaboration. By enabling simulation, prediction, and optimization across all project phases, the hybrid framework of BIM and AI can address challenges, ultimately driving the sector toward smarter and more resilient infrastructure development.

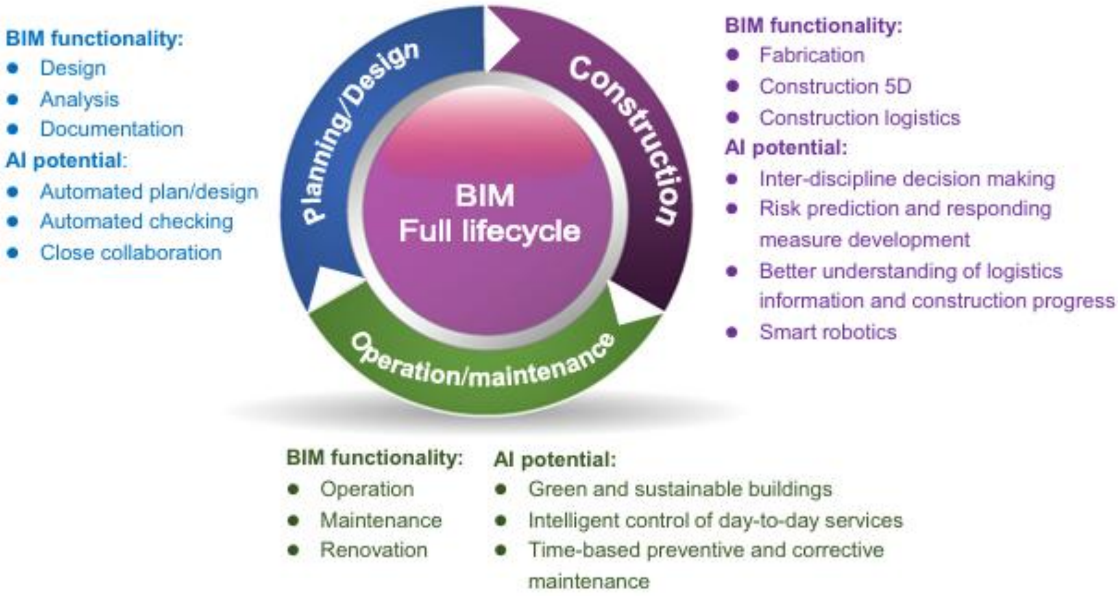

Figure 4 – Potential of BIM-AI integration in three phases throughout a project's lifecycle (Pan Y. at al., 2023)

As was mentioned by Pan Y. and Zhang L. (2023) "The nature of AI is to invent computer programs to automatically learn and think on its own, enabling complex problem-solving and smart decision making with fewer errors and higher efficiency". The article by Ajirotutu et al. (2024) investigates how the integration of AI reshapes infrastructure development: "The application of BIM ... has been significantly enhanced through AI-driven technologies such as ML and predictive analytics, enabling data-driven decision-making and process optimization.” AI by analyzing BIM data enable predictive modeling and risk assessment, to make more informed and strategic choices throughout the project lifecycle: “AI can leverage BIM data to predict potential project risks and develop contingency plans, ensuring that projects are delivered within budget and on schedule.” This proactive approach helps avoid delays and cost overruns, moreover: “AI enhances the decision-making process by identifying patterns and trends in data, facilitating evidence-based solutions to complex challenges in infrastructure development.” AI “fosters resource efficiency and minimizes environmental impact, aligning with global climate resilience goals”, and assess material lifecycles, recommend sustainable alternatives and supporting green building initiatives, what is useful and necessary in nowadays projects in terms of sustainability. Despite these benefits, the article highlights several challenges: data interoperability remains a major barrier for 'widespread adoption'; high implementation costs and ethical concerns, such as algorithmic bias and data privacy, complicate integration and 'can lead to inequities in resource allocation and project prioritization'; cybersecurity is another critical issue, as BIM-AI systems are vulnerable to digital threats, which means 'robust cybersecurity measures are essential' and requires investment in protective frameworks. The article identifies: “Future trends identified include the integration of generative AI, blockchain technology, and the Internet of Things (IoT) into BIM systems.” These technologies are expected to further expand the capabilities of BIM-AI systems and redefine infrastructure practices.

Looking beyond the construction site, AI and big data are also finding applications in building management systems. The article by Himeur et al. (2023) explore how AI and big data analytics transform Building Automation and Management Systems (BAMSs). The authors argue that while BAMSs traditionally manage HVAC systems, lighting, and security, they fail in tasks like performance evaluation, anomaly detection and predictive maintenance, and AI-big data tools considered as a solution to fill the gap. They apply term 'AI-big data' since " as the quantity of data collected in BAMSs is enormous". The paper presents a structured taxonomy of AI-big data frameworks categorized by learning methods, building environments, computing platforms and application domains (energy forecasting, fault detection, IEQ monitoring, water management, occupancy detection), and labeled each strength and limitations (Figure 5). For example, supervised learning offers high accuracy but requires labeled data, at the same time unsupervised learning - more flexible but less precise. Three real-world case studies demonstrate how AI can reduce energy consumption, improve comfort and enhance operational efficiency. Several challenges identified by the authors are data scarcity, lack of interoperability, legalization, cybersecurity risks and need for real-time analytics. Since these innovations aim to make BAMSs more intelligent, secure and scalable there are future directions which authors suggest such as multimodal data analysis, in-situ sensor calibration, smart building digital twins, transfer learning, blockchain integration and edge analytics.

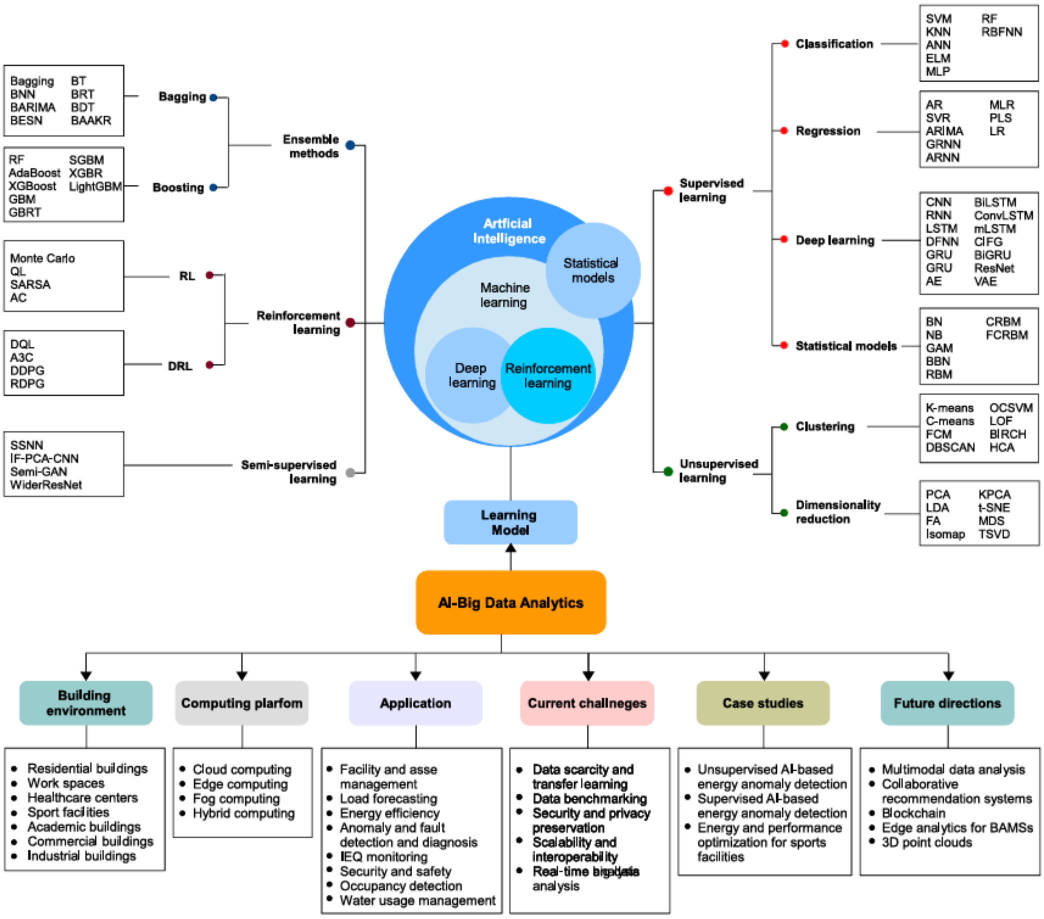

Figure 5 – Taxonomy of existing AI-big data analytics frameworks (Himeur Y. at al., 2023)

Ohakawa et al. (2024) explored the integration of digital tools, AI, and ML for design of affordable housing. The paper highlights how these technologies optimize cost, make housing more accessible, efficient and sustainable, while also improving design precision and reducing waste. It also demonstrates the successful application of Computer-Aided Design (CAD), BIM, and Virtual Reality (VR) with AI and ML in real-world small-scale projects, challenging the perception that advanced tools are only suitable for large-scale developments. Author emphasizes that traditional approaches often compromise quality due to budget constraints. For example, ML models can predict energy performance and optimize resource use, which are especially valuable in affordable housing, where cost-effectiveness and adaptability are critical. In the case studies predictive analytic was used to forecast housing needs and optimize resource allocation, and AI for layout optimization. Authors conclude that digital tools can radically improve design and construction of housing, making it more accessible, sustainable and responsive to diverse community needs. They note that despite challenges like data quality, technical expertise, and regulatory frameworks, nevertheless advocate for broader adoption and continued innovation.

The article by Amer, Jung, and Golparvar-Fard (2021) introduces a novel AI-based framework for automating the alignment between master schedules and look-ahead plans in construction projects. This alignment is critical for accurate progress tracking, payment applications and project coordination, yet traditionally requires manual effort and is prone to inconsistencies. Authors formalize the problem as a ranking task: given a look-ahead planning task, the system must identify the most relevant master schedule activity. Qualitative results show that generated tasks are coherent and contextually relevant, though inaccuracies still highlight the need for human oversight, which is why authors envision the tool as part of a human-in-the-loop system to assist planners and superintendents in aligning schedules and reporting progress. In conclusion, authors emphasize that this is the first formalization and NLP-based solution for automating the mapping between short-term and long-term construction plans. They highlight the model's dual capabilities: matching look-ahead tasks to master schedule activities and generating look-ahead tasks from master activity prompts. Fully automated approach is not developed strongly enough to deployment to industry with the precision 51.1%, on the other hand the semi-automated method shows strong potential with 76.5% precision. Expanding training datasets, refining hyperparameters and developing user-friendly interfaces that support human-in-the-loop workflows are the directions for future research. One of the suggestions of authors - integrating this model with automated progress monitoring systems to enable bidirectional updates between field-level tasks and master schedules, ultimately improving visibility and coordination across project teams.

2.3. Integrating Computer Vision into Construction

One of the most promising areas of AI application in construction is CV, which allows for automatic progress tracking, classification of objects and detection of deviations. AI in combination with BIM allows to recognize and classify automatically objects on a construction site. This capability is crucial for monitoring and managing safety.

In their work, Dolhopolov et al. (2024) propose a multi-stage classification method for objects that use AI tools like YOLOv5, and neural networks integrated with BIM to recognize construction elements and equipment. The study aims to develop a digital twin of construction sites by combining image-based modeling, neural networks, and IoT technologies. Authors propose a multi-stage classification system that uses Structure from Motion (SfM) to generate 3D models from site photographs, producing point clouds, mesh models, and Bill of Materials (BOMs). These representations are then analyzed using the YOLOv5 object detection algorithm and a Feedforward Neural Network (FFNN) to assess conformity with predefined standards. The system enables real-time monitoring and classification of construction elements across various stages of development. Two photogrammetry methods are used: static photo modeling, where cameras are fixed throughout the observation period, and dynamic photo modeling, where cameras move at intervals. It achieved a mean Average Precision (mAP) of 0.73, with the highest precision recorded - 0.88. It also explores the potential of Augmented Reality (AR) and cloud-based solutions for visualizing BOMs and managing large datasets. Authors argue that this approach supports scalable, real-time construction site modeling and opens new avenues for automation and digital transformation in the construction industry.

Kropp, Koch, and König (2018) integrate 4D BIM with computer vision techniques to automate indoor construction progress monitoring during renovation work at Ruhr-University Bochum. Interior finishing works, which account for 25–40% of total construction budgets, require effective project control with significant precision and frequency exposed challenges connected to labor-intensive manual progress tracking. The proposed method enabling accurate mapping of visual data to be scheduled construction registering each captured frame of on-site video sequences to the 4D BIM model tasks and determining the camera's position and orientation within the building's coordinate system. The framework supports object recognition (detecting installed radiators) and material and structure recognition (identifying drywall installation stages: paneling, plastering, painting). The system was tested on three distinct stages of drywall installation with unique visual challenges on each stage: panel installation which involves detecting screw spots and horizontal seams, plastering with low-contrast textures issues and painting in a uniform surface with minimal structural features. Authors used segmentation to photos to isolate the drywall from occluding objects, compass edge filters to detect structural patterns and performed contrast normalization to distinguish plastered from painted surfaces. Despite the variable and backlighting, partial occlusions, cluttered scenes and motion blur precision and recall rates achieved above 95%. Thus, the method is useful and effective for daily inspections and real-time decision-making, enabling frequent and accurate updates with minimal human intervention.

To successfully apply CV in construction, it is critical to understand how models are trained and what methods improve their accuracy. Shorten and Khoshgoftaar (2019) review techniques used to enhance image datasets for training deep learning (DL) models. Focused on CNNs which are central in modern computer vision tasks due to their ability to learn hierarchical representations of image data. They use parameterized, sparsely connected kernels that preserve spatial relationships while progressively down sampling image resolution and increasing feature depth. This architecture enables CNNs to extract low-dimensional, highly informative features that outperform hand-crafted descriptors. However, CNNs are highly data-dependent and prone to overfitting - when a model perfectly predicts training data but fails to generalize testing examples - when are trained on small or imbalanced datasets. One of the ways to mitigate it - data augmentation as a strategy to artificially expand and diversify training datasets. Augmentation techniques are categorized by authors into two main types: data warping and oversampling. Data warping includes geometric transformations (e.g., rotation, flipping, cropping), color space manipulations, kernel filters, and random erasing. These methods preserve labels while introducing variability that helps CNNs learn more robust features. Oversampling methods, such as image mixing and GAN-based synthesis, generate entirely new samples to balance class distributions and simulate rare conditions. There are also advanced augmentation strategies like adversarial training, neural style transfer and feature space augmentation. These techniques are useful in case of limited data availability and aim to improve model robustness and generalization. Experimental results demonstrate that augmentation techniques they used consistently improve CNN performance across benchmark datasets like CIFAR-10, ImageNet, and MNIST. The paper concludes that success of CNNs depends heavily on the quality and diversity of training data, and data augmentation remains one of the most effective tools for unlocking their full potential.

Deng et al. (2019) automate monitoring progress of tile installation indoor construction site environment. Need for accurate, timely and intuitive progress data to support decision-making is a key challenge in construction management, as authors address. Monitoring progress traditional methods rely heavily on manual measurement and reporting, which are time-consuming, error-prone and inefficient. The overcome for that is CV–based system, integrated with BIM that enables real-time, automated quantity tracking and visual progress updates. Authors created a 1,000 images tile image database, collected from both construction sites and online sources, and trained the system to detect tile and its boundaries through pixel coordinates which are then converted into real-world coordinates through camera calibration, aligning the detected tile areas with BIM model. The highest classification accuracy yielded to 91.17%. The system can eliminate misclassifications and determine the correct orientation of tiles to refine the accuracy of tile boundary detection according to geometric data and tile dimensions in

BIM model. The method also accounts for tile cutting and breakage: cut tiles are identified and quantified using area-based analysis, while broken tiles are detected through image uploads by workers and processed using the trained classifier. The system was tested in a real-world setting during the interior decoration phase with fixed camera captured images at 30-second intervals. Experimental results showed that the classifier performed well even in low-light conditions, and the improved edge detection algorithm successfully filtered out irrelevant lines. A visual progress plan was the final output for managers to view real-time updates and make informed decisions. In conclusion, the study demonstrates that integrating CV with BIM provides real-time quantity statistics, reduces manual effort and enhances the intuitiveness of progress visualization.

AI and CV are an innovative approach in automating the tracking of construction progress. In their study Jung et al. (2024) presented a transformer-based model that automatically maps construction schedule tasks to ASTM Uniformat classifications, via computer vision-based semantic segmentation, and due to that significantly increases the accuracy of 4D BIM, allowing to analyze reality capture data for up-to-date progress data. This innovation addresses the challenge of aligning schedule data with BIM and payment systems, which traditionally requires manual coding and coordination across departments. The training dataset includes over 35,000 activity sequence tuples from ten commercial building projects. Each tuple consists of a target activity and its surrounding context, enabling the model to learn from multiple perspectives. It shows consistent performance across projects, even with imbalanced data distributions. This reduces manual effort and enhances project planning and control. The study highlights the importance of fine-tuning pretrained language models for domain-specific tasks, comparing to ChatGPT and FastText-based embeddings, UNIFORMATBRIDGE demonstrated superior performance in understanding construction-specific terminology and abbreviations. Future work aims to expand its capabilities to include resource allocation, scenario analysis, and integration with reality mapping for enhanced process understanding.

The article by Braun et al. (2020) presents a framework for automating construction progress monitoring by integrating image-based 3D reconstruction, semantic BIM data and computer vision techniques. Instead of laser scanning, the system uses conventional camera images processed through SfM to generate point clouds, which are then compared with "as-planned" BIM model to assess the actual state of construction. To overcome occlusions and temporary structures like scaffolding and formwork, the system incorporates semantic relationships and technological dependencies between BIM elements, enabling inference of hidden components and dynamic adjustment of detection thresholds based on construction stage. A key innovation is the use of DL, specifically a Mask R-CNN model trained on thousands of labeled images, to classify construction elements such as walls, columns, and formwork in 2D image space, which achieves 0.92 F1 score, significantly outperforming geometry-only methods. The system also employs HSV color space analysis to distinguish materials like concrete and wood, enhancing classification accuracy. These techniques are embedded within a digital twin, combining geometric, semantic and visual data to reflect real-time conditions and support coordination, quality control, and progress verification. The use of 4D BIM enables temporal reasoning, linking construction elements to scheduled timelines and allowing logical inference about sequencing through Precedence Relationship Graph (PRG). By projecting BIM geometry into image space and applying visibility analysis, the system identifies which elements should be detectable from specific camera positions, refining detection accuracy.

Asadi et al. (2019) automates the registration of video frames to BIM in real time to improve indoor construction monitoring and address time-consuming and error-prone process of manually comparing as-built conditions on a site with as-planned data challenge in construction management. Traditional manual observations and documentation can consume up to 30% of daily effort on-site. Proposed by authors method aims to overcome these issues and integrates CV techniques with BIM geometry. The core of the approach combines augmented monocular SLAM (Simultaneous Localization and Mapping) with perspective detection and matching. SLAM is used to estimate the camera trajectory and generate sparse point clouds, while vanishing points and lines extracted from both video keyframes and BIM views are used to refine camera poses. Gradient descent-based optimization enables the system to localize the camera within the BIM coordinate system and align each keyframe with its corresponding BIM view, facilitating automated comparison between as-built and as-planned conditions with minimized alignment errors and improved accuracy of image-to-BIM registration. There were two case studies tested with robust performance, despite the challenges posed by low-feature environments and occlusions: a featureless hallway and a cluttered indoor construction site. It achieved real-time registration with average processing times of 0.65 seconds per keyframe in the hallway and 1.9 seconds in the construction site. The system was implemented on Jetson TX1 platform and showed that real-time performance is feasible in simple scenes, while more complex environments may require slower movement or more powerful hardware. One of the key innovations of the method is its use of vanishing geometry, specifically vanishing points and lines—as a means of refining camera localization. This is particularly valuable in indoor environments, where traditional SLAM methods often struggle due to lack of features and lighting variability. However, the study acknowledges several limitations: the accuracy of vanishing point estimation is sensitive to image resolution and scene clutter; high-resolution images yield better results but compromise performance. Authors note that the method is best suited for environments with straight edges and orthogonal geometry and may require adaptation for buildings with curved or irregular shapes. In conclusion, the proposed framework offers a scalable and efficient solution for automating visual data registration in construction, supporting progress tracking, quality assessment and decision-making directly within the BIM environment. It serves as an enabling technology for fully automating vision-based monitoring systems and has potential applications in autonomous data collection, robotic navigation and real-time project control.

2.4. Neural Networks in Construction

Munawar et al. (2022) in their study present a deep learning-based framework to detect cracks in buildings automatically with use of aerial imagery captured by drones. Authors propose a modified CNN architecture with 16 convolutional layers, integrated with CycleGAN for unsupervised image translation. The model was taught to detect and segment cracks using a dataset of 1300 images collected via Unmanned Aerial Vehicles (UAVs) and supplemented with open-source data in mid- and high-rise buildings in Sydney, Australia. The methodology includes image pre-processing, per-pixel labeling and segmentation using a U-Net and SegNet-based architecture. Authors apply data augmentation techniques such as rotation, cropping, and flipping to enhance model robustness and reduce overfitting. Crack pixels are categorized into significant and weak classes based on pixel width, and guided filtering (GF) and conditional random fields (CRFs) are used to refine predictions. The model is evaluated using multiple metrics with following results: global accuracy (0.990), class average accuracy (0.939), mean IoU (0.879), precision (0.838), recall (0.879), and F-score (0.8581), which outperform baseline methods results and other architectures like DeepCrack-BN, DeepCrack-GF, and SegNet. The proposed CNN demonstrates that the framework is scalable, efficient and suitable for real-time crack detection, offering significant improvements over manual inspection in terms of speed, cost and reliability. However, authors acknowledge limitations, particularly related to CycleGAN's potential instability and lack of photometric augmentation. They suggest future work should broadly explore geographic datasets, test HSV-based augmentation and implement transfer learning to generalize across different building types and cities.

Amer and Golparvar-Fard present a ML based method to automatically extract construction planning and sequencing data from archived project schedules. Authors introduce Dynamic Process Templates (DPTs), which are generative models built using Long Short-Term Memory (LSTM) neural networks. These models had been taught on real-world scheduling datasets to recognize activity dependencies and sequencing logic. To understand unstructured nature of schedule layout, authors use domain-specific word embeddings and tagging system that identifies key constituents such as action and object, and this way the scheduling data could be transformed into structured sequences suitable for training. The approach was validated on a dataset of over 78,000 activities from 32 diverse real-world construction projects, demonstrating powerful performance, with accuracy between 76% and 98% depending on the dataset scope. They were able to predict both successor and predecessor activities and showed robustness in handling variations in natural language expressions and unseen sequencing logic. Authors found that the DPTs could generalize well even abbreviated or rephrased expressions. The study highlights several practical applications of DPTs: automated look-ahead planning, schedule quality control and support for progress monitoring systems. However, authors also acknowledge limitations: assigning all dependencies as the same type (e.g., Finish-to-Start), ignoring more complex relationships; lack of ability to learn location-dependent workflows because location-specific information was removed to improve generalizability; model complexity and reducing in performance because of training on highly diverse datasets. In general, the results demonstrate that DPTs can effectively learn and represent construction sequencing knowledge, offering a scalable and flexible solution for digitizing planning expertise and improving project management practices.

In general, removing image degradations—such as construction visual noise, blur, or unwanted objects—does not consistently improve the accuracy of CNN-based image classification. The study of Pei Y. at al. (2024) found that many preprocessing methods aimed at restoring image quality (like dehazing or deblurring) often fail to recover meaningful features and can even distort important visual cues. As a result, CNNs trained on these processed images may perform worse than expected. The key insight is that degradation removal does not add new information to the image and may interfere with feature extraction, making it less effective for classification tasks. When both training and testing datasets undergo the same preprocessing—such as construction visual noise removal or object elimination, the CNN model can achieve better or comparable accuracy than using raw degraded images. The study shows that consistency in data preparation is crucial: training on images that match the degradation level or preprocessing style of the test set leads to significantly better performance. However, even in this case, the benefit of preprocessing is limited. Often, training directly on degraded images yields equivalent results to training on restored ones, suggesting that CNNs can learn to manage degradation if it is consistently present in the data.

2.5. Visual Monitoring Integration in Construction

Continuing with the topic of visual monitoring, the article by Shinde Y. at al. (2023) represents a systematic review of studies of the use of both images and videos 360° panoramic technologies within and out of AEC industry published between 2000 and 2022. Authors highlight that thanks to its lower cost, higher sense of presence and greater immersion 360° panorama technology has become more popular compared to VR and AR visualization technologies which have long been used to simulate realistic environment. It is a useful tool to remotely overview construction sites, get the current status of work and capture and solve potential issues without needing to be physically present. This is particularly valuable for large-scale projects or for international teams spread in various locations. Virtual walkthroughs or 360° panoramas due to their immersive quality provide more organized documentation exchange and communication among departments, more rich presentations for clients, more clear design reviews and coordination meetings, and enhancing understanding of spatial relationships and construction sequencing, which is often difficult to convey through 2D drawings or photos. It is used not only for visualization, but also to monitor construction progress over time, providing a comprehensive view. Teams can track changes by capturing panoramic views with regular intervals, compare planned versus actual time and stage, and maintain a visual record of structure installation. Virtual walkthroughs assist in dispute resolution or quality control with more transparency and accountability. There are still navigating challenges in large environments, data storage and rendering quality and compatibility with various hardware systems even though the technology is developed enough for practical use. Overall, 360° panoramic technologies offer significant potential for improving the construction industry.

Fang et al. (2023) use 360° panoramic images and DL techniques for a fast and cost-effective method to evaluate interior construction progress. The authors address the limitations of LiDAR and multi-image vision-based approaches, which are often impractical due to equipment cost, complexity and occlusion issues. They propose a method requiring only one or two panoramic images per room unit, making it highly efficient for data collection in real-world construction settings. The method was validated through a case study in a hotel renovation project in Hong Kong. HorizonNet - a neural network that estimates key corners and generates point clouds from panoramic images - reconstructs the room layout, then multiple point clouds views are merged to resolve occlusions and produce a unified spatial model by Bayesian Coherent Point Drift (BCPD). This step is crucial for maintaining spatial consistency. After PointNet++ - a hierarchical DL architecture - is applied to perform semantic segmentation on the sparse point cloud and identify specific interior trades such as tiling, waterproofing, and boarding. The segmentation is trained on manually labeled datasets, with each trade represented by millions of points. The segmentation performance varies by trade: boarding achieves an Insertion over Union (IoU) of 75.57%, waterproofing 61.62%, and tiling 52.99%. The lower accuracy for tiling is attributed to its visual similarity with surrounding surfaces and reflective properties that complicate feature extraction. There are acknowledged limitations by authors: the point clouds are 2.5D rather than fully volumetric, the DL models require fine-tuning and high-performance hardware and the lack of publicly available datasets for interior environments poses a barrier to broader adoption. Integration of AI and 360° imaging for interior progress monitoring which is scalable alternative to traditional methods, particularly in environments where consistent access to capture points is difficult. This is especially relevant considering challenges observed in platforms like OpenSpace.ai, where users may lose access to specific viewpoints when switching between dates due to route inconsistencies. Fang et al.'s approach directly addresses this issue by enabling multi-view registration and robust segmentation, ensuring that progress can be tracked even when capture conditions vary over time.

Amanda S. Barbosa and Dayana B. Costa (2022) use BIM, Unmanned Aerial Systems (UAS) and 360° cameras for tracking construction progress both indoors and outdoors . Authors highlight that traditional progress monitoring methods rely on manual visual inspections and are prone to errors due to human subjectivity, especially in residential projects, where many activities occur inside buildings and are harder to observe. To address these limitations, two exploratory case studies conducted. The first case of a residential complex with 20 buildings with five floors and four apartments per floor focused on testing the operational workflow for collecting and processing visual data using UAS for external site and 360° cameras for indoor site on safety helmets for immersive photographs. Aerial images were captured with DJI Phantom 4 drone and processed in Agisoft Metashape to generate point clouds and textured 3D models, the generating process for indoor point clouds from 360° images was too time-consuming and was declined by authors. The second case study BIM model of 13 buildings residential complex was created in Revit and linked to 4D schedule in Navisworks. External point clouds were manually aligned with BIM model, and internal progress was tracked by linking 360° photographs to certain viewpoints in BIM model. The team found that 360° cameras are effective for indoor progress tracking, demonstrating fast data collection and immersive visualization. However, photogrammetric reconstruction from these images is less practical for large-scale projects due to long processing times. Integration with BIM platforms like Navisworks and Synchro allowed visual comparison between planned and actual states, although each platform had limitations—Synchro does not support point clouds and Navisworks does not allow side-by-side model views. Based on these findings, authors proposed structured method that includes data collection, processing, integration with BIM and regular analysis. External data is collected via UAS and processed into point clouds, while internal data is captured with 360° cameras and stored in cloud platforms for immersive viewing, then linked to the BIM model to support progress tracking and decision-making, assessing whether activities are fully completed or require rework.

Along with images, new data sources – satellites, drones, video – are expanding monitoring capabilities. In addition, Glinka (2024) proposes Sat4BIM4D concept, which uses satellite remote sensing data combining with 4D BIM to monitor construction progress. Such approach opens new possibilities for large-scale infrastructure project tracking. The study reviews literature, highlights the advantages, limitations, and challenges of using satellite data for this purpose, and emphasizes the importance of precise and timely progress tracking in construction, noting that manual methods are inefficient, costly, time consuming and prone to delays. Glinka identifies a gap in the use of satellite data for automated construction progress monitoring, and proposes the Sat4BIM4D framework, which incorporates algorithms for processing satellite-derived data to complement BIM and enhance project management. While satellite imagery offers broad coverage and frequent revisit capabilities, its integration with BIM remains underexplored. Despite the promise of Sat4BIM4D, the study acknowledges that current satellite data may not support daily progress tracking due to cost and resolution constraints, cloud cover interference, and the need for high-frequency data acquisition. Instead, it recommends using satellite imagery for milestone-based monitoring at weekly or monthly intervals.

Literature review demonstrates potential progress in the integration of 4D BIM with AI, ML and CV to automate progress tracking, optimize scheduling and enhance decision-making. However, the clear gap between theoretical research and real-world implementation remains.

In most of the studies presented proof-of-concept models are tested under controlled conditions, when CNNs and other ML architectures achieve high accuracy in classifying construction progress. Although for real-world environments with factors such as lighting variability, occlusions, dynamic site conditions, and diverse construction practices, which are challenging the generalizability of these models, their robustness is limited. Additionally, construction projects are highly heterogeneous, thus, data scarcity remains a fundamental bottleneck: developing of large, annotated datasets requires considerable time, expertise, workflow insight and collaboration between stakeholders.

Another critical issue is the lack of attention to organizational and industry-related factors. Practical adoption of new technologies despite its technological capability requires overcoming barriers such as economic feasibility, workforce, software providers and developers' readiness, integration with legacy systems and alignment with project workflows processes, considering the cultural resistance within the construction sector, where traditional methods remain deeply embedded.

Overall, the technological maturity or possibility of integration showed by studies alone is insufficient to drive transformation of the sector. It must be supported by industry-wide training, interoperability standards, cost-benefit validation and stronger regulatory frameworks to implement, enhance and reach its full potential.

This page is intentionally left blank

Literature review demonstrated strong research interest and showed benefits of AI and ML. Nevertheless, the construction industry has not yet widely adopted these technologies due to the barriers which can be summarized as follows:

- Data Limitations – Inconsistent collection methods, lack of annotation and standardization, workflow insight, complexity of construction processes, fragmented ownership of project data challenge generation large-scale, standardized, diverse and high-quality datasets, influencing the reliability of training AI models.
- Workforce Readiness – The construction sector due to domain of conservative approach and traditional workflows, with resistance to changes, faces lack of trained professionals who combine expertise with both construction, digital and AI literacy further slowdown adoption.
- Integration Challenges – Many AI/ML solutions remain isolated expert-oriented tools, for their wide adoption to the industry it is necessary to provide seamless, user-friendly interoperability with existing BIM platforms, scheduling systems and other related software.
- High Costs and Investment – Deploying AI requires investment in computing infrastructure, software and skilled personnel. For small and medium-sized enterprises, such costs are often prohibitive, and the return on investment remain sceptical. The benefits of AI/ML are often long-term, while upfront costs are immediate, leading to hesitancy in decision-making.
- Regulatory and Ethical Concerns – Without clear standards, applied for AI tools in wide-spread BIM platforms or other tools, companies are cautious in implementation of AI due to issues of data security, privacy and potential algorithmic bias, especially in classified or infrastructure projects, which are often the most large-scale and are needed in such tools.

3. DATA COLLECTION

From the beginning of ML application, it has undergone a fundamental transformation, moving from models trained on pristine, controlled datasets of images to real-world imagery with its complexity, being able to leverage them - the transition indicated a pivotal shift in the field of CV. On preliminary stages computer vision systems historical development relied heavily on small-scale, meticulously curated images datasets, which were often captured in prepared laboratory environment - this was a foundation for a beginning of algorithmic research. However, those idealized images, without occlusions, diversity of lighting conditions and unpredictable angles, severely limited the generalizability and robustness of the resulting models to deploy them in practical, uncontrolled settings. In the early 2010s large-scale, created by crowd-source, datasets, such as ImageNet, including millions of photographs from a vast array of real-world sources, appeared and changed habitual datasets for ML, providing a rich, noisy and highly varied training ground. With such a base it became enabled to train the DL architectures, particularly CNNs, which demonstrated unprecedented performance in tasks like image classification and object recognition. Essentially, DL revolution was a direct result of this move from synthetic to authentic data. Contemporary research in machine learning deeply focuses on addressing the inherent challenges of real-world imagery and the way to apply it for real-world issues. This research develops robust model architecture, advanced data augmentation techniques and methods for semi-supervised and self-supervised learning that can function effectively with limited or imperfect annotations. The transition from idealized to authentic data has opened new frontiers for a wide range of applications of ML, such as using CV for progress monitoring in land development projects (Han J. et al. 2024), analyzing images to detect safety hazards and unsafe behavior (Nath N., 2020) and for automated quality control in construction (Garita-Durán H., 2025). This dissertation aims to go further with this trend to address the critical problem of tracking construction progress on-site using machine learning and photographic data.

This dissertation was developed in close collaboration with industry partner BIMMS, a leading BIM consultant company in the European market. The empirical foundation of this research rests on a dataset derived from BIMMS company through OpenSpace.ai platform - a reality capture and AI-powered analytics platform. It automatically matches 360° photos and videos to project plans, streamlining workflows and paperwork and improving communication between field and office teams by analyzing and comparing captured data with project models. The platform allows to revisit any point in time through an interactive map and tools for progress tracking. OpenSpace is used by owners, general contractors and traders to track progress, resolve conflicts and validate work-in-place.

From a technical perspective, OpenSpace.ai through its Vision Engine integrates and maps captured panoramic image with architectural 2D plan data, enabling precise user localization within the virtualized representation of a construction site and creating an accurate “as-built” spatial index. There are two main navigation paradigms: spatial navigation across the floor plan both within the same level and vertically between levels; and temporal navigation between multiple capture dates. Users can explore captured environments interactively, switching seamlessly between snapshots taken days, weeks or months apart. Building on the concept of dynamic digital twin that reflects real-time conditions, Braun et al. (2020) propose a complementary framework that enhances progress monitoring by combining geometric - point clouds, semantic - BIM data and visual data - computer vision (CV). Instead of relying on laser scanning, their system uses conventional camera images processed through SfM to generate point clouds, which are then compared with "as-planned" BIM model.

This category of technology has integrated into digital construction workflows, enabling the creation of navigable, time-linked visual records of complex construction site space. In the context of AEC industry, 360° image-based monitoring has been increasingly adopted as a cost-effective and scalable tool of organized documenting, remote overview construction progress, capturing potential issues with higher sense of presence and greater immersion, especially in large-scale and unique projects, enhancing understanding of spatial relationships and construction sequencing for variable departments, which is often difficult to convey through 2D drawings, periodic manual inspections or photos. Teams can track changes by capturing panoramic views with intervals, making it possible to revisit past site conditions virtually and to compare them with current states of a structure installation that supports both operational decision-making and retrospective analysis (Shinde et al., 2023).

However, while the OpenSpace.ai interface is notably intuitive and requires minimal training, the temporal navigation functionality presented a notable limitation during data collection. Specifically, when a user attempts to change the capture date from a particular viewpoint, if imagery for that exact location on the target date does not exist, the platform redirects the view to the nearest available capture point. Although this approach ensures that a user is always presented with some relevant visual context, it can disrupt analytical workflows where precise spatial consistency between timepoints is critical. The consequence is that fine-grained temporal comparisons, particularly those tracking incremental changes elements, can be interrupted. Similar challenges have been reported in other research utilizing 360° capture for progress monitoring, where variable capture density and route inconsistency hinder the ability to establish true one-to-one temporal correspondences in interior environments. As Fang et al. (2023) note, “only a limited number of locations can be chosen to capture images” due to "since indoor rooms are messy and full of various building materials during the construction period". Moreover, “to overcome the occlusion problem caused ... and the accumulation of building materials, multiple frames at different viewpoints are required.” These constraints make it difficult to maintain a stable set of capture points across dates, which in turn leads to situations where users may completely lose access to a previously visible object when switching between timepoints — a critical limitation precise progress tracking.

The dataset for this study was sourced from four construction projects of a data center, each with unique operational schedules and environmental conditions. The periodicity of capture varied: some sites were documented once per week, others twice per month, and in certain cases, multiple times per week. At the same time, the camera operator routes taken during each capture session were not always consistent, sometimes covering various parts of the site or recording the same areas from alternate angles - too far or too close, or missing some areas for one or several surveying. These inconsistencies are common limitation in real-world construction documentation workflows, where safety protocols, operational priorities or site access restrictions influence the capture process.

From the outset, it was evident that to build a robust dataset suitable for machine learning training, it was necessary to identify a construction element that:

- exhibited strong visual distinctiveness under variable lighting and environmental visual noise;

- was present in all four projects to collect a suitable amount of data;
- showed clear progression with multiple stages to identify classes;
- maintained a relatively consistent structural composition across sites.

The drywall (gypsum board) was selected as the target element for data collection. In contemporary construction practice, drywall partitions are ubiquitous, cost-efficient and relatively fast to install, making them a fixture of both residential and commercial projects. Structurally, such a partition consists of a supporting frame (either cold-formed steel studs or timber), insulation material, a rear gypsum board layer, a front gypsum board layer and a surface finish such as paint or textured coating. The fully completed wall (100%) is the one with smooth white or grey finishing. The installation process is inherently incremental, offering multiple visually distinguishable intermediate stages that can be readily identified in imagery.

For the purposes of machine training and testing, dataset with classification categories was created, and three classes A, B, C were defined based on stages of installation progress in percentage, from structure to finished wall:

- Class C (5–30% completion):
  This stage typically included bare framing or frames with partially or fully installed rear gypsum boards. Insulation may be partially put in place, but the rear board layer remains visible, with seeable structural studs. Some examples displayed partial installation of the front gypsum board or initial partial painting, yet the underlying board was still seeable (Figure 6).

Figure 6 – Class C wall example photos

- Class B (30–80% completion):
  At this stage, the rear gypsum board is not seeable, insulation is fully in place and the front gypsum board is either partially or completely installed. Visible seams between boards are common, sometimes with joint compound applied, and structural studs are still perceptible (Figure 7).

Figure 7 – Class B wall example photos

- Class A (80–100% completion):
  These walls are with mostly finished surfaces often painted white or gray depending on the interior design. The surface may have partially or fully compounded joints and, in some cases, communication elements such as electrical outlets, wiring or plumbing fixtures are installed (Figure 8).

Figure 8 – Class A wall example photos

Across all classes, the collected imagery contained substantial amounts of construction visual noise - extraneous elements such as stored materials, tools, machinery and personnel within the camera's field of view, door and window openings, sometimes revealing further clutter beyond (Figure 9). Furthermore, lighting conditions varied widely: some images were underexposed due to low-light interiors, some of them were overexposed with close related lighting (Figure 10). These variations are significant in the context of computer vision because they influence feature extraction and can introduce unwanted variance in the model's learned representations "Lighting biases are amongst the most frequently occurring challenges to image recognition problems" (Shorten & Khoshgoftaar, 2019). Camera viewpoint diversity was another factor. Variation in viewpoint can be beneficial for generalization for the model to learn and recognize an object from multiple perspectives - it also increases intra-class variability, requiring a larger dataset to achieve robust performance (Figure 11).

Figure 9 – Photos with construction visual noise

Figure 10 – Underexposed and overexposed photos

Figure 11 – Photos of a wall from different angles

Images were extracted as screenshots with baseline parameters of 1760 x 1080 pixels, 96 dpi, and 24-bit color depth from the OpenSpace.ai platform through 360° captured panorama. These parameters preserved a balance between resolution sufficient for fine detail recognition and manageable file size for dataset storage and model training and testing. In the process of model training with preliminary training datasets, it became clear that selective cropping could improve model accuracy. Cropping was applied to remove portions of the images containing non-target elements, such as walls from a different completion class, excessively sizable portions of ceiling or floor, or occluding objects such as equipment or building materials. The cropping was limited to lateral and vertical margins to maintain the central visual focus on the target wall (Figure 12). This technique is consistent with findings in the literature, where targeted removal of irrelevant regions in an image has been shown to reduce classifier confusion and improve accuracy in the case if those cropped photos are used in both testing and training datasets (Pei et al., 2019). As a result, the final dataset contained varied sizes of images with various dimensions, both within and between classes: in Class A 64 out of 200 were cropped, in Class B - 70 photos, in Class C – 88 photos. This introduces a degree of variability for training.

Figure 12 – Original with construction visual noise (a.) and cropped (b.) photo examples

The final dataset comprised 600 photographs, equally distributed across the three completion classes A, B, C (200 per class). The dataset evolved in incremental stages with the needs for the model training:

- Initially, 100 images per class were collected to enable rapid prototyping and training of the model.
- Following initial model evaluations, the dataset was expanded to 150 images per class to address overfitting and increase accuracy.
- Further increases to 200 images per class were driven by the need to improve accuracy to reach the target number.

During the earlier stages of collection, the internal distribution of subcategories within classes was highly uneven, also the classifications were different from each attempt (Table 1). By the final iteration of dataset construction, these subcategories had become more balanced and organized, ensuring that each class contained more representative variety of construction states.

Table 1: Classification of photos through attempts

# Class	Attempt 1		Attempt 2		Attempt 3	
	Number of photos	Criteria of the class	Number of photos	Criteria of the class	Number of photos	Criteria of the class
A	100	Walls with finished upper plasterboard layer and with painting on it from the top or from the bottom, or full wall, and with installed pipes or other communications: - 36 out of 100 with partly finishing; - 64 out of 100 with full finishing	150	The same stage of the structure as in Attempt 1 with the increase of number of photos: - 105 out of 150 with full finishing; - 45 out of 150 with partly finishing	200	The stage with compounded joints was added to this Class from Class B: - 100 of completed and partly finishing walls; - 100 of walls with plasterboards jointing; * 64 out of 200 were cropped
B	100	Walls with full insulation layer completed and the upper plasterboard layer started, also finished plasterboard layer and jointing: - 43 out of 100 with partly visible insulation; - 35 out of 100 with jointing; - 22 out of 100 with upper plasterboard	150	The same stage of the structure as in Attempt 1 with the increase of number of photos: - 53 out of 150 with visible insulation; - 50 out of 150 with jointing; - 47 out of 150 with upper plasterboard	200	The stage with compounded joints was moved from this Class to Class A: - 100 walls with finished upper plasterboard layer; - 100 walls with visible finished insulation layer; * 70 out of 200 are cropped
C	100	Walls with visible lower plasterboard and metal structure: - 25 out of 100 have visible insulation layer with/without upper plasterboard behind; - 75 with just lower plasterboard layer and structure frame	150	The same stage of the structure as in Attempt 1 with the increase of number of photos: - 72 out of 150 with part of insulation or/and upper plasterboard layer; - 78 out of 150 with just lower plasterboard layer and structure frame	200	The same as in Attempt 2: - 100 of walls with partly insulation or/and upper plasterboard layer; - 100 of lower plasterboard and structure frame walls; * 88 out of 200 are cropped

Data collection was constrained by several factors:

1. Not regular or skipped capture phases, when an object of study in the same location moved directly from Class C to Class A, or started from Class B skipping Class C between captures, eliminating some classes examples for that element.
2. Obstructed viewpoints, where the target wall was hidden behind temporary building materials storage, personnel, large equipment or other objects.
3. Non-replicated capture routes, which limit the ability to track the exact same element across time and lead to lose one class or classes examples of that element (Figure 13).

Figure 13 – Examples of routes inconsistency on OpenSpace.ai platform

These constraints mirror challenges commonly reported in real-world construction monitoring projects, where operational, logistical and environmental variables limit the completeness of longitudinal datasets "Real-world data from construction sites always introduces many occlusions, and non-modeled elements that make it nearly impossible to detect all elements on a construction site" (Braun et al., 2020). The challenges for the drywall images are common because of the wall features in general as Kropp C. et al. (2018) mentioned in their case study: “The main challenges for determining the state of the drywall are the low structure and the homogeneous color distribution over the wall. Furthermore, in indoor environments, walls may not completely be visible if a special lens is not applied”.

The methodology demonstrates that from 360° panoramic imagery can be produced a rich and diverse dataset for tracking construction progress over time. However, there are constraints to capture a big amount of data because of manual process and other mentioned obstacles that complicate getting better accuracy. The dataset captures certain variations in lighting, angles and occlusions while remaining generalizable across different construction sites. However, since it is not the main type of walls the combination of number of accessible elements for photo capture and overall elements produce still not much data.

Although panoramic capture and spatial indexing through OpenSpace.ai streamlined real-world data collection, reduce the need to search through data in open access, there are challenges in photo capture workflow such as inconsistent capture schedules, route deviations and missing viewpoints highlight the need for thoughtful dataset collection that anticipates such gaps. Ultimately, the resulting dataset offers a solid foundation for training CNN model capable of automated wall installation progress classification - an increasingly valuable tool for improving productivity, bill and quality control in digital construction management.

4. MACHINE LEARNING MODEL TESTING

Machine learning as the core component of AI that enables to analyze data, does not rely on explicitly defined rules as traditional programming but learns from data and builds predictive models capable of solving complex and powerful tasks, like image recognition, where rule-based approaches are often impractical due to the variability and complexity of visual inputs. AI according to Cambridge dictionary - “the use or study of computer systems or machines that have some of the qualities that the human brain has, such as the ability to interpret and produce language in a way that seems human, recognize or create images, solve problems, and learn from data supplied to them”. One notable example in construction of AI adoption is BuildAI - an AI-driven construction management platform. BuildAI automatically reports construction progress and provides real-time transparency into critical path activities, helping project teams improve coordination, safety, and efficiency across crane and site operations. It integrates intelligent software with on-the-ground construction experience, offering insights into productivity and enabling proactive decision-making. Baduge et al. (2022) in they work emphasize the transformative potential of AI in enabling smarter, data-driven construction practices, “the capability of AI to process massive amounts of data, recognize the pattern, and ability to build large-scale statistical models is a key facilitator of the building and construction industry 4.0 to process its digitized data.”

The decision to apply a Convolutional Neural Network was driven by the inherent characteristics of the dataset and the task at hand. Convolutional Neural Network (CNN) is known as a part of Artificial Neural Network (ANN), geared to work with data through grid-like structure. CNNs have proven to be one of the most effective approaches for tasks involving image recognition, classification and CV tasks. They mimic the visual cortex of the human brain, detecting local patterns such as edges, textures, and shapes before combining them into higher-level features.

There are three main types of layers in the architecture of CNN: convolutional layer, pooling layer and fully connected (FC) layer. In typical CNN, convolutional layer is followed by either another convolutional layer or a pooling layer, and the fully connected layer appears at the end of the network. Convolutional layer, as the core layer of CNN, uses filters and feature maps to extract information after input layer received the image data. In CNNs only neurons related to the filter are connected to the corresponding convolved neurons, this greatly reduces the number of connections, lowering memory usage and computational cost, making CNNs highly efficient for processing large inputs such as images, videos and audio. CNN architecture is determined by hyperparameters such as filter size, number of filters, padding and stride. Some of the most prominent CNN models have been proposed by various researchers over time are AlexNet, VGGNet, and ResNet. CNNs autonomously learn relevant features during training, compared to machine learning models, which often require manual feature extraction, thereby reducing the need for human intervention. This is especially valuable when working with construction site images, where environmental visual noise, lighting variability, and occlusions make feature engineering extremely challenging. As Munawar et al. (2022) emphasize, “Among the relevant techniques, CNN-based models have emerged as one of the most promising computer vision and DL methods for automatic learning and detection of image features.” This is reinforced by Shorten and Khoshgoftaar (2019), who note that CNNs, when trained on sufficient data, can achieve robust generalization even under challenging visual conditions.

Following the collection of the dataset described in Chapter 3, this chapter provides a comprehensive exploration of testing and evaluation process of CNN model applied to classify various stages of the study subject installation - drywall. Chapter 3 of the dissertation focuses on the data collection process essential for machine learning model training to classify the progress stages through installation process. Dataset was sourced from four real-world European projects under construction phases using the OpenSpace.ai platform, which holds 360-degree panoramic imagery mapped to 2D architectural plans and 3D models. Drywalls were selected as the target element due to their universality, visual distinctiveness and gradual installation process. Based on visible construction features such as framing, insulation, gypsum board layers and finishing painting, screenshots from 360-degree panoramic imagery were manually classified into three classes according to each stage — Class C (5–30% completion), Class B (30–80%) and Class A (80–100%). The data set was built in three attempts, expanding to 200 images per class with selective cropping applied to improve model focus and accuracy, captured sufficient diversity in lighting, angles and environmental visual noise to support robust model training.

While Chapter 3 focused on the rationale for dataset composition, this chapter shifts the emphasis toward testing strategies, iterative experimentation, model adjustments and performance evaluation. Testing is a fundamental transitional stage in machine learning between theoretical assumption and practical application. It enables to determine whether the model is capable of generalizing knowledge from training real-world construction data to unseen testing images and to critically assess the potential of AI-driven approach in AEC industry. Testing process was iterative, consisting of multiple attempts. Gradual improvement of results on each attempt was a result of lessons learned. This shows an important principle of AI-based research: models evolve through continuous refinement, do not achieve optimal performance in one step but through hyperparameter tuning and dataset expansion. The chapter documents these iterative steps in detail and positions them within the broader context of research on computer vision in construction.

In the case of this dissertation, the CNN was expected to differentiate between subtle progress stages: bare framing (Class C), partially completed drywall (Class B) and nearly or fully completed walls (Class A). These stages are visually similar, yet CNNs can extract fine-grained details such as joint compound, visible seams, or the presence of finishing paint. The architecture was therefore deemed the most suitable for this application.

Google Colab provides cloud-based access to GPU acceleration and model development, and testing were performed using it. This environment was chosen for its efficiency, ease of integration with TensorFlow libraries, and ability to run multiple experiments within reasonable computational timeframes. The base code was adapted from Tensorflow.org, ensuring that the model structure was built on a reliable, community-tested foundation. Modifications were progressively applied to tailor the architecture and training procedures to the specific dataset used in this research.

Evaluation of the CNN relied primarily on training accuracy and validation accuracy, plotted across epochs to track learning behavior. Accuracy is a straightforward and widely adopted metric for classification problems, providing an immediate indication of how many predictions were correct. However, for this research, validation accuracy is particularly critical, as it represents the model's performance on unseen data rather than memorized training examples. Overfitting was a persistent risk:

CNNs with small datasets can achieve nearly perfect training accuracy while performing poorly on validation sets.

For this dissertation, a validation accuracy of 70% was set as the target threshold. While modest compared to the near-perfect results sometimes reported in computer vision literature, this threshold is realistic and meaningful in the context of real-world construction monitoring. As Kropp et al. (2018) highlight in their study, drywall installation stages are difficult to distinguish visually due to low texture and uniform surfaces. Moreover, Shorten & Khoshgoftaar (2019) stress that environmental factors such as lighting and occlusion can drastically reduce classification reliability. Thus, reaching 70% validation accuracy demonstrates the model's practical potential for deployment in site monitoring applications.

4.1. Attempt 1 – Model with 100 images per class

The initial experiment employed 100 images per class, totaling 300. The used CNN consisted of two convolutional layers, followed by pooling and dense layers. No advanced augmentation techniques were applied, and dropout layers were not included. As expected, the model rapidly memorized the training set, achieving high training accuracy but peaking at only 56% validation accuracy after 5 runs. The lack of augmentation limited the model's exposure to variability, while the small dataset restricted generalization.

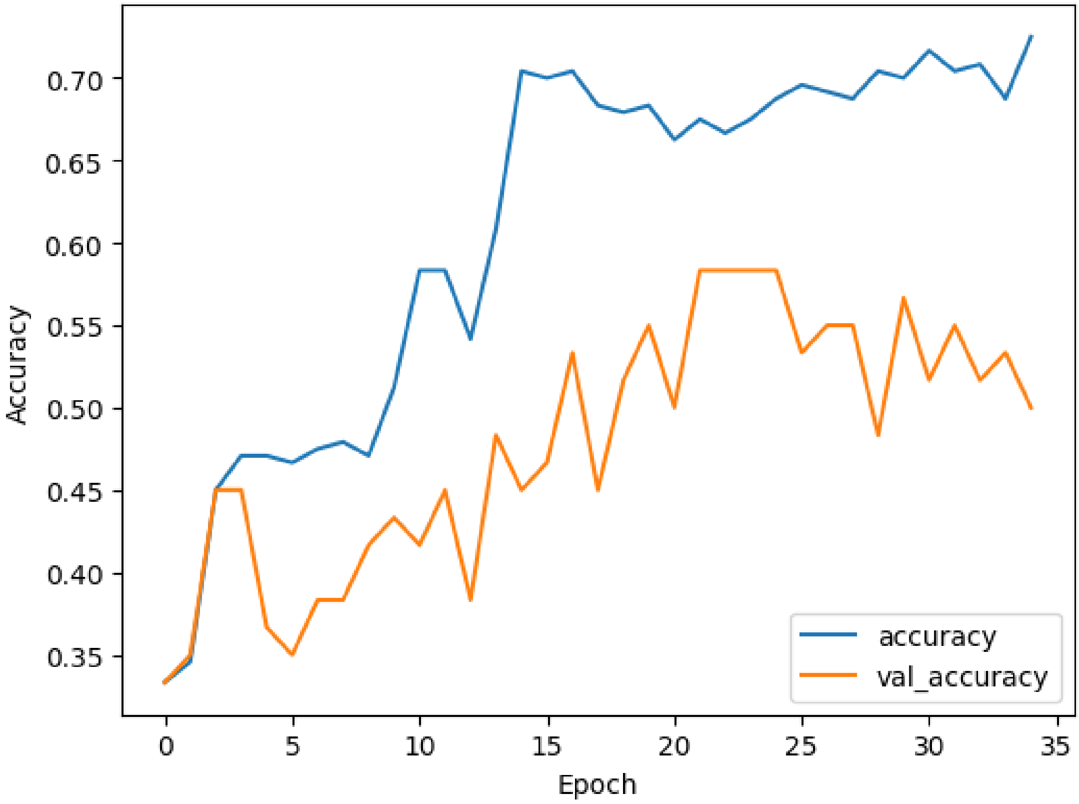

Figure 14 – Graph of training accuracy and validation accuracy versus epochs for 1st attempt

Performance of the 1st attempt graph (Figure 14) shows the validation accuracy (val_accuracy) peaks around 57% after epoch 20. During next epochs the training accuracy (accuracy) continues to rise, while the validation accuracy fluctuates and even drops. These curves show clear signs of overfitting. The model is performing well on the training data but fails to generalize on the validation data, leading to a large and growing gap between the two curves. The poor generalization is expected given the code for current attempt. It uses a basic ImageDataGenerator with no data augmentation and lacks Dropout layers in the model architecture.

ImageDataGenerator is a tool used for reading and loading image data from a specified directory path, categorizing them based on subfolder names and loading them into memory in batches. It splits the data into training and validation subsets and scales pixel values to a range between 0 and 1, dividing them by 255. Data augmentation is a technique that applies random, but realistic, transformations to training images to create new, modified, much larger and more diverse datasets to reduce overfitting, by learning more general and robust features which aren't focused on image orientation or position, improving its ability to generalize new, unseen data. Transformations may include rotation; horizontal or vertical shifting; shearing along an axis; random zooming; horizontal flipping. The Dropout layer is a regularization technique that randomly and temporarily "drops out" percentage of neurons during the training of a neural network, thereby it cannot rely on them to learn specific features. These layers are used to prevent overfitting, and force other neurons to pick up the slack, leading to a more robust and less co-dependent network.

Thus, the model is less sensitive to specific features and generalizes new data worse. The model, without these regularization techniques, memorizes specific training examples rather than learning general features, especially with a small number of images. Low accuracy of maximum 0.56 after 5 runs suggests the model is not powerful enough to learn the features from the data effectively.

4.2. Attempt 2 – Expanded 150 images dataset with augmentation

The main changes from attempt 1 to attempt 2 focus on improving the model ability to generalize, to learn features from the images and avoid overfitting because these were the issues of 1st attempt. The second attempt expanded the dataset up to 150 images per class (450 total), and applied basic augmentation techniques such as horizontal flips, random rotations and zooming, while dropout layers were added to reduce overfitting.

In attempt 1, the ImageDataGenerator only performed pixel normalization. In attempt 2, a variety of data augmentation techniques are implemented: rotation_range - to rotate the images either clockwise or counter-clockwise on a range in degrees; width_shift_range, height_shift_range - to randomly shift the images horizontally or vertically to a fraction of the total width or height; shear_range - a sort of random "tilting" of the image, creates a parallelogram shape from a rectangle; zoom_range - to randomly zoom into or out of the images; and horizontal_flip - to randomly flip images horizontally. The network architecture was changed to be deeper and more complex. In attempt 1, the model had four convolutional layers with an increasing number of filters (64, 128, 128, 256), attempt 2 has more filters (64, 128, 256, 512) this results in improvement of the model ability to extract more complex and hierarchical features from the images. Attempt 2 introduces Dropout layers after each convolutional block and after dense layers. This is a key reason the validation accuracy in Attempt 2 is higher and more stable than in Attempt 1. The patience parameter for EarlyStopping increased from 5 epochs in attempt 1 to 10 epochs in attempt 2. This gives the model more time to find a better minimum on the validation loss before stopping the training process.

This resulted in moderate improvements: validation accuracy increased to 64%. While still below the 70% target, this attempt demonstrated the value of augmentation and dataset size. It also suggested that further scaling would yield better performance.

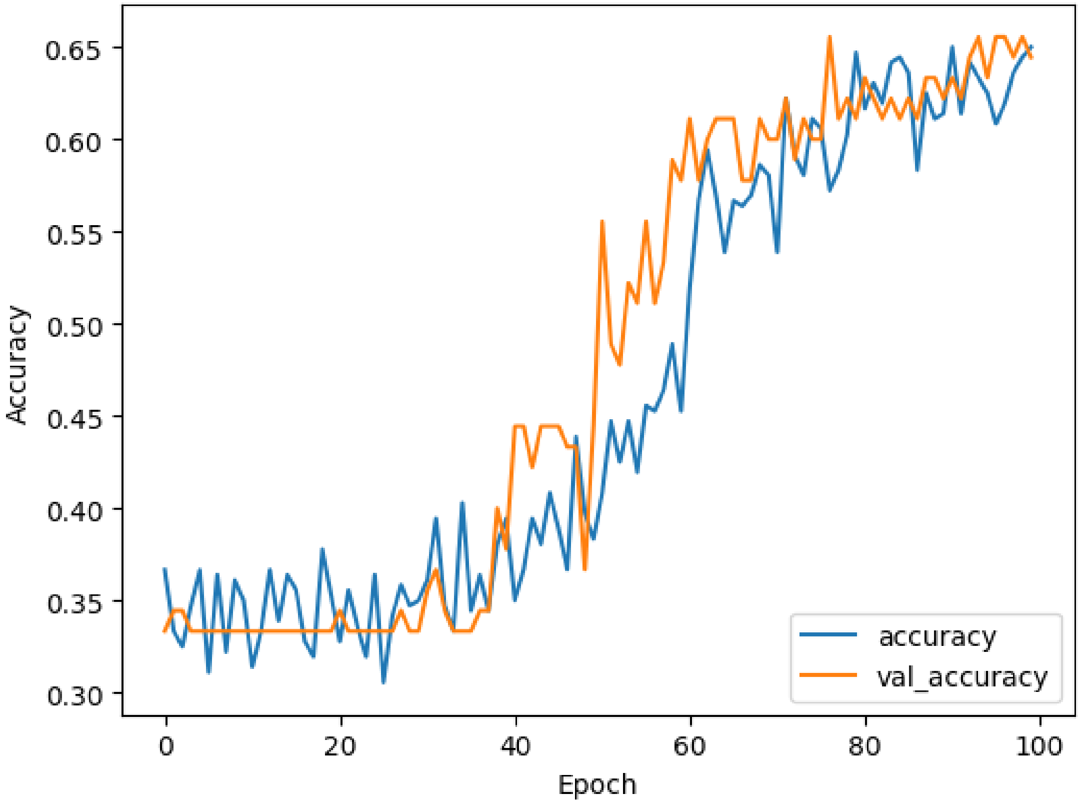

Figure 15 – Graph of training accuracy and validation accuracy versus epochs for 2nd attempt

The 2nd attempt graph (Figure 15) shows the model achieves a peak validation accuracy of about 64% by the end of training. The training and validation accuracy curves show more stable, albeit slightly jagged, increase over the epochs. The number of epochs used in each attempt is not a fixed value, rather, it is determined by the model complexity and the training process. In 2nd attempt instead of setting a high number of epochs, the code uses EarlyStopping function as an intelligent mechanism to automatically halt training when the model stops improving. EarlyStopping with a patience of 5 was implemented into the code in 1st attempt to stop training before overfitting became critical. The graph (Figure 14) confirms that the model reached its performance peak early at 35th epoch, and any further training would have been ineffective, although there are 50 epochs according to the code. The code for 2nd attempt became more complex with the addition of data augmentation and deeper architecture. Data augmentation effectively expanded the training set, which required more time for the model to learn. Consequently, the maximum epochs were increased to 100, and patience of EarlyStopping was raised to 10 to give the model more time to converge and to allow for small fluctuations in validation accuracy and then stop the training when no further progress was made. The gap between the two curves on the graph is much smaller than in the previous attempt. The attempt 2 code improved the model performance by introducing more complex architecture with additional convolutional layers and Dropout layers. Crucially, it also added extensive data augmentation for the training data. The smaller gap between the training and validation accuracy lines confirms that the model is no longer severely overfitting, creating more robust model that could better generalize new images. The increase in accuracy to 0.64 shows the positive effect of these changes.

4.3. Attempt 3 – 200 images dataset with advanced augmentation

The most meaningful change from attempt 2 to attempt 3 is the move from a custom-built CNN to a pre-trained model with transfer learning. This approach leverages knowledge from a massive dataset ImageNet through pre-trained MobileNetV2, which is a major reason for leap in effectiveness. Additionally, final attempt utilized the full dataset of 200 images per class - 600 images in total, moreover, photos in the dataset were cropped as described in Chapter 3. The range of data augmentation parameters (e.g., rotation_range, width_shift_range, zoom_range) and patience for EarlyStopping were increased compared to attempt 2.

Instead of building a model from scratch - a process that is often computationally expensive and requires very large dataset to be successful, attempt 3 uses a pre-trained MobileNetV2 model as a feature extractor that was already trained on millions of images and thousands of classes (e.g., different types of dogs, cars, planes, etc.). The "top" of the model is a set of fully connected (Dense) layers that take the high-level features learned by the convolutional base and use them to make these final classifications. The include_top=False parameter is used to load the model without its final classification layers. The idea is to leverage the features learned by MobileNetV2 from the large ImageNet dataset, which can significantly improve performance on smaller datasets. This provides a powerful set of pre-learned features that are highly effective for general image recognition. Retaining the core of the model, which consists of all the convolutional layers which are the "brain" of the network that has already learned to recognize and extract a wide range of useful visual features (like edges, shapes, and textures) from millions of images and then attaching new classification layers from study dataset on top of this pre-trained base. This process is called transfer learning, where you transfer the knowledge learned by a model on a large task to a new, specific task. It is a highly effective way to achieve remarkable results on a custom dataset, especially when you have a limited number of images. Custom layers were added on top of the pre-trained MobileNetV2 base model, including:

- GlobalAveragePooling2D is responsible for compressing the data from the convolutional part of the network. It takes each channel from the previous layer and calculates the average value of all its pixels, to reduce the spatial dimensions of the data while preserving important feature information - which helps reduce the number of parameters, which in turn reduces the risk of overfitting;
- Dense layer is a fully connected layer where each neuron is connected to all neurons in the previous layer. It takes the vector created by GlobalAveragePooling2D and processes it with 128 neurons, which in this layer performs a linear combination of the input data, and then applies to the result a relu (Rectified Linear Unit) activation function that helps the network learn faster and more efficiently. This hidden layer allows the model to learn more complex, non-linear combinations of features that have been extracted from the images;

- Dropout layer with a 0.5 rate to prevent overfitting;
- Final Dense layer with a softmax activation for classification. This layer has three neurons, one for each class (A, B, and C). The softmax activation is ideal for multi-class classification tasks as it directly outputs probabilities for each class, it transforms the output data into a probability distribution, where the sum of all probabilities equals 1.

Attempt 3 uses a different two-step training strategy process with fine-tuning - transfer learning, which involves a two-stage training process. At first step training the top layers was performed with 10 epochs to quickly adapt the new top layers to the data. The weights of the MobileNetV2 base model were initially frozen, and only newly added top layers were trained. At second step Fine-Tuning was executed. After top layers finished training, last 40 layers of the base model were unfrozen, and training was performed already on the entire model with 50 epochs according to code. The number of epochs decreased, compared to 2nd attempt, because of the timing of learning and testing. 2nd attempt performance took hours to complete a test while this attempt allowed it to be finished in 15-20 minutes. "Fine-tuning" allows pre-trained model weights slightly adjust to the new dataset with an incredibly low learning rate, often leading to better results. This allows the model to adapt its pre-learned features specifically for the new task without corrupting its already-strong feature extraction capabilities The training continues until EarlyStopping with a patience of 20 is triggered. One of obtained graphs for Attempt 3 demonstrates this by showing the training process stopping around epoch 30. This is far more efficient than setting a fixed high number of epochs, as it ensures you get the best performance without wasting time or risking a poorly generalized model. This configuration finally pushed validation accuracy above the 70% threshold, meeting the target benchmark and even outperformed.

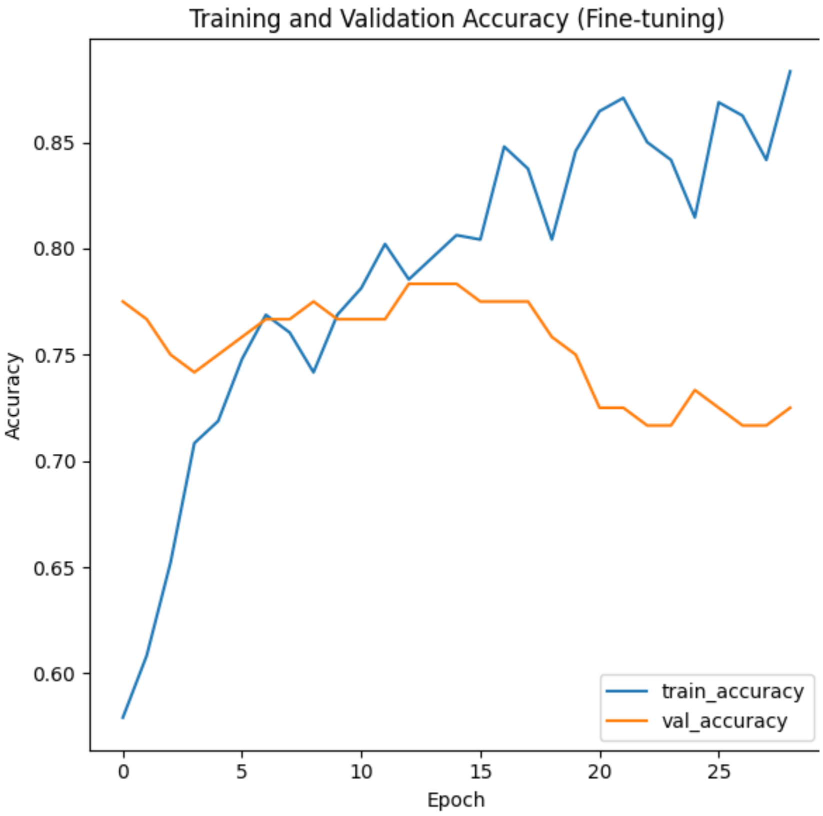

Figure 16 – Graph of training accuracy and validation accuracy versus epochs for 3rd attempt

The performance of 3rd attempt graph (Figure 16) shows a significant jump. The validation accuracy (val_accuracy) reaches almost 80%. This dramatic increase in performance is a direct result of increasing the amount of data and cropping, using transfer learning with the MobileNetV2 model, which leverages a powerful feature extractor pre-trained on a massive dataset. The model quickly adapts to the new image data through two-phase training process (training new layers first, then fine-tuning), resulting in a much higher accuracy of 0.775 (and with average accuracy after 5 runs – 0.755).

The graph shows how the validation accuracy decreases because of a classic and important reason - overfitting, which begins to appear at a certain stage of training. At the very beginning, the model quickly learns general, basic features from photographs that are useful for both training and validation data, so both curves grow in parallel. After about the 16th epoch, the model has already learned all the basic patterns. Further training makes it pay attention to more complex, but noisy and specific features that are critical for a certain class from the training set, and are also less obvious. The model begins to overfit the training set, rather than learn to generalize. Those "special" features that it finds do not work on the validation data, which contain slightly different examples. As a result, the accuracy on the training data continues to grow (since the model has perfectly overfitted the data), and the accuracy on the validation data begins to fall, because the model loses its ability to generalize. This is why the EarlyStopping function was integrated into the code - to monitor and stop the validation accuracy if it does not improve, but either drops or does not change. In this case, EarlyStopping was supposed to be triggered when the validation curve went down, to keep the model weights that were the best before the retraining started.

4.4. Discussion of Results

A principal component of this research was the application of convolutional neural networks to the collected dataset of drywall images. In the evaluation of model performance confusion matrices provided transparent and detailed representation of classification outcomes. The confusion matrices revealed areas in which the model struggled and showed how many times the model correctly and incorrectly classified each class. The diagonal values represent correct predictions, while off-diagonal values are incorrect ones. Through study evolution the confusion matrices reveal how the model performance improved with each code and data change.

The matrix (Figure 17) corresponds to the code with an accuracy of around 56% from 1st attempt, when the model predictions are inconsistent and often wrong. As it is seen, on this stage of development, the basic code and bad classified data, the model struggles to correctly identify Class A. It does not spread between all the classes inside Class A but clearly shows Class B with 11 misclassifications, which is worse than the spread numbers between all classes. The highest numbers are scattered across the matrix, concentrated on the diagonal only for class B and C. For example, the model predicts Class B a considerable number of times for Class C images, with 7 misclassifications. Which means that the model is struggling to recognize all the classes at this point of study with small number of correct classifications across the board.

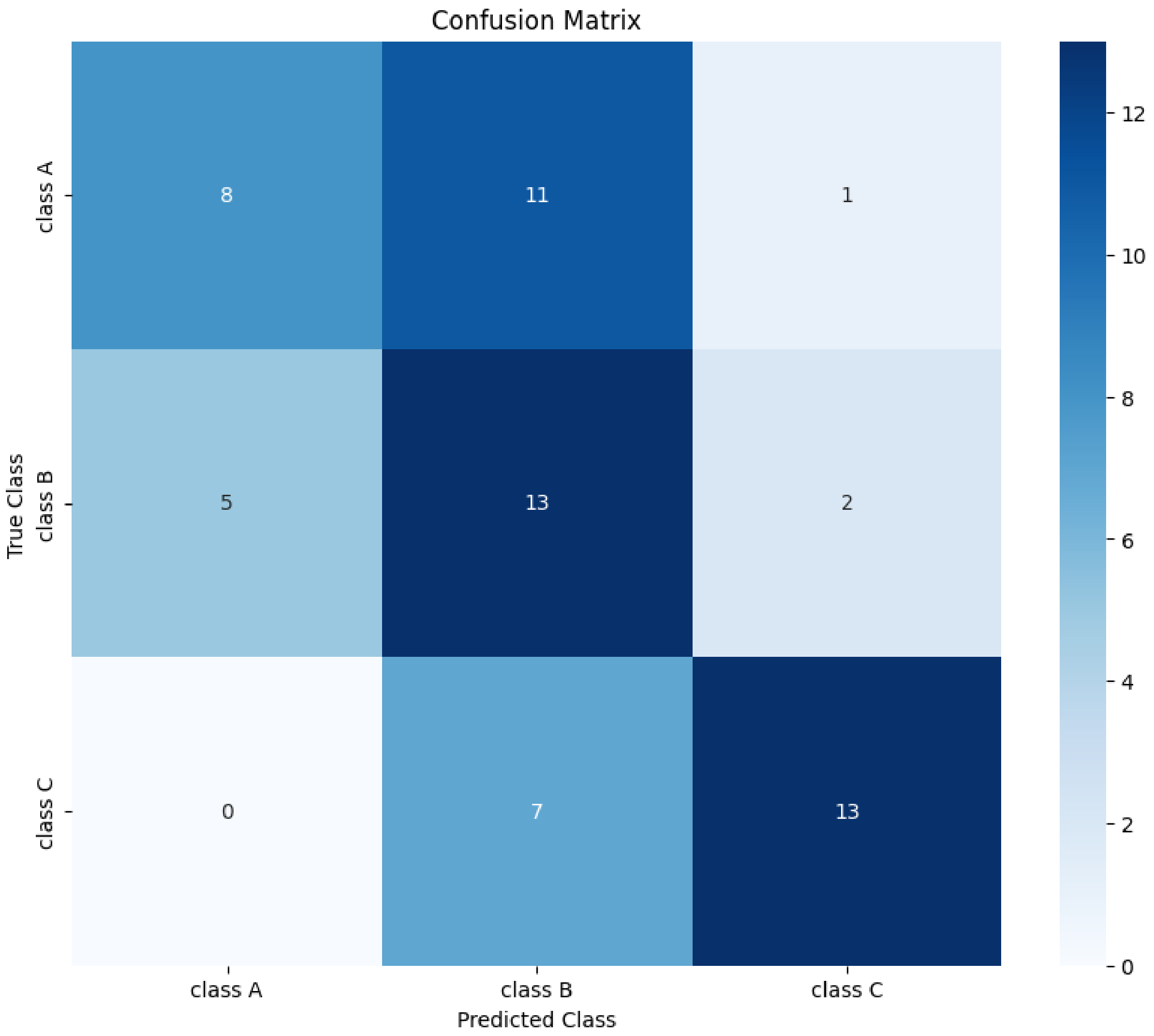

Figure 17 – Confusion Matrix 1st attempt

The matrix (Figure 18) corresponds to the code with an accuracy of around 64% from 2nd attempt, when the model still struggles with misclassifications. It is not clearly noticeable that the matrix shows an improvement in confusion after data expansion and code improvement, significant problems remain. The model still cannot clearly recognize Class B, placing the substantial number of photos into Class A and the smallest number to the actual Class B, thereby particularly predicting the class wrongly. This is also the result of unclear classification inside that class data. The highest numbers are concentrated on the diagonal except Class B. The model misclassifies Class A as Class C for 2 times, classify the actual class 28 times, and Class C as Class A for 2 times, as Class B for 1 with right classification of 27 times. Since Class A and Class C are more organized into visual recognition, such confusion indicates that the model manages to distinguish between all the classes effectively, despite the added data augmentation and Dropout layers into the code. Although Class B remains hard to classify, because of the mess in the dataset, this is a frequent problem with custom CNN on limited datasets and requires further developing of the code and working with data.

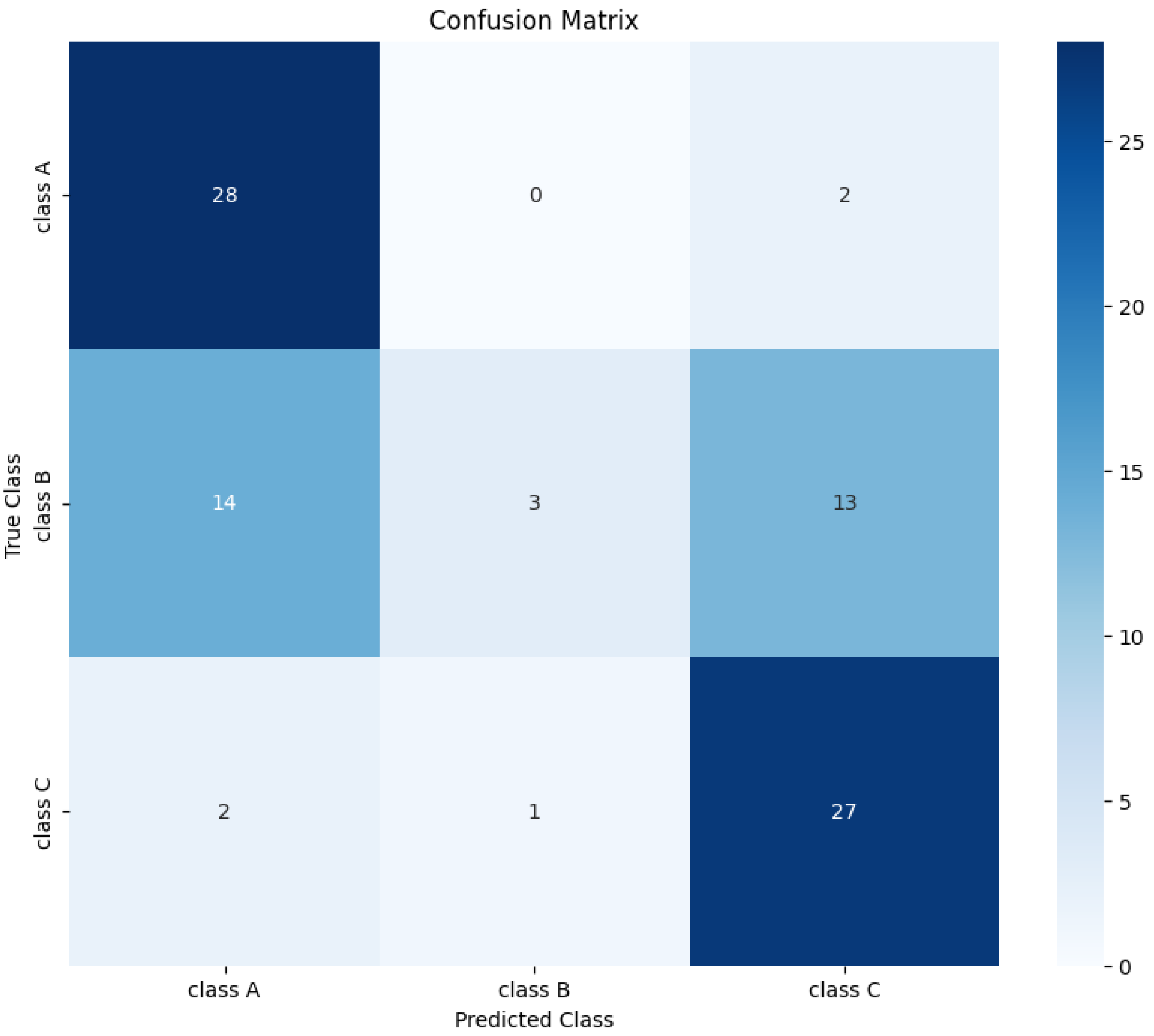

Figure 18 – Confusion Matrix 2nd attempt

The matrix (Figure 19) corresponds to the code with the highest accuracy of around 77.5% from 3rd attempt, when the stated performance is much better, some misclassifications still occur. This matrix shows a dramatic improvement into Class B classification after more balanced dataset and improvements in the code. It shows the biggest number – 23 – on a particular class into Class B data and spread the mistakes between other classes almost equally, which is the sign of better dataset. The highest numbers are now again on the diagonal, indicating high rate of correct classifications. The model correctly identifies 36 images of Class A and 34 Class C images with the same trend of misclassification of those classes with each other which is the result of dataset classification inside of those classes. It is still possible the photos quality or the stage difference confuse the model – in Class A there are bunch of photos with visible upper sheets and jointing of them which can be considered by CNN as structured frame stage (Figure 20), and Class C includes bunch of photos with upper plasterboard which can be consider as jointing stage (Figure 21). This matrix indicates that the transfer learning approach with MobileNetV2 can influence the model performance but the model to learn and distinguish between the different classes with much greater accuracy, but the data quality and clearness of classification are the same important.

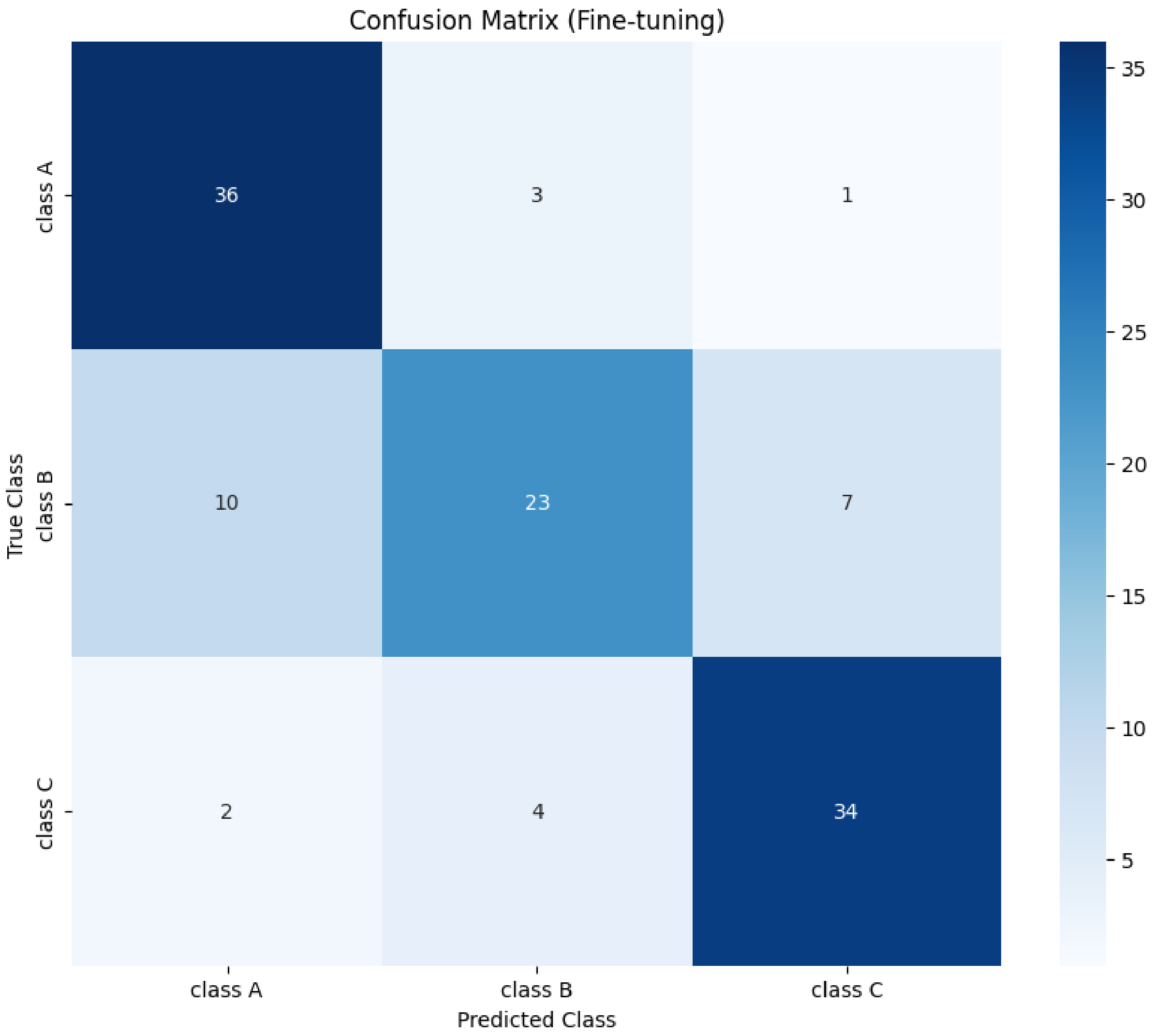

Figure 19 – Confusion Matrix 3rd attempt

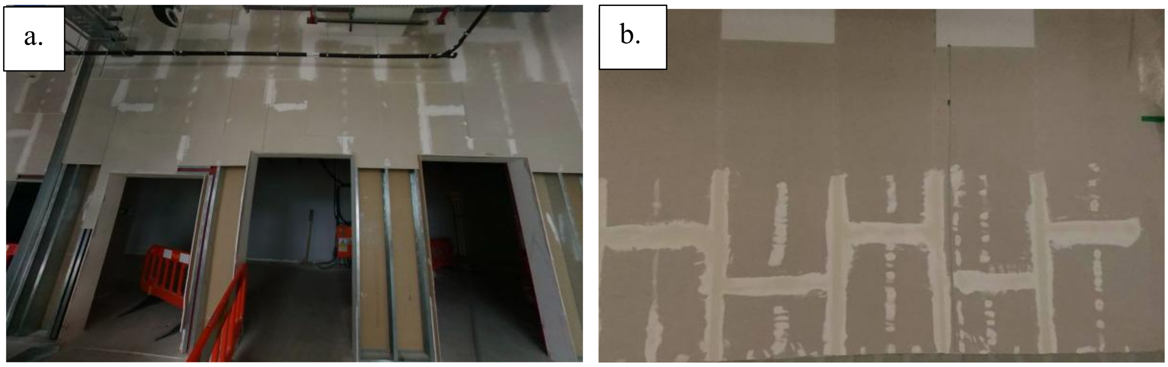

Figure 20 – Class C (a.) and Class A (b.) possible misclassifications examples

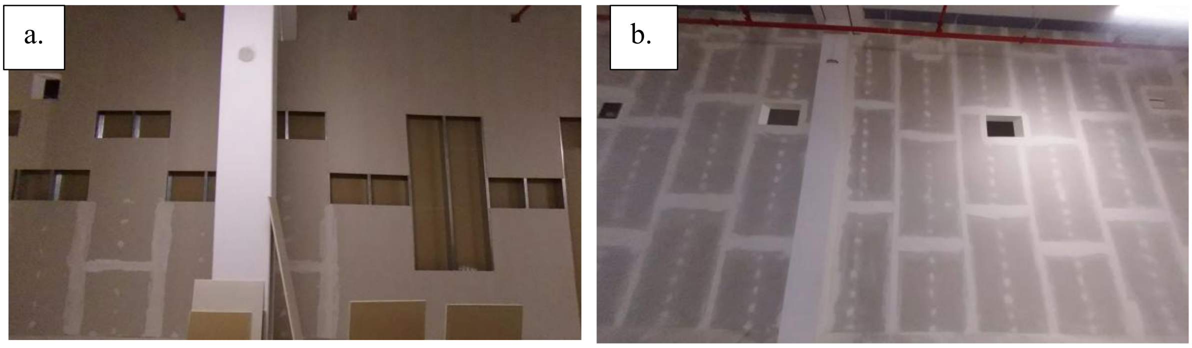

Figure 21 – Class C (a.) and Class A (b.) possible misclassifications examples

Analysis of confusion matrices clearly demonstrates how the model performance evolved from one attempt to next one, confirming the findings from the accuracy graphs. Matrices show that as changes were made to the code and the amount of data was increased, the model became a little more capable of generalization and more confidently distinguish between classes.

Matrices show progression in model performance and code developing. The initial model was inconsistent and prone to errors. The second model, with regularization and data augmentation, showed slight improvement but still struggled with a clear misclassification bias, which in any case showed the direction for future development. The final model, which used transfer learning, demonstrated a significant leap in accuracy and confidence, correctly classifying many of the images and confirming the effectiveness of that approach. However, the results are still needed to be developed and observed.

Misclassifications were most observed in Class B (Figure 22) where visual similarities between adjacent stages blurred the distinction even for a human eye because construction is a highly multi-threaded process, so it is even difficult for a person sometimes to determine to which class a particular recorded state of the object belongs to the current stage. The findings from matrices underscore the limitations of work with real-world data: environmental visual noise, lighting variability, occlusions, real progress workflow and inconsistent capture routes inevitably affected the results. Nevertheless, these findings contribute to further improvements of the model and contribute to future research.

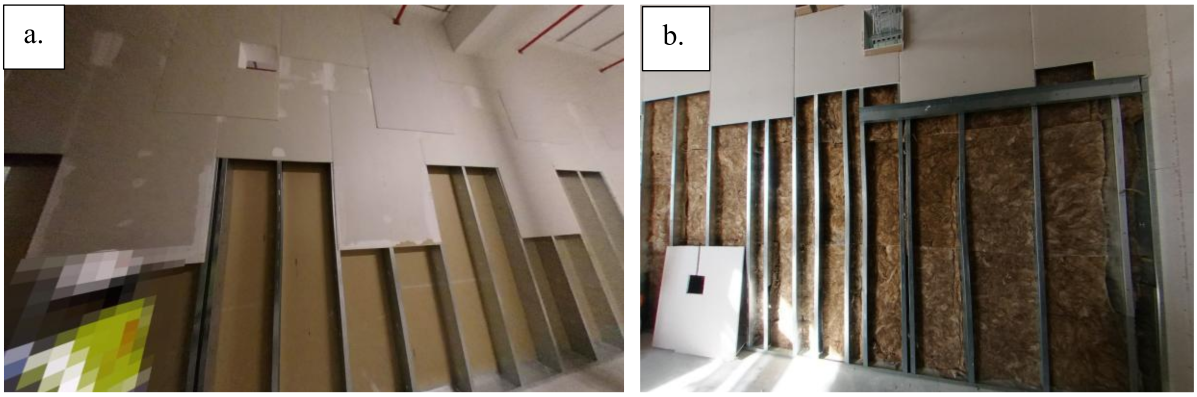

Figure 22 – Class C (a.) and Class B (b.) possible misclassifications examples

From a methodological side, confusion matrices provided more than an accuracy score, but a deeper understanding of error distribution and category-specific weaknesses of the model, in some way the thinking process of the model. For example, the confusion of the model partially completed walls (Class B) with more advanced finishing stages (Class A) highlights the need for further dataset expansion and augmentation, also increase in quality of photos since the Class A includes smooth surfaces but without visible drywall sheets and Class B includes smooth surfaces but with visible sheets, so the difference between surfaces is highly fine (Figure 23). As well as a solution could be potentially introducing additional classes or subcategories to reflect the gradual nature of construction progress. Such refinements could substantially enhance classification precision but also require more data.

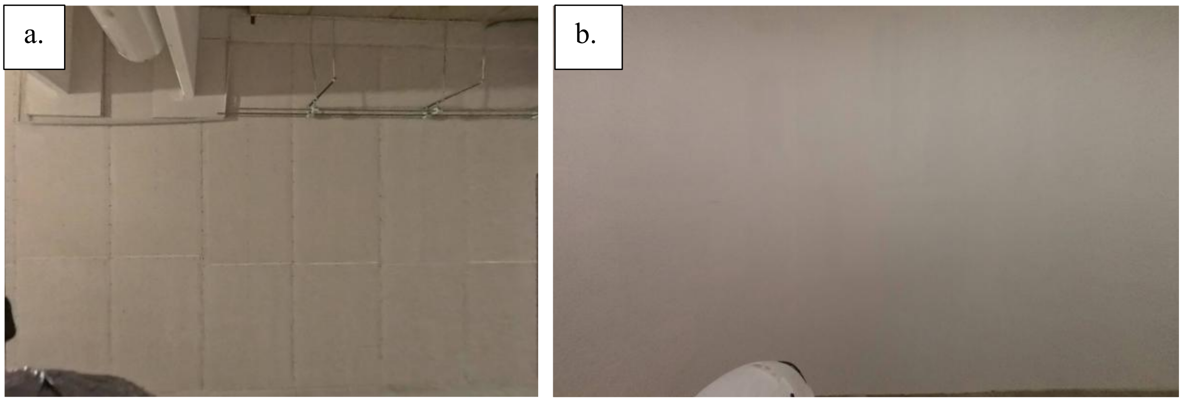

Figure 23 – Class B (a.) and Class A (b.) possible misclassifications examples

The performance of CNN model across three experimental attempts demonstrates both challenges and opportunities involved in applying deep learning to real-world photos of construction. The classic problem of overfitting in the baseline model in 1st Attempt highlighted with characteristics of rapid grow of the training accuracy, approaching near-perfect levels, while validation accuracy stagnated at around 56%. The divergence between training and validation accuracy curves, approximately 20%, clearly illustrates that the model memorized the training data rather than learnt generalizable features for validation. Such behavior is common occurrence for CNNs, training on small datasets, and it underlines the importance of both size and diversity of a dataset in achieving meaningful performance.

The second attempt showed incremental improvements when the dataset was expanded to 150 images per class and code was modified with data augmentation. Although the validation accuracy rose to 64%, the model still struggled with generalization, because of small size of dataset. Importantly, this attempt confirmed that strategic augmentation of the data — such as flipping, rotating or zooming — was effective in reducing overfitting, as the graph shows the gap between training and validation accuracy at the end of training became smaller and the curve lines repeated each other. The modest gain suggests that while augmentation plays an essential role, it cannot fully compensate for the limitations in size of dataset. Still, the improvement demonstrated a trajectory toward a more robust model.

The final third attempt, which used the complete dataset of 600 more prepared photos - classes were more organized and balanced inside, some photos were cropped to highlight the study object images - in total augmentation improvements was the most successful. Validation accuracy finally surpassed 70%, reaching the threshold defined as the benchmark for this research. The result is not only validated the methodological approach but hereby confirmed that with systematic refinement, CNNs can be reliably applied to classify drywall installation stages, the real-world images of structures in general. While the final model did not reach extremely prominent level of accuracy (90–95%), often reported in laboratory-based studies, its performance is highly commendable given the complexity and variability of real-world site conditions. In fact, the ability to achieve over 70% accuracy on noisy, heterogeneous and low-quality images is itself evidence of the practical robustness of CNN approach.

Another key point from the performance analysis is the relationship between dataset preparation and model outcomes. Each step of expansion and augmentation directly correlated with measurable improvements. This highlights iterative and experimental nature of machine learning: progress does not occur in a single breakthrough, but through systematic refinement, informed attempts and lessons learn, and careful adjustment of both data and code architecture.

Placing these results in context with related research makes it able to underscore both strengths and limitations of this dissertation approach. Kropp et al. (2018) in their study managed precision above 95% in detecting drywall installation progress. However, their experiments were conducted under tightly controlled conditions (when construction activities were either completed or temporarily paused to document progress) on construction sites inside buildings during inspection phases. Progress was captured in real-world conditions with shadows, clutter, uneven lighting and partial obstructions. They used a monocular camera system—meaning a standard digital camera without stereo capabilities with high resolution (4288×2848 pixels). The camera was manually moved through the space to capture certain processes. The operator had to be mindful of how the movement would affect trajectory reconstruction algorithms. By contrast, data for this dissertation was captured from general inspection photos of the buildings without focusing on one or another structure. Operator was not aiming to film by a particular trajectory and did not consider further study. Work on the site was not stopped specifically for the survey as we can see by captured workers. The fact that the model still achieved 70% validation accuracy under these conditions highlights its practical applicability from even not prepared data. Braun et al. (2020) advanced the field by integrating semantic BIM data with Mask R-CNN, achieving high accuracy in detecting structural elements with an F1 score of 0.92. Their work, while technically impressive, required well-structured, vast (5000 photos) datasets and computationally intensive workflows with photos to create a point cloud previously, which limit immediate scalability for everyday construction monitoring. The present study, by contrast, demonstrates that a lighter CNN architecture trained on relatively small, noisy datasets without additional steps to analyze and preparation of photos can still yield actionable insights, making it more adaptable for real-time, on-site use. Deng et al. (2019) achieved 91% accuracy in monitoring tile installation using 1,000 images. Tiles, however, offer distinct texture and visual contrast with clear geometry patterns compared to the more uniform, mate surfaces of drywall, making the classification problem comparatively easier. Additionally, the progress of laying is more gradual and consistent than the drywall installation. This dissertation's focus on drywall, where subtle differences between stages are difficult to detect, demonstrates progress in tackling a more visually ambiguous problem.

Taken together, these comparisons show that while absolute accuracy levels achieved here are lower than some of the controlled or large-scale studies, the contribution of this dissertation lies in robustness under realistic conditions. The ability to achieve 70% accuracy with a small, imperfect dataset demonstrates the feasibility of deploying CNNs in real projects without the need for massive annotation or heavily curated datasets. This positions the research as a bridge between highly experimental work and scalable, practical applications.

Moreover, this dissertation adds value by emphasizing the practical integration potential of CNNs with existing BIM/4D platforms. Rather than focusing solely on technical optimization, it demonstrates how models can be embedded into workflows to deliver semi-automated progress reports. This represents an essential step toward closing the gap between experimental research and real-world adoption in the AEC industry.

This page is intentionally left blank

5. CONCLUSION

5.1. Outputs and Discussion

This dissertation sets out to investigate the integration of Artificial Intelligence and Machine Learning techniques into 4D Building Information Modeling for the purpose of automating construction progress monitoring. By focusing on the recognition of drywall installation stages through computer vision, the study has combined theoretical research with practical experimentation, providing both conceptual insights and empirical validation.

The first key contribution lies in confirming the relevance of AI-enabled automation in the Architecture, Engineering and Construction (AEC) sector. As the literature review has demonstrated, while BIM and 4D modeling are already well-established tools, AI and ML are still in the experimental or limited use stage but nevertheless show the potential of exponential increase of BIM and 4D usage when combined with machine learning and visual monitoring. The developed approach has shown that real-world construction images, despite their complexity, noise, and inconsistencies, can be systematically utilized to train models capable of classifying progress stages with promising accuracy.

Despite these challenges, the relatively low rate of false negatives is particularly important for potential practical application, as it ensures that completed work is not mistakenly recorded as incomplete — a factor that directly influences reporting accuracy, billing processes and project management decisions. In case of more wide classification the misunderstanding between close-related classes may have less effect rather than not recognizing a stage at all. This reliability emphasizes the potential future applied value of the approach for adoption within industry workflows.

The broader significance of the findings is twofold: the study contributes to the growing body of knowledge on AI-BIM integration, illustrating how computer vision and deep learning can extend the role of 4D and 5D BIM from visualization into automated process validation, tracking and prediction; it provides a proof of concept for semi-automated progress monitoring, offering a pathway toward reducing reliance on labor-intensive manual inspections, documenting and enabling more timely, evidence-based decision-making in planning, budgeting, procurement, logistic and management.

Nevertheless, there are limitations that need to be considered. The collected dataset remains relatively small and project-specific, uncapable to teach even a baseline CNN model with high accuracy. Expanding the scope to include larger and more diverse datasets across multiple construction typologies will be essential to improving model generalization. The integration of multisource data, such as sensor inputs, LiDAR or schedule-linked metadata — could enhance robustness and enable more comprehensive automation. However, it is still relevant to observe will the additional data be helpful together with the collected dataset or even without it.

In conclusion, the dissertation has achieved its initial objectives: the relevance and feasibility of AI integration in 4D BIM for progress monitoring has been identified, a CNN-based classification model has been upgraded and tested, evaluating its performance through graphs and confusion matrices, and has been critically reflected on both its successes and shortcomings. The results demonstrate that implementation of such an integration into construction is not only possible but also beneficial, while simultaneously outlining clear directions for refinement, development and investment in the field.

Ultimately, this work lies in its dual contribution developing a proof of concept proposed by industry partner BIMMS: it validates the potential of AI to support more efficient, accurate, and proactive construction management, while also highlighting the challenges and open questions that remain. The collaboration of BIMMS was necessary for the development of this investigation, providing guidance and support from its team throughout the process, as well as required data from the construction cases. Confusion matrices and graphs, as both evaluative and diagnostic tools, have played a pivotal role in the analysis, providing evidence that the developed approach is a promising step forward, though not yet definitive. As such, this dissertation should be seen not as a final statement, but as a foundation for future research which will build upon these findings to move closer toward fully automated, intelligent and sustainable construction practices.

5.2. Future Developments

Based on the outcomes, the concept of applying machine learning to construction progress monitoring is not only validated but also highlight several directions for further development and discussion. Those are critical to ensuring that methodology evolves into a scalable, reliable, and widely applicable solution or give ground to novel research for the Architecture, Engineering and Construction (AEC) industry.

1. **Expansion of datasets and improved generalization**
   The most immediate step to developing future work is the collection, processing and organization of larger, more diverse datasets. The current dataset, while sufficient for proof of concept still with low accuracy, remains limited in size and variety, which constrains the model ability to generalize to new projects, structures and environments. Incorporating images from different construction typologies and structures, climates, with obstacles and ideal environments, timing and lighting conditions would substantially increase robustness. Additionally, the use of synthetic data generation and advanced augmentation techniques could help to overcome data scarcity and improve balance across classes. But it is needed to consider the influence of synthetic data on the results.
2. **Refinement of classification categories**
   The present study classified progress into three broad stages (Classes A, B and C). However, construction processes are inherently more gradual and complex, so the current stage of a structure is hard to identify even to a human, which is the main reason in report mistakes. For instance, drywall is two-sided element and if from one side it may be included into A class then from another side it still may be C class. Future research may explore more granular classification schemes, subdividing stages into finer categories or using hierarchical labeling systems, which could be developed on initial stages of a project and be included into Building Execution Plan (BEP). Such refinements could reduce confusion between adjacent stages, as evidenced by the misclassifications revealed in the confusion matrices, and provide stakeholders with more precise progress information.
3. **Integration of multisource and sensor data**
   Visual data alone, while powerful, may not capture the full complexity of construction progress. Combining computer vision outputs with other data sources—such as LiDAR scans, IoT sensor

streams, schedule-linked metadata —would enable a more comprehensive understanding of on-site conditions, from different perspectives, angles and providing more details for training to enhance accuracy. One of the issues with installation of IoT into construction sites is not all the time convenient and possible in this environment.

4. **Embedding models into 4D BIM and Digital Twin platforms**
   Trained models seamless integration into existing BIM environments and Digital Twin frameworks is a long-term aim. Such integration would allow automatically update project schedules and reports, generate real-time progress dashboards and issue alerts for deviations or delays based on captured classification results of any object in real-time. In this way, machine learning models could evolve from stand-alone prototypes into core components of dynamic decision-support systems, enhancing collaboration across project teams and providing transparent reporting for clients.
5. **Optimization of computational performance**
   Although convolutional neural networks have proven effective, their deployment in real-world construction projects requires consideration of computational cost and scalability. Future work could explore lightweight model architectures suitable for edge devices or cloud-based distributed processing, ensuring that the technology can be adopted in both large-scale infrastructure projects and smaller residential sites without prohibitive hardware requirements.
6. **Human-in-the-loop systems and practical adoption**
   While automation is a long-term goal, the current state of technology suggests that hybrid approaches—where human expertise complements algorithmic predictions—are the most realistic near-term solution. Developing interfaces that allow project managers to review, validate, and adjust model outputs will increase trust, facilitate adoption, and generate new annotated data to further improve accuracy. This human-in-the-loop paradigm bridges the gap between academic prototypes and operational industry tools.
7. **Contribution to sustainable and resilient construction practices**
   Finally, as global priorities shift toward sustainability and resilience, future work should investigate how AI-driven progress monitoring can contribute to reducing delays, minimizing resource waste, and improving lifecycle efficiency. By linking automated monitoring to sustainability indicators (e.g., energy use, material efficiency, or carbon emissions), the methodology can support not only productivity but also environmental responsibility in the construction sector.

Taken together, these directions outline a research agenda that builds on the foundations laid by this dissertation. By expanding datasets, refining classifications, integrating multimodal inputs, and embedding models into BIM and Digital Twin environments, future work can transform this study from a proof of concept into a fully operational framework. The ultimate vision is a construction industry where real-time, AI-driven monitoring enhances efficiency, reduces risk, and supports both economic and environmental sustainability.

This page is intentionally left blank

REFERENCES

Ajirotutu R., Adeyemi A., Ifechukwu G, Ohakawa T., Iwuanyanwu O., Garba B. (2024). Exploring the Intersection of Building Information Modeling (BIM) and Artificial Intelligence in Modern Infrastructure Projects. Journal of International Journal of Science and Research Archive, 2414–2427. IJSRA. https://doi.org/10.30574/ijsra.2024.13.2.2421

Amer F., Golparvar-Fard M. (2021). Modeling dynamic construction work template from existing scheduling records via sequential machine learning. Journal of Advanced Engineering Informatics 4. Elsevier Ltd. https://doi.org/10.1016/j.aei.2020.101198

Amer F. , Jung Y., Golparvar-Fard M. (2021). Transformer machine learning language model for auto-alignment of long-term and short-term plans in construction. Journal of Automation in Construction 132. Elsevier B.V. https://doi.org/10.1016/j.autcon.2021.103929

Asadi K., Ramshankar H., Noghabaei M., Han K. (2019). Real-Time Image Localization and Registration with BIM Using Perspective Alignment for Indoor Monitoring of Construction. Journal of Computing in Civil Engineering. ASCE. DOI: 10.1061/(ASCE)CP.1943-5487.0000847

Baduge S., Thilakarathna S., Perera J., Arashpour M., Sharaf P., Teodosio B., Shringi A., Mendis P. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. Journal of Automation in Construction 141. Elsevier B.V. https://doi.org/10.1016/j.autcon.2022.104440

Barbosa A., Costa D. (2022). Use of BIM and visual data collected by UAS and 360° camera for construction progress monitoring. In Earth and Environmental Science. IOP Publishing. doi:10.1088/1755-1315/1101/8/082007

Braun A., Tuttas S., Borrmann A., Stilla U. (2020). Improving progress monitoring by fusing point clouds, semantic data and computer vision. Journal of Automation in Construction 116. Elsevier B.V. https://doi.org/10.1016/j.autcon.2020.103210

Daniotti B., Gianinetto M., Della Torre S. (2020). Digital Transformation of the Design, Construction and Management Processes of the Built Environment. Springer Open. https://doi.org/10.1007/978-3-030-33570-0_1

Deng H., Hong H., Luo D., Deng Y., Su C. (2019). Automatic Indoor Construction Process Monitoring for Tiles Based on BIM and Computer Vision. Journal of Journal of Construction Engineering and Management. ASCE. DOI: 10.1061/(ASCE)CO.1943-7862.0001744

Dolhopolov S., Honcharenko T., Terentyev O., Savenko V., Rosynskyi A., Bodnar N., Alzidi E. (2024). Multi-Stage Classification of Construction Site Modeling Objects Using Artificial Intelligence Based on BIM Technology. In The 35th Conference Of FRUCT Association. IEEE. DOI: 10.23919/FRUCT61870.2024.10516383

Fang X., Li H., Wu H., Fan L., Kong T., Wu Y. (2023). A fast end-to-end method for automatic interior progress evaluation using panoramic images. Journal of Engineering Applications of Artificial Intelligence. Elsevier Ltd. https://doi.org/10.1016/j.engappai.2023.106733

Fazeli A., Banihashemi S., Hajirasouli A., Mohandes S. (2022). Automated 4D BIM development: the resource specification and optimization approach. Journal of Engineering, Construction and Architectural Management. Emarald Publishing Limited. https://doi.org/10.1108/ECAM-07-2022-0665

Glinka S. (2024). Sat4BIM4D—the concept of using satellite remote sensing to monitor construction progress in conjunction with BIM. Journal of Reports on Geodesy and Geoinformatics. De Gruyter. DOI: 10.2478/rgg-2024-0023

Himeur Y., Elnour M., Fadli F., Meskin N., Petri I., Rezgui Y., Bensaali F., Amira A. (2023). AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Journal of Artificial Intelligence Review. Springer Nature. https://doi.org/10.1007/s10462-022-10286-2

Iordanova I., Valdivieso F., Filion C., Forgues D. (2020). SCHEDULE OPTIMIZATION OF A LARGE HOSPITAL PROJECT – 4D BIM STARTING WITH THE DEMOLITION. In 28th Annual Conference of the International Group for Lean Construction (IGLC28). . doi.org/10.24928/2020/0048

Jung Y., Hockenmaier J., Golparvar-Fard M. (2024). Transformer language model for mapping construction schedule activities to Uniformat categories. Journal of Automation in Construction 157. Elsevier B.V.. https://doi.org/10.1016/j.autcon.2023.105183

Khondoker T., Hossain M., Saha A. (2024). 4D BIM integrated optimization of construction steel bar procurement plan for limited storage capacity. Journal of Construction Innovation. Emarald Publishing Limited. DOI 10.1108/CI-12-2022-0310

Kropp C., Koch C., König M. (2018). Interior construction state recognition with 4D BIM registered image sequences. Journal of Automation in Construction 86. Elsevier B.V. https://doi.org/10.1016/j.autcon.2017.10.027

Moeisra D., Juliastuti, Insyira A., Haripriambodo T. (2023). Accelerating the middle – rise building construction duration using BIM 4D, a case study in central Jakarta. In The 7th International Conference on Eco Engineering Development 2023 (ICEED 2023). IOP Publishing. doi:10.1088/1755-1315/1324/1/012022

Munawar H., Ullah F., Heravi A., Thaheem M., Maqsoom A. (2022). Inspecting Buildings Using Drones and Computer Vision: A Machine Learning Approach to Detect Cracks and Damages. Journal of Drones . MDPI. https://doi.org/10.3390/drones6010005

Ohakawa T., Adeyemi A., Okwandu A., Iwuanyanwu O., Ifechukwu G. (2024). Digital Tools and Technologies in Affordable Housing Design: Leveraging AI and Machine Learning for Optimized

Outcomes. Journal of International Journal of Engineering Inventions, 255-264. . DOI: 10.13140/RG.2.2.15481.07523

Pan Y., Zhang L. (2022). Integrating BIM and AI for Smart Construction Management: Current Status and Future Directions. Journal of Archives of Computational Methods in Engineering. Springer Nature. https://doi.org/10.1007/s11831-022-09830-8

Pei Y., Huang Y., Zou Q., Zhang X., Wang S. (2019). Effects of Image Degradation and Degradation Removal to CNN-Based Image Classification. Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence. IEEE. DOI: 10.1109/TPAMI.2019.2950923

Sanon S., Boton C. (2024). Comparative study of three 4d simulation software. In Creative Construction Conference. Budapest University of Technology and Economics & Diamond Congress Ltd. https://doi.org/10.3311/CCC2024-157

Shinde Y., Lee K., Kiper B., Simpson M., Hasanzadeh S. (2023). A Systematic Literature Review on 360° Panoramic Applications in Architecture, Engineering, and Construction (AEC) Industry. Journal of Information Technology in Construction. . DOI:10.36680/j.itcon.2023.021

Shorten C., Khoshgoftaar T. (2019). A survey on Image Data Augmentation for Deep Learning. Journal of Big Data. Springer Open. https://doi.org/10.1186/s40537-019-0197-0

Yin J., Luo S., Rui J., Yao J., Zhang F., Fu B., Kassem M. (2024). BIM-based detection and optimization of spatial-temporal clashes in underground pipeline construction. Journal of Automation in Construction 166. Elsevier B.V.. https://doi.org/10.1016/j.autcon.2024.105616

Zhang L., Li Y., Pan Y., Ding L. (2024). Advanced informatic technologies for intelligent construction: A review. Journal of Engineering Applications of Artificial Intelligence 137. Elsevier Ltd. https://doi.org/10.1016/j.engappai.2024.109104