Building and Testing a Digital Twin Using Machine Learning Within a Mining Environment in South Africa

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

A dissertation submitted in fulfilment of the requirements for the degree

Master of Science in Engineering.

Author: Unerson Govender (1110925)

Supervisor: Prof. Turgay Celik

Co-Supervisor: Prof. Bekir Genc

School of Electrical and Information Engineering,

University of the Witwatersrand, Johannesburg

September 2025

Abstract

This study presents the development and testing of a Digital Twin (DT) for the crushing circuit in a South African mining environment, using advanced machine learning (ML) methods. The key objective was to select the best ML algorithm for the crushing circuit DT development. Data was collected over a 12-month period from sensors and Supervisory Control and Data Acquisition systems across key components of the crushing circuit, then pre-processed and integrated for analysis.

The methodology involved defining the requirements and establishing data exchange protocols to select, train and validate ML models. Multiple approaches were evaluated, including traditional statistical methods (Autoregressive Integrated Moving Average) and advanced neural networks, including the Gated Recurrent Units (GRUs), Long Short-Term Memory (LSTM), and Transformer models. Hyperparameter optimisation using frameworks like Optuna was employed to fine-tune model performance, with predictive accuracy assessed via the coefficient of determination, root mean square deviation, and mean absolute error metrics. The results highlighted that the DT reliably captures equipment behaviour, with Recurrent Neural Network-based models (LSTM and GRU) achieving high predictive accuracy. The study concluded that the GRU model was the best ML approach and provides the mining industry with an efficient DT to predict the percentage product material under 80% of the target particle size distribution, which is crucial to maximise throughput and downstream milling efficiency. This paves the way for future advancement of DT in the mining industry.

The complexity of this work lies in its comprehensive integration and comparative analysis of diverse ML models to develop a high-fidelity DT tailored for a mining application, an area with limited prior research. By addressing challenges unique to the South African mining context, this study offers a robust DT framework that can be used for DT development in the crushing circuit.

Future work is recommended to expand DT applications to other mining processes and further validate the framework under real-time operational conditions. These advancements will contribute to sustainable mining practices and accelerate the adoption of digital transformation initiatives within the industry.

Declaration

I, Unerson Govender (Student Number: 1110925), declare that this dissertation is my own, unaided work, except where otherwise acknowledged. It is being submitted for the degree of Master of Science in Engineering to the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination to any other university.

Signed this 23 September 2025

U. Govender

Acknowledgements

I humbly express my compelling gratitude to those who have guided, encouraged, and inspired me throughout my masters studies. First and foremost, I humbly thank the Supreme Personality of Godhead, Lord Krishna, whose grace, wisdom, and blessings have been my guiding light, providing me with the strength and clarity to pursue and complete this endeavour. To my supervisors, Prof Celik and Prof Genc, your invaluable insights, patient guidance, and unwavering support have significantly shaped my academic growth. Thank you for sharing your knowledge and expertise generously, enabling me to achieve this milestone.

Finally, my deepest appreciation goes to my family, whose continuous encouragement, love, and understanding have been my greatest source of motivation and strength. Your belief in me has made all the difference, and I am eternally grateful for your unwavering support.

Table of Contents

List of Acronyms

AI	Artificial Intelligence
ARIMA	Autoregressive Integrated Moving Average
Bi-GRU	Bidirectional Gated Recurrent Unit
BPS	Business Process System
CNN	Convolutional Neural Network
CPS	Cyber-Physical System
DBN	Deep Belief Network
DT	Digital Twin
EAM	Enterprise Asset Management
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
IoT	Internet of Things
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
PSD	Particle Size Distribution
R^2	Coefficient of Determination
RNN	Recurrent Neural Network
ROM	Run of Mine
RMSE	Root Mean Square Error
SCADA	Supervisory Control and Data Acquisition
SVM	Support Machine Vector

List of Figures

List of Tables

1 Chapter 1: Introduction

This chapter defines Digital Twins (DTs) and their application in the mining industry. It also motivates that the growth of DT applications has the potential to solve mining-related challenges. It further presents the key area of this research, which is a crushing circuit within a process plant of a mine in South African. The flow of the chapters is presented to provide a concise overview of the thesis structure and the interconnectedness of the key sections of the study.

The South African mining industry significantly impacts the economy, the workforce, and surrounding communities [1]. An illustration of a mining value chain is shown in Figure 1.1 [1]. When prospectors identify a resource and determine its economically feasible to mine, a mining right authorisation must be secured from the mandated authority. Following which, a mine can be established and a processing plant can be built. Processing of ore then commences. This thesis focuses on the crushing circuit of the plant.

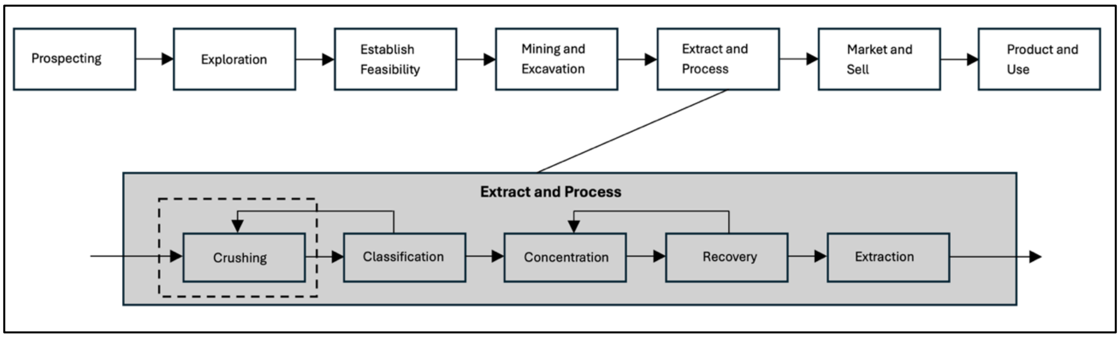

Figure 1.1: A general mine value chain, modified from Rajagopalan [1]. This study focused on the dashed line.

The mining value chain, as shown in Figure 1.1, consists of seven main stages: Prospecting, Exploration, Establish Feasibility, Mining and Extraction, Extract and Process, Market and Sale and Product and Use [1]. Each stage represents a critical component in the journey from identifying a mineral resource to delivering it to market [1].

1. Prospecting: This is searching for mineral deposits and establishing mineral characteristics of the deposit.
2. Exploration: This phase involves searching for mineral deposits using advanced techniques such as geological mapping and geophysical surveys.

3. Establish Feasibility: This study is conducted to assess whether the mining project is economically viable and to secure financial investment for its development.
4. Mining and Extraction: Minerals are extracted from the earth through various mining methods, including surface or underground mining.
5. Extract and Processing: This stage involves crushing, grinding, and processing the extracted materials to separate valuable minerals from waste.
6. Market and Sale: This stage links the outputs of the mine to the customer base.
7. Product and Use: At this point, the product is refined, and it is ensured that responsible producer principles are applied until end-of-life.

According to the Minerals Council South Africa [2], 477 000 people were employed in the mining sector in 2023, contributing substantially to employment in the country. Beyond job creation, the industry invested in education and training programmes, benefiting employees and local communities alike. The combined contributions to company taxes and royalties totalled R104 billion in 2023, directly supporting the nation's fiscal capacity and development initiatives [2].

Figure 1.2 illustrates the mining industry's economic contributions in 2023 across multiple dimensions, including its direct gross domestic product contribution of R425.6 billion, revenue of R786.2 billion, and mineral exports valued at R781.6 billion [2]. This value add underscores the critical role of mining in driving economic activity, generating revenue, and sustaining livelihoods in South Africa [2]. Moreover, the employee earnings of R186.5 billion highlight the sector's ability to uplift household incomes, while the pay-as-you-earn contributions of R31.3 billion reflect its substantial impact on government revenue through taxation [2].

From the context of this thesis, the data in Figure 1.2 is directly relevant to understanding the need for efficiency improvements and technological advancements in mining operations [2]. With mining operations accounting for a significant portion of South Africa's economic activity, it is imperative to modernise and optimise processes to sustain this contribution amidst challenges like ageing infrastructure, which increases inefficiencies and risks; depleting mineral resources, requiring more energy-intensive processing; and rising operational costs due to fluctuating commodity prices, energy expenses, and labour disputes. These challenges threaten the sustainability of the industry's economic and social contributions, as depicted in Figure 1.2.

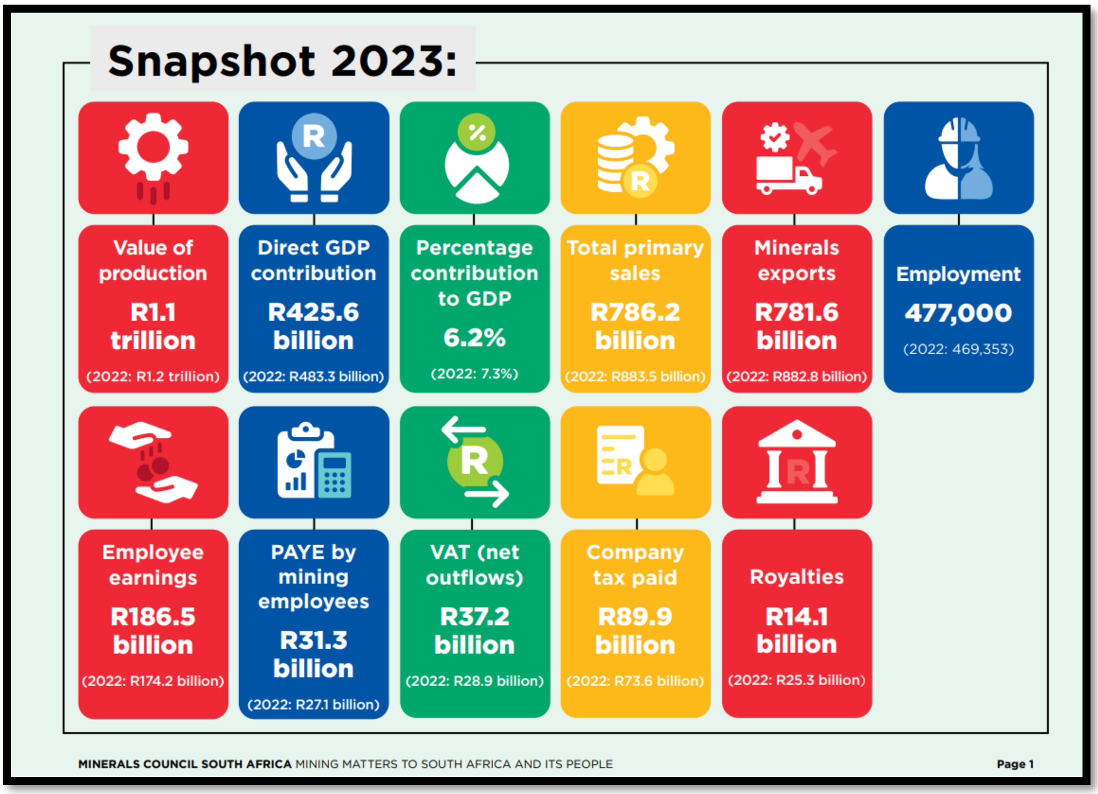

Figure 1.2: Mine contribution in South Africa for 2023 [2]

A focus on modernising the mining sector, as highlighted by the Mine Modernisation Programme, aligns closely with the adoption of digital technologies such as DTs. These technologies offer solutions to optimise critical components of mining operations, such as crusher circuits, which play a central role in ore processing [1]. By implementing predictive analytics and process optimisation, DTs could help address inefficiencies, improve safety, and ensure the sustainability of mining operations, directly contributing to the objectives of modernised, safe, and fulfilling workplaces.

1.1 Modern Mines

The mining sector has significantly increased its contribution to the economy by leveraging technology to build connected systems [2]. These advancements have enabled more efficient mining operations, early detection of failures, and improved safety performance [3]. As part of this transformation, digital mines are being developed using a combination of advanced technologies, such as sensors and software, to create integrated systems that enhance operational efficiency, optimise processes, and support data-driven decision-making [4].

The five characteristics of a digital mine are shown in Figure 1.3. A digital mine is integrated and technology driven [4]. It transforms traditional mining operations by leveraging advanced systems and data analytics [4]. It starts with real-time data capturing and feedback, where sensors continuously gather information on operational metrics and environmental conditions, enabling operators to adjust processes instantaneously [4]. Enhancing real-time monitoring of the environment and assets to continuously track equipment health and environmental safety, allowing issues to be identified and resolved proactively before they escalate. The use of digital twinning further enhances efficiency by creating virtual replicas of the mine, allowing for the simulation of different scenarios and testing of options without disrupting the actual mining process [4]. A digitally connected workforce enables effective collaboration and real-time data access for both on-site and remote team members. Autonomous mining further increases productivity and safety by using robotics and automated systems to handle high-risk tasks with little human involvement. Together, these five characteristics streamline operations and enable more sustainable and resilient mining practices [4].

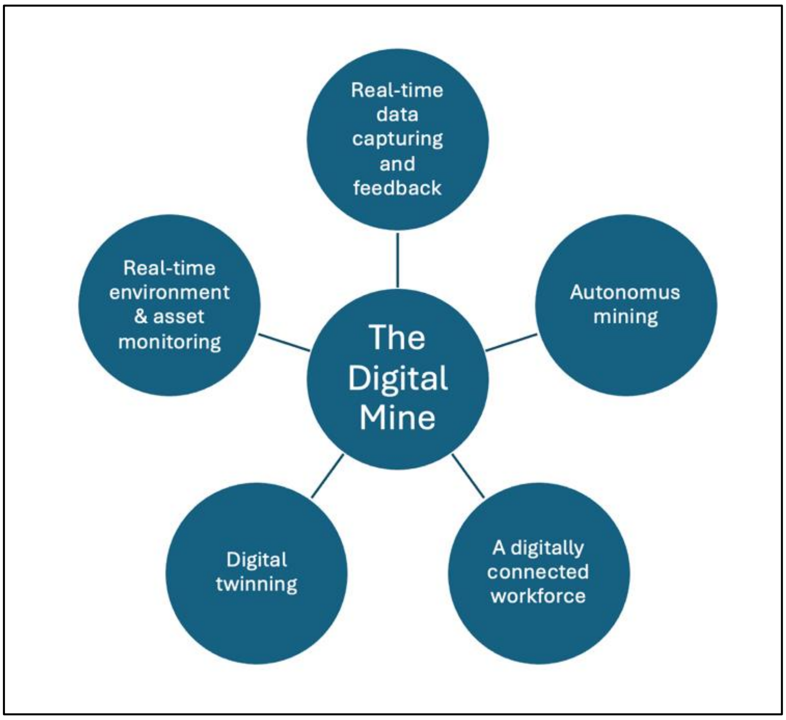

Figure 1.3: Characteristics of a digital mine [4]

Machine Learning (ML), a subcomponent of Artificial Intelligence (AI), plays a pivotal role in enabling digital mines [5]. ML is based on creating algorithms that can learn from the data and gradually enhance their performance with additional information [5]. These algorithms analyse past experiences, which are primarily represented as datasets, to create predictive models. Once trained, these models can be used for various tasks, including pattern recognition, anomaly detection, classification, regression, and forecasting. By utilising algorithms and statistical models, ML systems can analyse vast amounts of data and make autonomous decisions with minimal human intervention [6]. This capability improves operational efficiency and enhances safety by enabling higher-order controls and more proactive risk management approaches, a concept commonly referred to as intelligent mining [6].

In digital mines, sensors are strategically installed at key locations to facilitate continuous monitoring of environmental conditions and operational parameters [4]. These sensors allow for the rapid detection of changes, improving decision-making processes and minimising risks. Additionally, these sensors track equipment performance and location in real time, enhancing asset utilisation and operational efficiency. When integrated with DT technology, ML-powered systems further elevate predictive capabilities by analysing sensor data to forecast potential equipment failures [4]. These integrated ML-powered systems result in reduced unplanned downtime, an optimised maintenance schedule, and enhanced overall productivity [4].

In digital mines, sensors are placed at critical points to continuously monitor environmental and operational conditions, enabling quick identification of changes, supporting better decisions, and reducing risks. These sensors also provide real-time data on equipment performance and location, improving asset use and efficiency. With DT technology, ML-driven systems analyse sensor inputs to predict equipment failures, leading to less unplanned downtime, more efficient maintenance, and higher productivity.

By incorporating these advanced technologies, the mining sector is moving toward smarter, safer, and more efficient operations, positioning itself to address industry challenges while maximising its economic contributions.

1.2 Digital Twins

DT applications are a growing technology area within the industry [7]. Dr Michael Grieves first highlighted the DT concept in his 2002 lecture [7]. He described DT as a digital representation that mirrors the actual product with three main parts, viz., the real-world product, the virtual counterpart, and the link that connects them [7]. The connection component enables the tracing of the equipment's behaviour in real time to allow the operator to improve the utilisation and efficiency of the equipment. DTs will also indicate if scheduled maintenance is necessary based on the performance of real equipment and benchmarked data. Scheduled maintenance would result in savings for the business as it would decrease unnecessary service costs. For example, General Electric used digital twins to monitor and manage engine performance across its own fleet and that of a partner airline [8]. DT technology has been extended into mining applications, including maintenance optimisation for multi-unit systems. In such applications, a DT framework integrates discrete event simulation with system dynamics modelling to replicate mining operations and optimise maintenance policies [9]. This integrated approach enables a detailed evaluation of equipment performance and maintenance scheduling, which, in turn, helps reduce maintenance visits and saves millions of dollars in unnecessary maintenance stops through actionable DT insights [10].

The 2019 Gartner survey on DT entry into mainstream reported that 75% of organisations implementing IoT (Internet of Things) are already (13%) or planning to use (62%) DTs [11]. The Deloitte study on expected DT added that DT technology is becoming more widespread and is expected to grow with a 38% combined annual growth rate to reach 16 billion dollars by 2023 [10].

Through the implementation of DT technology, mining equipment can be continuously monitored by creating a virtual replica that mirrors its historical data and real-time performance [12]. By comparing the operational data from the physical equipment with its digital counterpart, any discrepancies or deviations in performance can be detected early. When the real equipment fails to operate at optimal levels, the DT can signal potential issues, prompting timely maintenance interventions [9]. This approach, known as condition-based maintenance, enables proactive fault detection, preventing equipment failures before they escalate. As highlighted in previous studies, the ability to compare real-time data with an optimised virtual model allows for immediate identification of performance anomalies [12]. If delays or inefficiencies arise in the physical system compared to the expected performance of its DT, maintenance teams can respond swiftly, reducing downtime and conserving operational resources. This predictive approach to maintenance enhances the efficiency, reliability, and lifespan of critical mining equipment [9].

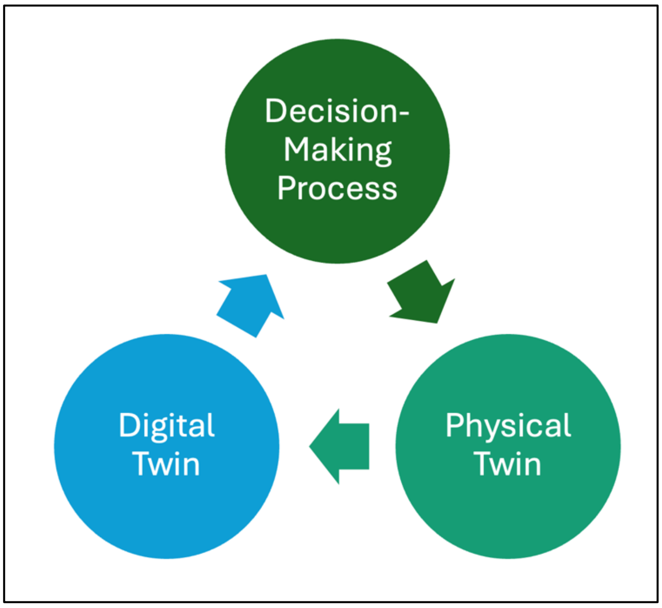

Figure 1.4: Digital Twin decision-making framework [13]

The DT framework in Figure 1.4 illustrates the continuous, cyclical relationship between the Physical Twin, its DT virtual counterpart, and the decision-making process that supports both [13]. The Physical Twin refers to the real-world equipment or operation, such as mechanical equipment, a piece of infrastructure, or the mine. The performance and condition can be measured through sensors or other data-gathering methods [12]. These measurements and observations feed directly into the DT, a virtual model that mirrors the status quo of the physical system with a high level of accuracy [7].

Once the DT is updated with real-time inputs from the Physical Twin, it can be used to analyse performance, simulate various scenarios, and predict outcomes under different conditions or operational changes [13]. Insights collected from this model are then fed into the decision-making process, which can involve automated algorithms, human experts, or a combination of both. Based on the insights provided by the DT, decisions are made. This would involve optimising performance, scheduling preventative maintenance, or implementing new operational strategies [13].

These decisions, once finalised, are then applied back to the Physical Twin, adjusting how it operates, updating parameters, or making necessary changes on the ground. Once these changes are implemented, the process continues with the Physical Twin's new data flowing back into the DT, which updates accordingly, and the cycle repeats. This closed-loop feedback system allows for continuous improvement, helping organisations respond rapidly to changing conditions, reduce downtime, optimise efficiency, and make better-informed, data-driven decisions [13].

The rapid growth in the adoption of DTs is attributed to delivering business value and is an integral part of the organisation's digital strategy [11]. The drawback to DTs is that building an impactful DT is a complex and iterative task due to the need to integrate many interconnected technologies and data sources [14]. One major challenge is defining clear requirements, including identifying the DT's purpose, data needs, and constraints, such as data accuracy and infrastructure limitations. [14]. The attractiveness, however, is that DT applications can be scaled up by starting small, i.e., single process applications within the industry.

The complete benefits of DT technology in the mining sector are summarised in Table 1.1. These contributions span multiple areas, including technological advancements, operational efficiency improvements, sustainability efforts, academic knowledge, and practical industry solutions. By leveraging DT technology, mining operations can achieve real-time, data-driven decision-making and significantly enhance safety, efficiency, and sustainability.

Table 1.1 highlights how DT supports the optimisation of processes, predictive maintenance, and energy efficiency while also addressing knowledge gaps in the context of South African mining operations, particularly for critical systems like crushing circuits. This comprehensive approach demonstrates how DT can be a transformative tool for achieving safer and more efficient mining operations.

Table 1.1: Key areas of contribution and applications of Digital Twin technology in mining

Area of Contribution	**Application**	**Reference**
Technological advancements	Optimising the process, predictive maintenance of key equipment and automation of plant in the mining sector	[7]
Operational efficiency improvements	Contributing to increased throughput, reducing costs and better energy efficiency of the mechanical equipment	[10]
Sustainability efforts	Reducing material waste and lowering the overall carbon footprint of the process plant	[8]
Practical industry solutions	Enhancing process plant safety and operational efficiency	[4]

1.3 Purpose of the Study

Although there are documented benefits of DT to industry, a study by Lottermoser and Barnewold in 2020, which analysed 158 active surface and underground mines, found no significant benefits of using AI and ML [15]. Furthermore, there was poor data quality, insufficient infrastructure, skill shortages, cultural resistance, strict regulations, high costs with uncertain return on investment, data security concerns, and challenges in integrating with legacy systems [15]. These issues, therefore, hindered the effective adoption and realisation of AI and ML's full potential in the mining industry.

Considering these difficulties, the application of DT technology emerges as a suitable solution to address the limitations currently hindering the mining industry's adoption of AI and ML. By focusing on a single application and applying a quantitative approach with historical data, the current study sought to identify the most effective ML model to support the development and deployment of a DT tailored for mine operations.

This study focused on the crushing circuit within the processing phase, as described in Section 1.1. This phase is critical as it transforms raw materials into a manageable size for subsequent processes. The crushing circuit was selected because it leads to increased complexity in maintenance due to the constant motion of the equipment, making it a prime candidate for optimisation [16].

Furthermore, the crushing circuit contributes heavily to the energy usage, maintenance, and overall efficiency of mining operations. DTs can contribute to improving energy efficiency by optimising equipment performance and reducing material waste [13]. This not only supports the sustainability goals of mining operations but also enhances productivity, making the crushing circuit a cornerstone for the implementation of DT technology.

1.4 Problem Statement

There are limited applications of DT in South African mines [17], particularly in the crushing circuit. Poorly maintained or inefficient crushing circuits can create bottlenecks in production, reduce plant efficiency, and increase operational costs. Additionally, the dynamic and mechanical complexity of the crushing circuit makes it highly susceptible to wear and tear, further emphasising the need for process optimisation of the equipment [18].

Most mines do not have a structured approach for the collection of connected and offline datasets [14]. However, it is important to have an accurate set of input information from instruments at the mine to implement production optimisation.

Hence, this study focused on developing a DT for an existing mine in South Africa by leveraging advanced ML approaches. The research emphasises the collection, preprocessing, and organisation of relevant data from the crushing circuit to facilitate the development of a sophisticated ML model within the DT framework.

1.5 Research Question

To develop a DT for the crushing circuit of a process plant, the following research question was addressed in this study:

What is the best ML approach to build a DT for the crushing circuit as part of the process plant at the mine?

1.6 Aims and Objectives

To answer the research question, this research was directed at understanding the ML approaches available for DT applications. The study explored DT solutions for the crushing circuit. The performance of the DT solution was quantified through a clear definition of the validation and testing criteria.

Therefore, the research objectives were to:

1. Collect the multivariate time sensor data for the crushing circuit and pre-process the data for ML;
2. Develop modelling approaches to capture the dynamic behaviour of the crushing circuit equipment using ML and;
3. Identify the best ML method for developing a DT for the crushing circuit by comparing the different experimental results.

1.7 Research Significance

Along with the limited applications of DT technology in the mining industry [18], this study aimed to support the growing body of knowledge by exploring the effectiveness and potential of DT applications in a mining environment. While DT technology has seen success in other industries, such as manufacturing and energy management, its adoption in mining remains limited, with only a 4% share of published papers [19]. The lack of research on DT is particularly evident in critical mining processes such as the crushing circuit [19]. This study sought to fill this gap by demonstrating how DTs have the potential to monitor, forecast, and optimise the performance of the crushing circuit. The insight gained from the study is expected to provide guidance for future DT innovation, which would accelerate digitalisation in the mining sector.

Furthermore, the study is particularly relevant to the South African mining industry, where the sector faces unique challenges such as ageing infrastructure, declining ore grades, and increasing operational costs. By leveraging DT technology, this study sought to deepen the understanding of equipment behaviour, enabling more informed decision-making and proactive maintenance strategies.

1.8 Limitations

The limitation of this study is the relatively narrow scope, focusing specifically on a single aspect of the mining process: the crushing circuit. While this targeted approach allows for a detailed exploration of the application of DT technology and ML methods, it does not encompass the broader aspects of the mining value chain or other critical equipment. As a result, the findings may not be directly generalisable to other mining operations or equipment types without further validation.

1.9 Thesis Outline

The thesis is structured into five chapters. Chapter 1 introduces the research topic, outlining the research problem, aims, objectives, hypothesis, limitations, and questions while also highlighting the study's significance. Chapter 2 provides a comprehensive literature review, discussing the background of DTs and various ML models and identifying gaps in current research. Chapter 3 details the methodology, including exploratory data analysis and the application of different ML models. In Chapter 4, the results are presented and discussed, with a comparison of the accuracy of each model's predictions to determine the most suitable approach for building a DT. Finally, Chapter 5 concludes the thesis by offering recommendations for the optimal ML model for DT development and suggesting directions for future research and further applications of DT technology.

1.10 Chapter Summary

This chapter introduced the concept of DT and its potential applications in the mining industry, particularly in addressing challenges related to operational efficiency, predictive maintenance, and energy optimisation. The contribution of the mining sector to the South African economy was highlighted, emphasising the need for technological advancements to sustain its contributions amidst pressing challenges such as ageing infrastructure, rising costs, and depleting mineral resources.

The focus of this study was on the crushing circuit within the processing phase (Figure 1.1). The justification for selecting the crushing circuit is that it plays a critical role in mining operations and presents unique challenges compared to other equipment. The research problem, aims, objectives, and research question were clearly articulated, outlining the scope and direction of this study.

2 Chapter 2: Literature Review

This chapter presents a literature review to examine key research areas, unpack the research gaps, and establish foundational knowledge for this study. Three main topics are highlighted, viz., introduction to DT, ML methods, and ML applications in the mining environment. This chapter also bridges the gap between DT and ML in the mining industry by exploring current use-case examples.

The previous chapter explored the definition and benefits of DT. This section focuses on the development of a protocol to build a DT, as well as verify and validate that the purpose is being achieved. The DT uses various information sources to collate real-time data for accurate modelling of equipment [20]. The equipment is modelled using analytics in a virtual setting [21]. DTs operate as proxies for the combination of inputs from people and traditional monitoring devices and controls. The use case to optimise the fleet maintenance of the operations in a mining environment showcases both the maintainability and profitability of the DT. The DT identifies optimal fleet configurations and maintenance schedules, which improve equipment utilisation and minimise maintenance costs. This highlights DT's effectiveness in a mining environment, though it application is currently limited to specific use cases. A further use case for DT applications is explored in this chapter.

2.1 Digital Twin Development

The development of a DT is guided by a use-case scenario. The proposed nine principles or high-level value statements for an effective DT, as per the University of Cambridge, are shown in Figure 2.1 [22]. There must be a clear purpose for the development of the DT. This is demonstrated through strategic needs, such as social, financial, operational, governance or environmental. As such, a DT must feature the organisation's digital transformation strategy with clear ownership defined [22].

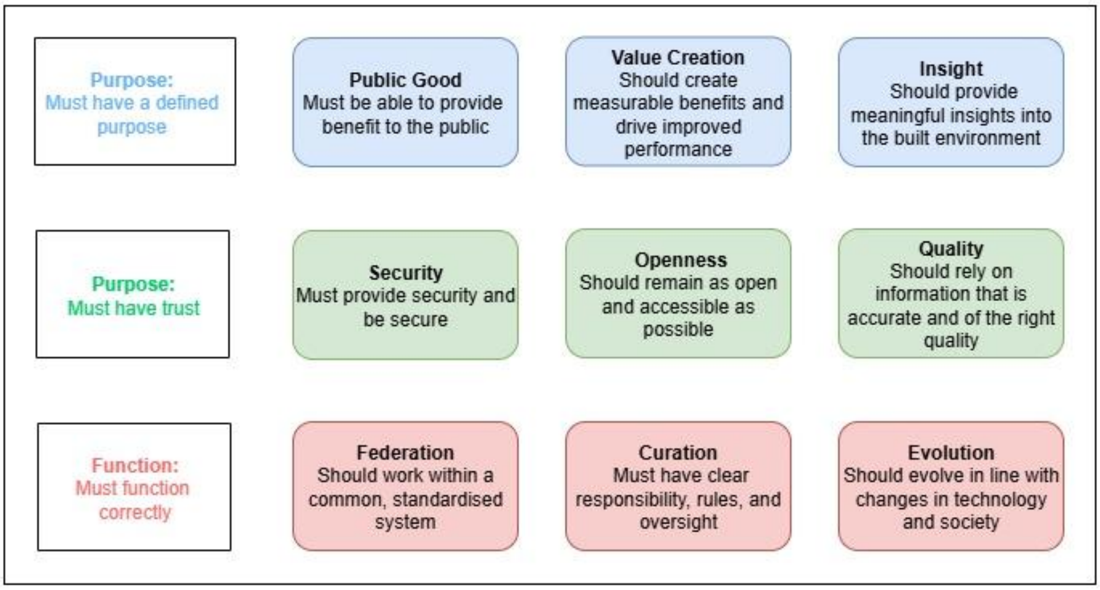

Figure 2.1: Adapted from the University of Cambridge Gemini nine-step principles [22]

In addition, there is an accepted methodology for collating and integrating real-time data for a dynamic, robust, and agile DT [20]. The limitation of the current methodology is that it is not extended into the mining industry, and it is still in the infancy stage of implementation. The methodology shown in Figure 2.2 indicates a streamlined and integrative process for developing and validating data requirements within an Enterprise Asset Management (EAM) system. The steps are as follows [23]:

1. Identify and Categorise Requirements: Extract and organise organisational needs.
2. Develop Asset Functions and Systems: Create a classification system for assets and their functions.
3. Identify Organisational Information Requirements: Align the information needs with goals of the organisation.
4. Develop Functional Information Requirements: Define detailed functional data requirements.
5. Develop Asset Information Requirements: Specify the detailed asset-related information needed.
6. Validate Information Requirements: Review and ensure the Develop Asset Information Requirements are fit for purpose. If not, refine and revalidate.
7. Document and Communicate: Finalise, document, and share the validated requirements for implementation.

This process ensures that the data requirements are aligned with organisational goals, validated for accuracy, and effectively communicated for use within the EAM system.

Furthermore, by following this framework, a DT can effectively replicate physical assets, enabling real-time monitoring, predictive maintenance, and optimised decision-making in a mining environment [23]. The structured classification of asset functions ensures that relevant operational parameters, such as equipment performance trends, energy consumption patterns, and failure indicators, are systematically integrated into the DT. The validation and documentation process ensures that the data feeding into the DT is reliable, eliminating inconsistencies that could affect predictive accuracy.

In addition, the methodology aligns with industry best practices in EAM by ensuring that maintenance and operational strategies are data-driven [23]. Being data-driven enhances the DT's capability to detect inefficiencies, predict failures before they occur, and support condition-based maintenance strategies. By leveraging these structured data flows, mining operations could achieve improved equipment reliability, reduced downtime, and enhanced overall process efficiency, contributing to increased productivity and cost savings [13].

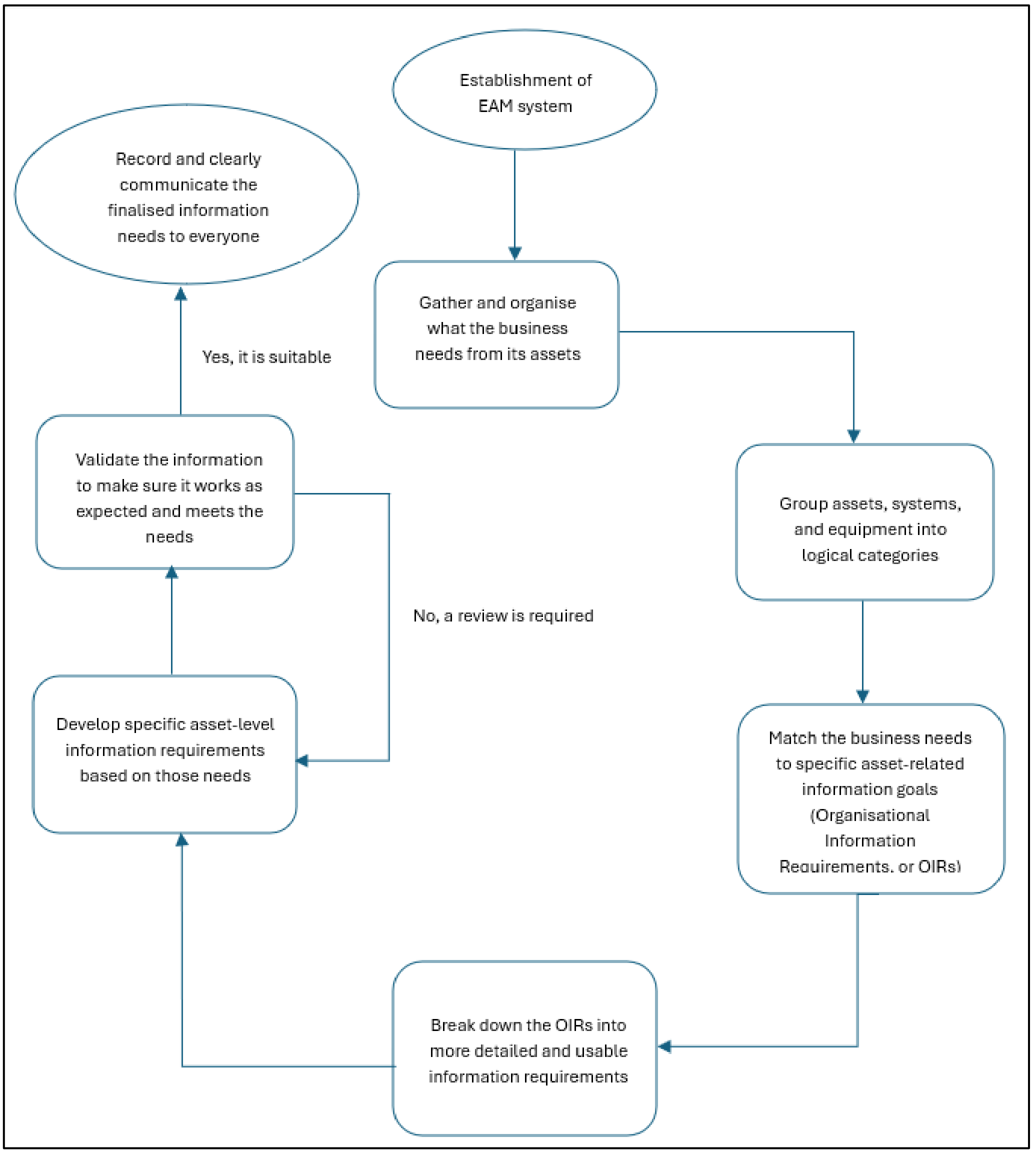

Figure 2.2: Adapted from J. Heaton, A. K. Parlikad and J. Schooling Asset Information Requirements (AIR) method using inputs from the Enterprise Asset Management (EAM) and Organisation Information Requirements (OIR) to improve the output of physical assets through a Digital Twin [23]

A DT is created using four steps, as shown in Figure 2.3. Through this four-step process, an accurate snapshot of the Physical Twin's current state is determined, providing a systematic framework for developing a DT for the crushing circuit. By identifying the real-life system, integrating sensors to collect real-time and historical data, and developing a virtual model using ML algorithms, the DT enables predictive analytics to inform human operators on site or remotely about potential issues and solutions [24]. This capability allows for the prediction of imminent breakdowns and supports preventative actions, thereby reducing downtime, optimising maintenance schedules, and improving equipment reliability [8]. Furthermore, the DT can simulate various operational scenarios, offering a virtual testing environment for analysing equipment behaviour under different conditions. These simulations can be used for training purposes, enabling the mine to prepare for potential shutdowns, optimise processes, and enhance decision-making [3].

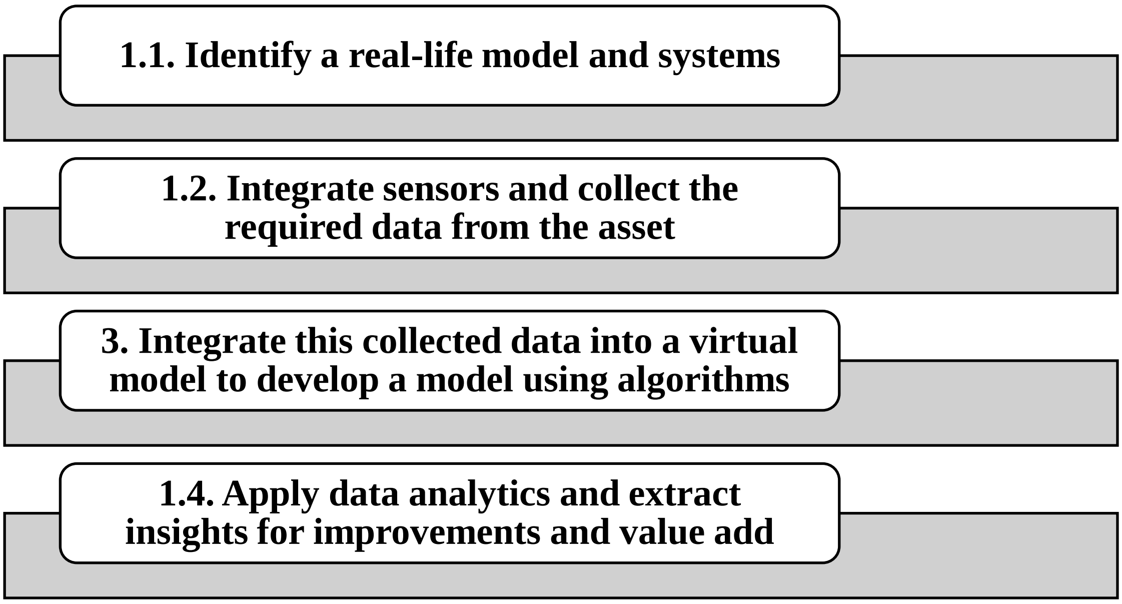

Figure 2.3: Four-step process for creating a Digital Twin [24]

The DT relies on ML algorithms to significantly enhance and expand its predictive, analytical, and decision-making capabilities [13]. ML models empower DTs by efficiently processing large volumes of real-time and historical data, identifying complex patterns, accurately predicting equipment failures, and effectively simulating diverse operational scenarios [13]. Furthermore, ML's proficiency in handling vast, high-dimensional datasets enables DTs to manage the complexity and in dynamic mining operations, leading to more informed, and data-driven decisions [5]. The overview of ML algorithms, as well as use-case integration with DT, follows.

2.2 Overview of Machine Learning

ML is a type of AI that helps systems to learn and make decisions depending on the data [5]. The ML can recognise patterns and trends, which can be used to make predictions, automate systems, and continuously improve the efficiency of systems [5].

DTs rely on advanced modelling techniques to develop an accurate digital model of actual systems, enabling real-time monitoring, prediction, and optimisation [5]. Statistical methods and ML techniques are widely used to analyse sensor data and model equipment behaviour [5]. These methods can be broadly categorised based on their tasks: prediction, classification, and anomaly detection.

2.2.1 Statistical Methods

Statistical approaches, which include the Exponential Smoothing and the Autoregressive Integrated Moving Average (ARIMA), are commonly used for time-series forecasting. ARIMA, for example, has been widely applied in DTs for tasks to identify trends and make short-term forecasts of equipment performance [8]. ARIMA is suitable for modelling time-series data with stationary trends, making it effective for simpler systems with limited data and linear behaviour [25].

2.2.2 Traditional Machine Learning Methods

Linear regression and decision trees, as regression models, are used to predict numerical outputs. These outputs include energy consumption, production outputs, or equipment wear [13]. These methods are computationally efficient and interpretable but limited in capturing complex patterns. Support Vector Machines are effective for classification tasks, such as detecting anomalies or failures in equipment based on sensor data [10]. Traditional ML models are useful for initial analysis where datasets are small or moderately complex. They provide insights into basic relationships within the system.

2.2.3 Advanced Machine Learning Methods

Artificial Neural Networks capture non-linear relationships in data and are suitable for predicting complex behaviours of physical assets. Long Short-Term Memory (LSTM) Networks, a type of Recurrent Neural Network (RNN), are commonly used for DTs to analyse multivariate time-series data due to their ability to retain long-term dependencies in sequential data [16]. Similar to LSTMs, Gated Recurrent Units (GRUs) are computationally efficient and are often applied to predict equipment behaviour and remaining useful life.

Transformer models excel at processing sequential data and have demonstrated improved performance in modelling time series, particularly in applications where long-term relationships and high-frequency data are critical [14].

Advanced ML methods such as LSTMs, GRUs, and Transformer models were selected for the current study due to their superior ability to model dynamic, non-linear, and time-dependent behaviours of systems like the crushing circuit [16].

2.2.4 Hybrid Methods

Hybrid approaches combine statistical models and ML techniques. For instance, ARIMA-LSTM models integrate statistical forecasting (ARIMA) with deep learning (LSTM) to enhance accuracy in complex systems [17]. Hybrid methods are useful for improving prediction accuracy when data exhibits both linear and non-linear characteristics.

The selection of RNNs (LSTM and GRU), ARIMA, and the Transformer model makes sense for this study as they align with the data type and objectives of the study [17]. The crushing circuit generates multivariate, time-dependent data, which requires models capable of handling sequential patterns and long-term dependencies. LSTM and GRU, as advanced RNNs, are well-suited for capturing non-linear relationships and predicting equipment behaviour, making them ideal for predictive maintenance. ARIMA, on the other hand, serves as a baseline statistical model for analysing cyclical and trend-based patterns, providing interpretability and efficiency for simpler data structures [25]. Lastly, the Transformer model, with its self-attention mechanism, excels in managing large datasets with complex interdependencies, offering scalability and accuracy for modern DT applications.

Together, these models allow for a comprehensive comparison between traditional statistical methods, modern ML approaches, and best practice techniques, ensuring the identification of the most effective solution for optimising the crushing circuit's performance.

2.3 Recurrent Neural Network

An RNN stores information over time using its internal memory [26]. Unlike standard feed-forward neural networks, RNNs rely on hidden states to retain data from prior steps, enabling it to capture long-range relationships between consecutive elements in a sequence [26]. This ability makes them well-suited for tasks like time-series prediction and speech recognition [27]. The model is designed for sequential data, which uses memory from previous predictions [26]. The ML is suited for time-series forecasting, which makes it an excellent choice for predicting the behaviour of the mechanical equipment [26]. GRU and LSTM, two types of RNN, are reviewed in subsequent sections.

In each time step, an RNN processes the current input along with the hidden state from the previous step to generate a new hidden state and an output [26]. This approach allows the network to incorporate information from earlier parts of the sequence, helping it identify and learn patterns over time across the sequence [26]. The RNN architecture is shown in Figure 2.4.

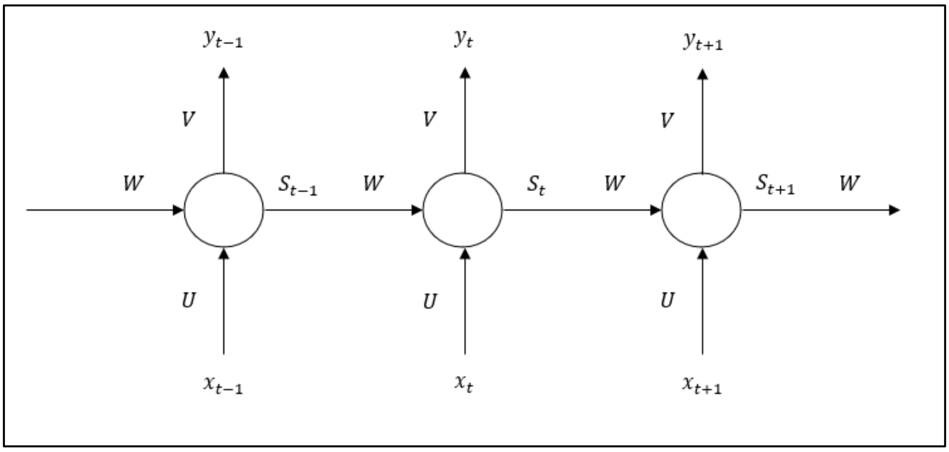

Figure 2.4: Recurrent neural network architecture [26]

2.3.1 Hidden State Update

$$\boldsymbol{S_t = f(Wx_t + US_{t-1} + b)} \tag{2.1}$$

In an RNN, the input at the time t is denoted as x_t [26]. At each of the time steps, the hidden state, S_t is processed alongside the previous state indicated as S_{t-1}. This computation involves applying a weight matrix, which is denoted as W for the inputs and U for the previous state. Additionally, the bias term, b is included to shift the resultant values before they are passed through the activation function f, which is typically in the form of hyperbolic tangent function, $tanh$. This function allows the neural network to collect and learn complex temporal dependencies in the data [26].

2.3.2 Output Calculation

$$\boldsymbol{y_t = g(VS_t + c)} \tag{2.2}$$

The output of the function, y_t at time t is derived from the hidden state S_t. The hidden state is transformed by the weight matrix V, which projects it into the output space [26]. The bias term, c is added to introduce non-linearity [26]. The function, g is typically in a softmax function for classification tasks and linear models for regression problems. This allows the neural network to generate outputs that can capture complex relationships and patterns presented in the hidden states.

However, RNNs face challenges when dealing with long-term dependencies due to issues like vanishing and exploding gradients. These problems can make training difficult for sequences with extended time spans [26]. To resolve these limitations, advanced architectures such as LSTM networks and GRUs were developed [26]. These models include mechanisms that improve their ability to handle long-term dependencies and enhance training stability [26].

2.4 Gated Recurrent Unit

The GRU architecture addresses the exploding/ vanishing gradient by incorporating gating mechanisms to help regulate the information flow through the network [28]. The conventional setup of the GRU architecture is presented in Figure 2.5.

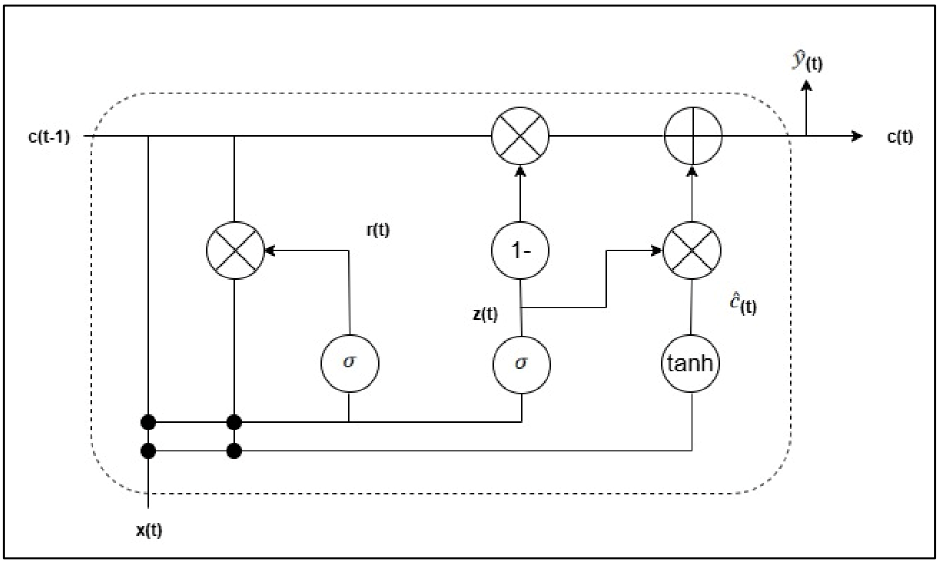

Figure 2.5: Vanilla Gated Recurrent Unit architecture block [28]

The GRU network is made up of two gates, viz., update and reset gate [28]. The gates are used by the GRU network to retain only the important information over time and remove the irrelevant information [28]. This allows the model to manage long-term dependency effectively without requiring any complex memory mechanisms [28].

The forward-pass equations for the GRU model are as follows:

$$\boldsymbol{z(t) = \sigma(W_z x(t) + U_z h(t-1) + b_z)}, \tag{2.3}$$

$$\boldsymbol{r(t) = \sigma(W_r x(t) + U_r h(t-1) + b_r)}, \tag{2.4}$$

$$\boldsymbol{\hat{h}(t) = tanh(W_h x(t) + U_h (r(t).h(t-1)) + b_h)}, \tag{2.5}$$

$$\boldsymbol{h(t)} = \ \boldsymbol{(1 - z(t)).h(t-1) + z(t).\hat{h}(t)} \tag{2.6}$$

The update gate, $z(t)$ determines the quantity of the previous hidden state that is retained. The reset gate, $r(t)$ determines the amount of previous data to be forgotten [28]. The candidate hidden state, $\hat{h}(t)$ includes the new input and resets the previous state. The new hidden state, $h(t)$ is the combined old state and new states [28]. Furthermore, W_z, U_z, b_z, W_r, U_r, b_r, W_h, U_h, and b_h are the weight matrices and bias for the respective update gate, reset gate, and candidate state.

2.5 Long Short-Term Memory

The exploding/ vanishing gradient challenges are overcome with the introduction of the LSTM architecture, as the network uses a constant error carousel to retain the information over long-term dependencies [2]. The typical vanilla architecture of the LSTM model is presented in Figure 2.6.

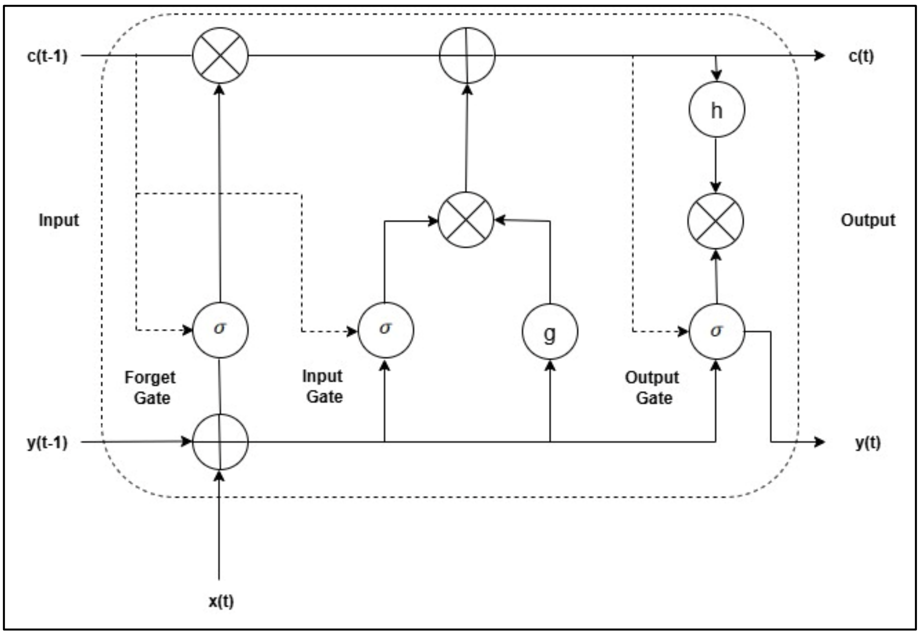

Figure 2.6: Vanilla Long Short-Term Memory architecture block [27]

The LSTM network is made up of three gates, viz., the input, output, and forget gates [27]. The gates are used to forget values that are not important or output the value [27]. The LSTM blocks or passes information based on its significance to the model [27].

The forward-pass equations for the LSTM are:

$$\boldsymbol{i(t)} = \ \boldsymbol{\sigma(W_i x_t + U_i h_{t-1} + b_i)}, \tag{2.7}$$

$$\boldsymbol{f(t) = \sigma\left(W_f x_t + U_f h_{t-1} + b_f\right)}, \tag{2.8}$$

$$\boldsymbol{\sigma(t) = \sigma(W_o x_t + U_o h_{t-1} + b_o)}, \tag{2.9}$$

$$\boldsymbol{\hat{c}(t) = \tanh \sigma(W_c x_t + U_c h_{t-1} + b_c)}, \tag{2.10}$$

$$\boldsymbol{c(t) = (f(t).c_{t-1}) + i(t).\hat{c}(t)}, \tag{2.11}$$

$$\boldsymbol{h(t) = \sigma(t).\tanh(c(t))} \tag{2.12}$$

The input gate, $i(t)$ determines the quantity of new information that is written into the cell state. The forget gate, $f(t)$ determines the quantity of information from the prior state that is retained. The output gate, $\sigma(t)$ determines the influence of the cell state on the hidden state [27]. The candidate cell state, $\hat{c}(t)$ incorporates the current input and the previous hidden states and is used to update the cell state, $c(t)$ by combining it with the previous cell state. The new hidden state, $h(t)$ is the combined of the output gate, $\sigma(t)$ with the transformed cell state [27]. Furthermore, W_i, U_i, b_i, W_f, U_f, b_f, W_o, U_o, b_o, W_c, U_c, and b_c are the weight matrices and bias for the respective input gate, forget gate, output gate, and candidate state [27].

2.6 Long Short-Term Memory VS Gated Recurrent Unit

The summary of the key features of the LSTM and GRU is shown in Table 2.. The link for how these models supports this research is also depicted in the table, with models showing benefits to the research.

Table 2.1: Key difference between Long Short-term Memory (LSTM) and Gated Recurrent Unit (GRU) and the link to this research [26]

Feature	LSTM	GRU	Link to this research
Controlled Memory Exposure	The amount of memory seen by other network units is controlled by the output gate.	The entire memory content is exposed to the network.	The LSTM's output gate assists by manging the selection of the previous states. This would be beneficial for complex crushing circuit patterns.
New Memory Computation	No standalone control for the flow of information from the prior time step.	Determines the amount of information from the previous activation used.	GRU simplifies training, as it directly controls the information flow. This provides benefits when dealing with repetitive time-series data from the crushing circuit.
Complexity vs Performance	Higher complexity due to the additional gate, potentially offering better performance for intricate tasks.	Fewer parameters, faster training, and requires less data to generalise.	The LSTM captures the subtle long-term dependencies, while the GRU is more efficient, as it reduces computation demands that is required from the crushing data.

2.7 Autoregressive Integrated Moving Average

The ARIMA is a statistical method used for time-series analysis [25]. The ARIMA offers further insight into understanding the cyclical patterns, which can be used to better understand the equipment during operational cycles or maintenance schedules [25]. The model is broken up into the following three components:

- AR for Autoregressive: This combines the dependencies between the parameter and the lagged value
- I for Integrated: This operation ensures the data is stationary by removing trends and stabilising the mean
- MA for Moving Average: This extracts the relationship between the observations and the previous step residual error

The ARIMA model is represented by an order of three parameters (p, d, q), where p represents the autoregression, which refers to past values in the time series, and q refers to the moving average, which illustrates the error for the order of q number of terms [29]. The d parameter is used to make the time series stationary. The general ARIMA model can be expressed as [29]:

$$\nabla^d x_t = c + \sum_{i=1}^{p} \varphi_i \nabla^d x_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{2.13}$$

Where ∇^d is the differencing operator of order d, φ_i and θ_i are the model parameters, and ϵ_t is the white noise error term (29).

The parameters must be optimally selected for successful forecasting [29]. The ARIMA presents the advantage of forecast time-series data that contain linear trends. The model is also flexible and easy to develop [29]. The limitation of the ARIMA is that it is less effective for complex non-linear time series [29].

It would be best to combine the ARIMA with the RNN models to have the benefits of capturing the linear and non-linear trends. This hybrid method would result in improved accuracy and adaptability of the crushing circuit application [29].

2.8 Transformer Model

The Transformer model is a deep-learning architecture that was introduced in 2017 [30]. The Transformer model requires less training time than RNN models as it processes input sequences in parallel [31]. These models contain an encoder and decoder to process the input and output data, respectively [30]. The model has a self-attention mechanism, which allows it to focus on certain parts of the data sequence [30]. This would be beneficial for large datasets with multiple interrelated parameters [31]. The encoder converts the input into a vector of a fixed length, and the decoder translates it into an output sequence. The architecture block of the Transformer model is presented in Figure 2.7.

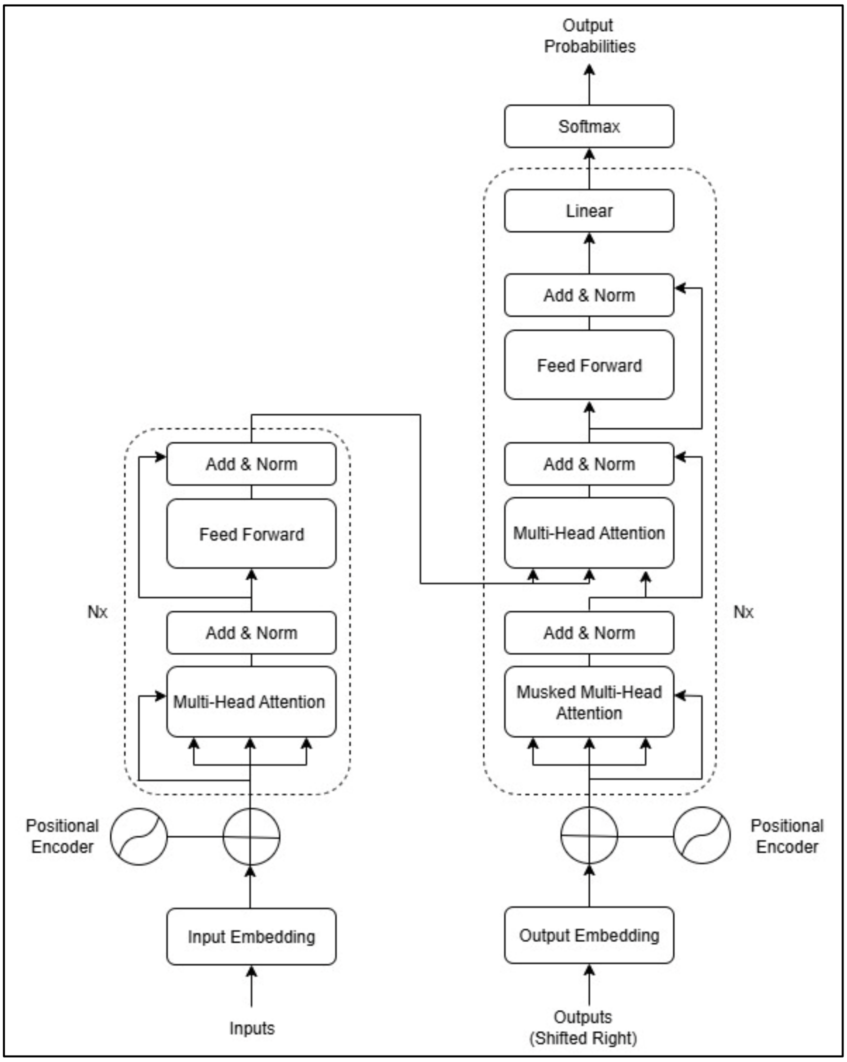

Figure 2.7: Transformer model architecture block [30]

The encoder on the left-hand side of Figure 2.7 is broken down into the following [30]:

- Multi-Head Attention: This mechanism allows the model to concentrate on different segments of the inputs in parallel and capture the relationship between tokens and their distance in the sequence.
- Feed-Forward Neural Network: This component adds non-linearity and allows for complex feature extraction.
- Add & Norm: These operations are key to stabilising training and ensuring efficient learning.

The decoder on the right-hand side of Figure 2.7 is broken down into the following [30]:

- Masked Multi-Head Attention: This allows the model to predict sequentially by masking output sequence future positions.

- Multi-Head Attention: This uses the input sequence to make accurate predictions.
- Feed-Forward Neural Network: This layer further processes the combined information from the attention mechanism
- Add & Norm: This layer is like the encoder as it stabilises training and ensure efficient learning.

The scaled dot-product attention equation is shown below [30]:

$$\boldsymbol{Attention(Q,K,V)} = \boldsymbol{softmax}\left(\left(\frac{\boldsymbol{QK^T}}{\sqrt{\boldsymbol{d_k}}}\right)\right)\boldsymbol{V} \tag{2.14}$$

The scaled dot-product attention equation plays a key distinct role in transforming and combining the information. The query matrix, Q consists of vectors that look for the required information, while the key matrix, K consists of vectors that provide an index to locate or match the information requested by the queries, and the value matrix, V contains the actual information to be aggregated once a match is established [24]. The dot product, QK^T measures the similarities between the keys and queries. The score is normalised by dividing by the dimensionality of the key vectors, $\sqrt{d_k}$, which ensures numerical stability [24]. The $softmax$ function then converts the scores into a probability distribution that determines how the values (V) are weighted in the output.

The Transformer model processes entire sequences in parallel, which would be beneficial for large-scale projects [30]. However, the model requires high computational costs and memory usage. These models are also more complex to train compared to RNN models [32]. In the context of the current study, the model presents the opportunity to build a high-fidelity DT for the crushing circuit.

2.9 Integration of Digital Twins and Machine Learning

The integration of DTs and ML techniques has proven to be a powerful solution for improving predictive maintenance, process optimisation, and operational efficiency in complex industry applications [33]. Several studies demonstrate the effectiveness of integrating DTs with statistical and ML approaches, highlighting their relevance to this research [33].

2.9.1 Hybrid Models: Autoregressive Integrated Moving Average and Recurrent Neural Network Integration

A hybrid prediction model that combined the ARIMA and RNN approaches was proposed [34]. The process flow, as illustrated in Figure 2.8, demonstrates how mechanical equipment data is processed through both models, and the results are averaged to improve the prediction accuracy for switch machines. The hybrid ARIMA-RNN model shown in Figure 2.8 is particularly useful in systems like the crushing circuit, where time-series data contains both linear trends (captured by ARIMA) and complex, non-linear dependencies (modelled by RNNs). Combining these models enhances accuracy in predicting equipment behaviour, improving predictive maintenance, and minimising downtime.

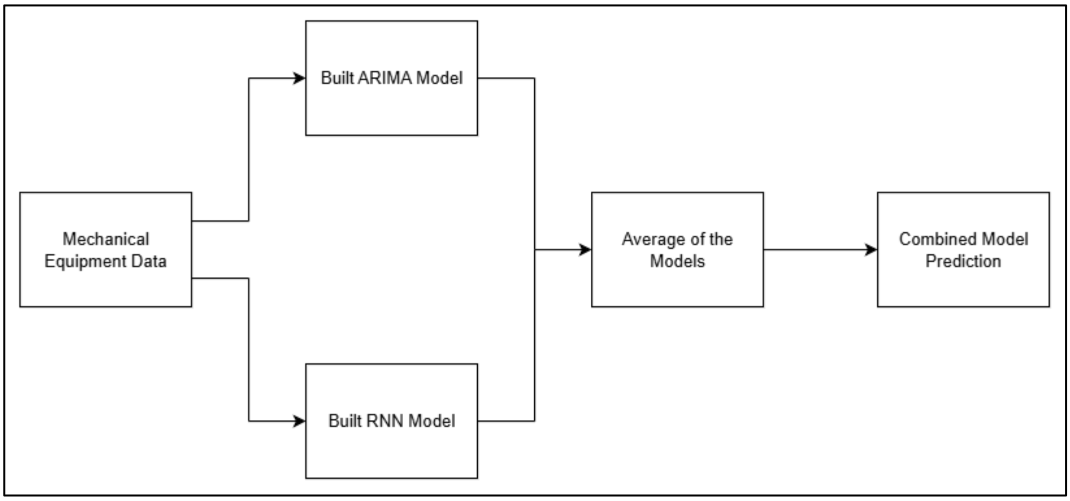

Figure 2.8: Combined Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Network (RNN) model process flow [35]

2.9.2 Bidirectional Gated Recurrent Unit

Bidirectional GRU (Bi-GRU) was used to analyse mechanical equipment behaviour for Shearers [35]. Bi-GRUs are extensions of RNNs that process data in both directions, capturing historical and future information to enhance predictive accuracy [35]. For the crushing circuit, where dynamic operational data is constantly updated, Bi-GRUs can process past and future states simultaneously to provide more accurate predictions. This dual analysis enhances the reliability of DTs in identifying wear patterns and predicting failures.

2.9.3 Long Short-Term Memory for Capturing Long-Term Dependencies

The LSTM was used for a study on rolling bearing vibration signals to predict their remaining useful life [36]. The method effectively captured temporal dependencies and the key degradation patterns from the sensor data, which significantly improved the model's robustness and forecasting accuracy [36].

LSTMs are particularly effective for the crushing circuit, where data has long-term trends, such as equipment degradation over time. The ability of LSTMs to retain historical patterns makes them ideal for predicting failures and modelling non-linear behaviours in mechanical systems.

2.9.4 Transformer Model for Complex Dependencies

A recent study demonstrated the potential of a Transformer-based model that enhanced the system-to-system transferability of crystallisation prediction and control tasks [35]. Transformers use a self-attention mechanism to assign different levels of significance within the input sequence, enabling them to handle complex and long-term dependencies [30].

In mining systems like the crushing circuit, where datasets are large and multivariate, Transformers are highly efficient for processing interdependencies across parameters, such as vibration, temperature, and throughput. Their scalability and ability to model relationships across the entire sequence make them well-suited for DT applications.

2.10 Relevance to This Research

The integration of DTs and ML techniques, as demonstrated in the examples above, highlights the potential for improving operational performance. This study builds on these approaches by:

1. Applying ARIMA to capture linear trends and seasonal behaviours in equipment performance.
2. Utilising LSTM and GRU models to handle non-linear, sequential data and predict dynamic behaviour.
3. Exploring the Transformer model for its ability to process large, complex datasets with multiple interdependent variables.

By testing and comparing these methods, this research aimed to identify the most robust ML method for developing a DT for the crushing circuit. The outcomes will provide actionable insight into the best ML algorithms for the actual crushing data.

2.11 Gaps and Opportunities of Digital Twins that Informed the Current Research

There is a plethora of research on DTs and their application in industry. Table 2.2 provides a summary of recent research on DTs, the type of ML applied, and their shortcomings.

Table 2.2: Summary of Digital Twin applications and applied machine learning techniques in industry

Area of Mine Value Chain	Type of Machine Learning Applied/ Study Focus	Outcome	Shortcomings	Contributions	Areas of Improvement	Reference
Geochemical Mapping	Deep Learning - Autoencoder, Convolutional Neural network (CNN), Generative Adversarial Network (GAN), Deep Belief Network (DBN)	Enhanced identification of geochemical patterns and anomalies, improving mineral exploration	Limited focus on real-time applications and interpretability of deep-learning models	Demonstrated the application of deep learning to complex geoscientific data, aiding mineral exploration	Integrating real-time data processing and improving model interpretability for more actionable insights	[37]
Switch Machine Maintenance	LSTM, ARIMA	Improved predictive maintenance accuracy for railway switch machines, enhancing visualisation	Limited applicability beyond railway systems; challenges in adapting model to different equipment	Developed a predictive maintenance model using Digital Twin (DT), showing increased accuracy for railway switch machines	Expanding the model to other types of equipment; enhancing model adaptability across different systems	[34]
Process Industry (General)	Systematic literature review on enablers and barriers	Identified key enablers and barriers for DT implementation, providing a framework for adoption	Broad overview; lacks specific implementation case studies and quantitative data	Organised fragmented literature on DT enablers and barriers, offering a foundational framework for industry adoption	Conducting case studies and quantitative assessments to validate framework applicability in real settings	[38]
Surface Mining	Six-layer DT architecture for Mining 4.0	Comprehensive DT framework enabling real-time data	Limited application of AI-driven optimisations;	Developed a six-layer DT framework for surface mining,	Implementing AI and machine learning for real-time decision-	[39]

		exchange and cognitive decision-making	focuses on conceptual architecture rather than empirical results	setting the stage for Mining 4.0 advancements	making; performing field trials to validate architecture	
Shearer Key Parts Maintenance	RNN, Bidirectional Gate Recurrent Unit (bi-GRU)	Achieved accurate predictive maintenance for shearer key parts with real-time state visualisation	Limited focus on broader mining equipment; heavily reliant on sensor data quality	Combined high-fidelity simulation with deep learning for effective remaining useful life prediction, enhancing maintenance accuracy	Expanding to other types of mining equipment; refining predictive accuracy through advanced data integration	[35]
Petrochemical Production Optimisation	IoT and Machine Learning Framework for Real-Time Control	Enabled agile, real-time production control and adaptability to changing conditions in the petrochemical sector	Complexity in data integration and real-time adjustments; high dependency on data quality	Introduced a framework for production optimisation integrating IoT and machine learning for dynamic production control	Enhancing predictive capabilities and refining data integration processes for more robust real-time control	[13]
Tailings Dam Stability Monitoring	ARIMA, Random Forest Classifier	Improved early warning system for tailings dam failure through real-time monitoring and prediction	High dependency on sensor data quality; potential for false positives due to environmental factors	Proposed a DT and machine learning framework for real-time monitoring of tailings dams, enhancing safety	Incorporating additional environmental and geotechnical parameters; refining predictive accuracy with deep-learning models	[40]
Building Performance Simulation	Systematic Review on AI, ML, DT, Cyber-Physical System (CPS), IoT, and Data Mining	Highlights conceptual overlaps, strengths, and limitations across simulation technologies	Limited empirical data on integration; overlapping terms with existing Business Process System (BPS) concepts create confusion	Provided an extensive literature review to outline how BPS can benefit from AI, DT, and other digital domains	Need for empirical work to assess BPS integration with AI/DT; clearer definitions and boundaries for simulation concepts	[41]
Conveyor Belt Systems	CNN (Support Machine Vector SVM), IoT, and Computer Vision	Improved fault detection, predictive maintenance, and optimisation for conveyor belts in mining and manufacturing	Dependent on high-quality data; challenges in integrating diverse AI models for unified monitoring	Consolidated AI techniques to enhance conveyor system reliability, safety, and efficiency	Improving data quality and developing integrated AI frameworks for seamless predictive maintenance	[42]

Production Systems Readiness	Naive Bayes, Decision Trees, Random Forests, and Logistic Regression	Enables efficiency assessment for production systems, improving planning and resource allocation	Limited to non-computerised systems; dependent on archival data rather than real-time inputs	Provides a framework for evaluating production efficiency, aiding decision-making in manufacturing contexts	Adapting the model for real-time data integration and broader applications across computerised systems	[43]

Based on the outcomes presented in Table 2.2, DT applications across industries face notable limitations, particularly in complex environments. Key challenges include limited real-time data integration, making rapid decision-making difficult, and issues with interoperability in large plants requiring seamless data exchange. DT models also struggle with generalisability, as they are not easily transferable across equipment and operational settings. Additionally, there is heavy sensor dependency, where data quality and accuracy can vary, and there is a lack of comparative studies on advanced ML models for optimal DT performance. The absence of empirical validation further reduces confidence in DT effectiveness, while the complexity and cost of implementing cognitive DT solutions hinder widespread adoption.

Despite these challenges, DT technology presents significant opportunities in the mining sector:

1. Enhanced Predictive Maintenance: Real-time monitoring can identify faults and prevent failures, minimising production downtime.
2. Process Optimisation: DTs provide data-driven insights to improve production efficiency and reduce energy consumption.
3. Safety Improvements: DTs enhance safety by simulating hazardous environments and monitoring equipment integrity.

The success of DTs requires a culture change where people will understand the maintenance requirements of the physical equipment and systems and actively collaborate with data scientists and Information Technology professionals to add value to the business.

This thesis addresses these gaps and opportunities by applying DT technology to the crushing section of a mining plant. By testing and comparing different ML techniques, the study identifies the most effective approach for predicting equipment behaviour and optimising operations. The research contributes empirical insights, enhancing understanding of model effectiveness and demonstrating the practical value of DTs in complex process plant environments.

2.12 Chapter Summary

This chapter explored the application of ARIMA, RNNs, and Transformer models for developing a DT, with each method offering distinct strengths and limitations. ARIMA provides a robust and interpretable foundation for linear time-series forecasting, effectively capturing trends and cyclical patterns. However, its reliance on linear assumptions limits its ability to handle the non-linear dynamics inherent in complex systems like the crushing circuit.

RNNs, including advanced variants such as LSTMs and GRUs, excel at collating temporal dependencies and patterns in sequential data, making them particularly effective for modelling non-linear behaviours and predicting equipment performance over time. However, these models face challenges with long sequences due to issues such as vanishing gradients and the inefficiencies of sequential data processing.

Transformers overcome many of the limitations of RNNs through their self-attention mechanism, which supports parallel processing and effectively manages the long-term dependencies in large, multivariate datasets [30]. Their ability to model complex relationships and prioritise critical data sequences makes them particularly suited for systems requiring real-time analysis and scalability. Despite these advantages, the higher computational cost of Transformers remains a key consideration.

Each of these models contributes unique value to the development of a DT for the crushing circuit. However, this study evaluated the models based on their accuracy in predicting equipment behaviour and development time to determine the most suitable ML approach. This comparative analysis provides a practical solution for the operational efficiency of the crushing circuit, thereby addressing critical challenges in the mining process plant.

3 Chapter 3: Methodology

This chapter explores the requirements for developing a DT tailored to the performance of the mine's crushing circuit to evaluate different ML methods. This study is important as it offers insights into using different ML algorithms for a conventional zinc mine in South Africa and responds directly to the research question. The analysis focuses on identifying the critical parameters that influence the production of the crushing circuit. These parameters are collected from real-time sensor data and historical datasets and form the foundation for developing the DT.

To achieve the study's objectives, the different ML models that were defined in Chapter 2 were employed to model the dynamic behaviour of the crushing circuit. The methodology investigated and compared the effectiveness of the ARIMA, RNNs (including LSTM and GRU), and the Transformer model in capturing the relationships between operational parameters and target production output. The selection of these models is driven by their proven capabilities in time-series forecasting, anomaly detection, and handling multivariate, sequential data [17].

Key stakeholders were engaged to understand the current plant performance parameters and maintenance requirements. These stakeholders included the plant manager, the chief digital officer, and the general manager. These engagements were conducted to ensure the correct parameters were selected for the crushing circuit DT and that the senior executives supported the research.

In addition to the guidance of the key stakeholders, ML was utilised for complex non-linear mapping of the predefined benchmarked model to the actual live data [35].

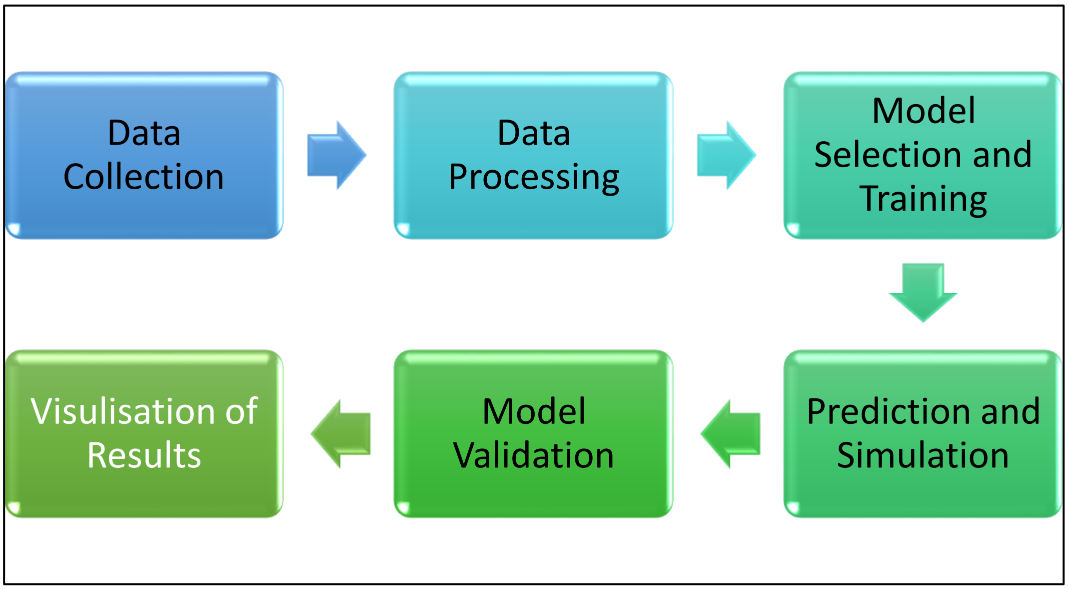

Figure 3.1: Data-driven workflow for developing and validating a Digital Twin

The workflow in Figure 3.1 outlines the process of developing and validating a DT model from the data collection stage to the final visualisation. The process begins with Data Collection, where raw instrument and sensor data is gathered from the physical asset. This data then undergoes Data Processing, where it is cleaned and prepared for analysis. Following this, Model Selection and Training take place, where the selected ML model is chosen and trained on historical data to learn patterns and trends of the asset. Once trained, the model is used for forecasting the target parameter. The outputs of the model are then subjected to Model Validation, where predictions are compared against actual data to assess accuracy and reliability. Finally, the results are presented in the Visualisation of Results stage, providing a clear and interpretable view of how well the DT model represents the physical asset's behaviour.

3.1 Mine Site

This section provides the context in which the DT was conceptualised. The selected mine is located in the Northern Cape, South Africa, and primarily extracts and sells zinc as a mineral concentrate. The process flow diagram, shown in Figure 3.2, illustrates the mechanical flow of material through the crushing circuit. The mine has been in operation for over 10 years, with site photos of the crushing equipment presented in Figures 3.3 to 3.7. The crushing plant is a critical component of the mine's operation, designed to process raw material efficiently while ensuring the durability and reliability of the equipment.

The process begins with a constant feed of material from the mining operations, delivered to the Run of Mine (ROM) bin trucks. The ROM bin shown in Figure 3.7 is critical in preventing large rocks from falling directly into the crusher circuit, as oversized ore could damage the equipment. From the ROM bin, material is fed into the apron feeder shown in Figure 3.4, which regulates the flow of material into the crusher using a Variable Speed Drive for precise control. This ensures a steady and optimised feed rate, which is critical for maintaining the crusher's efficiency and preventing overloading.

Once the material enters the crusher (shown in Figure 3.3), it undergoes size reduction at a target ratio of 4:1 post-crushing, as specified by the mine design planning. Since the incoming ore of the mine is expected to measure around 800 mm, the goal is to produce a 200 mm crushed product. In parallel, smaller particles bypass the crusher entirely and drop directly into the conveyor (shown in Figure 3.5) via a grizzly screen, which separates finer material from the bulk feed. The crushed material and bypassed smaller particles are then transported by the conveyor to the stockpile, where they await further processing or transport. This system ensures a smooth, controlled, and efficient workflow, minimising downtime and maximising throughput with the product-crushed material moved to the milling circuit, which is where the material is crushed to a finer particle and continues for further processing.

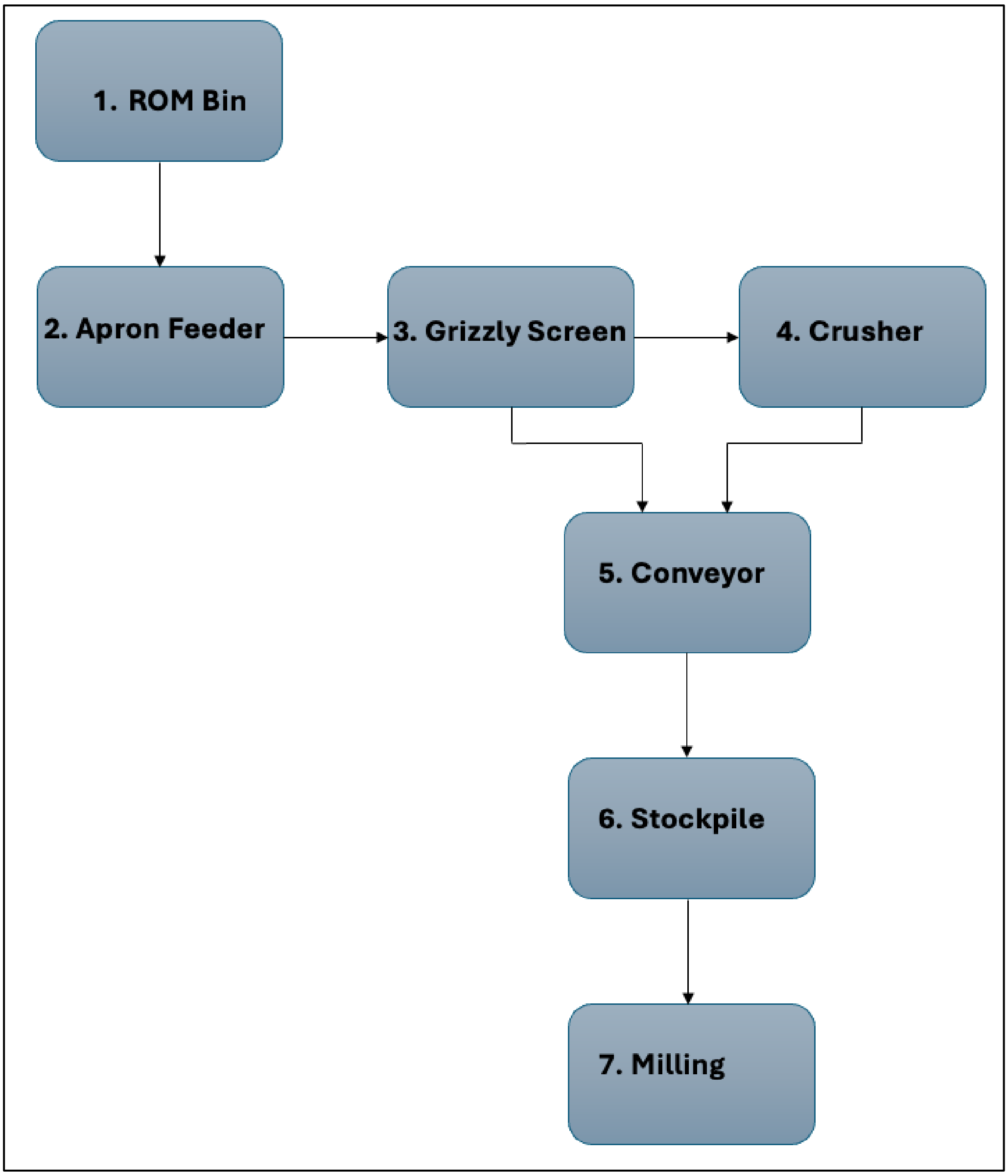

Figure 3.2: Simplified flow diagram of the zinc mine's crushing circuit

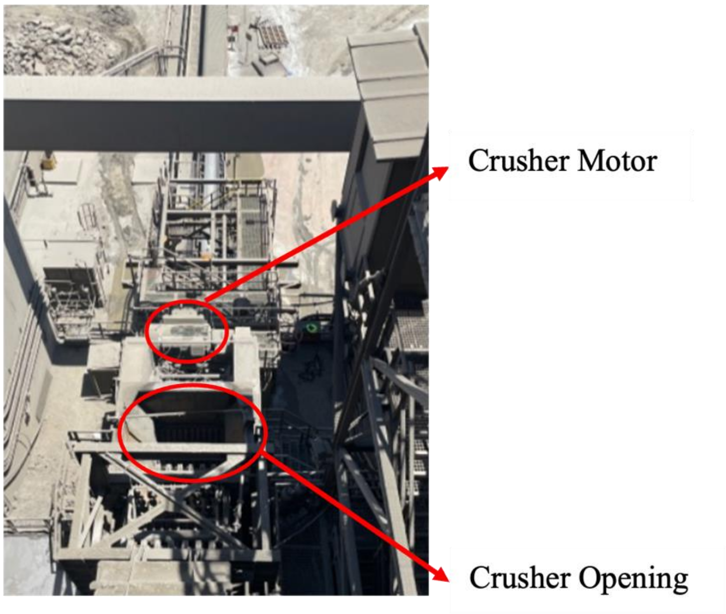

Figure 3.3: On-site crusher and motor with key sensors location labelled

Figure 3.4: Site apron feeder and motor with key sensor locations

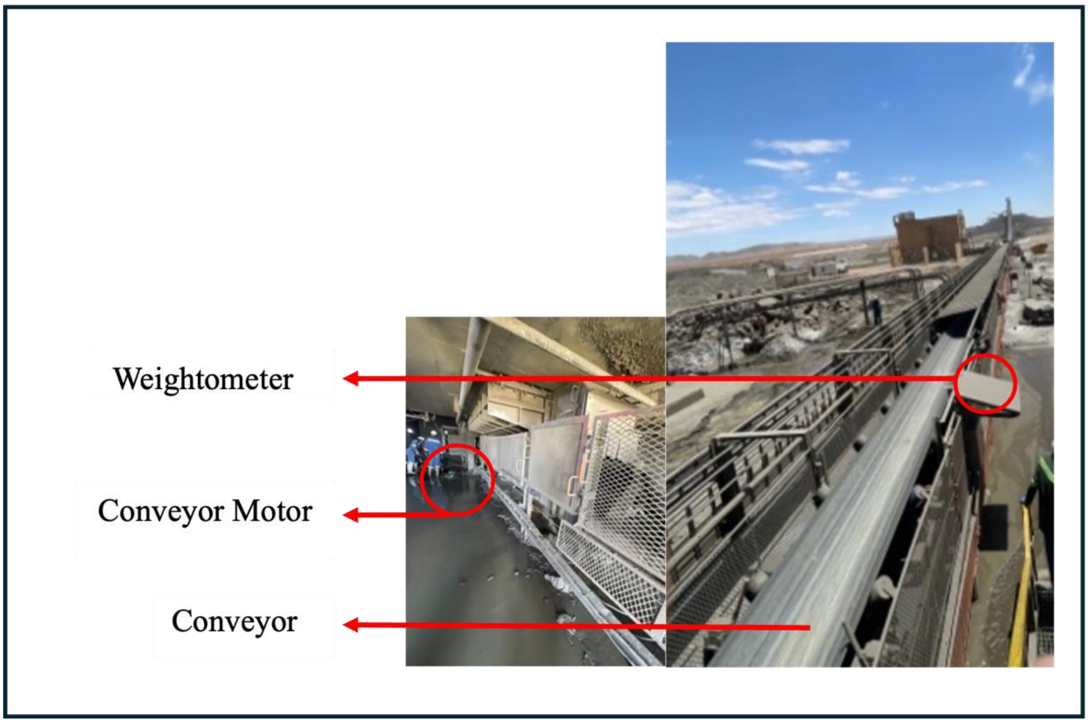

Figure 3.5: Site conveyor and motor with key sensor data

Figure 3.6: Particle size distribution (PSD) sensor for production material monitoring

Figure 3.7: Site Run of Mine (ROM) tip for initial ore unloading

3.2 Research Design

The research design is presented in Figure 3.8. The research design consists of three sections: Data Collection and Data Processing, Model Training, and Forecasting, which consists of implementing the different ML algorithms, including the hyperparameter optimisation, and, lastly, the Performance Evaluation of the different ML methods.

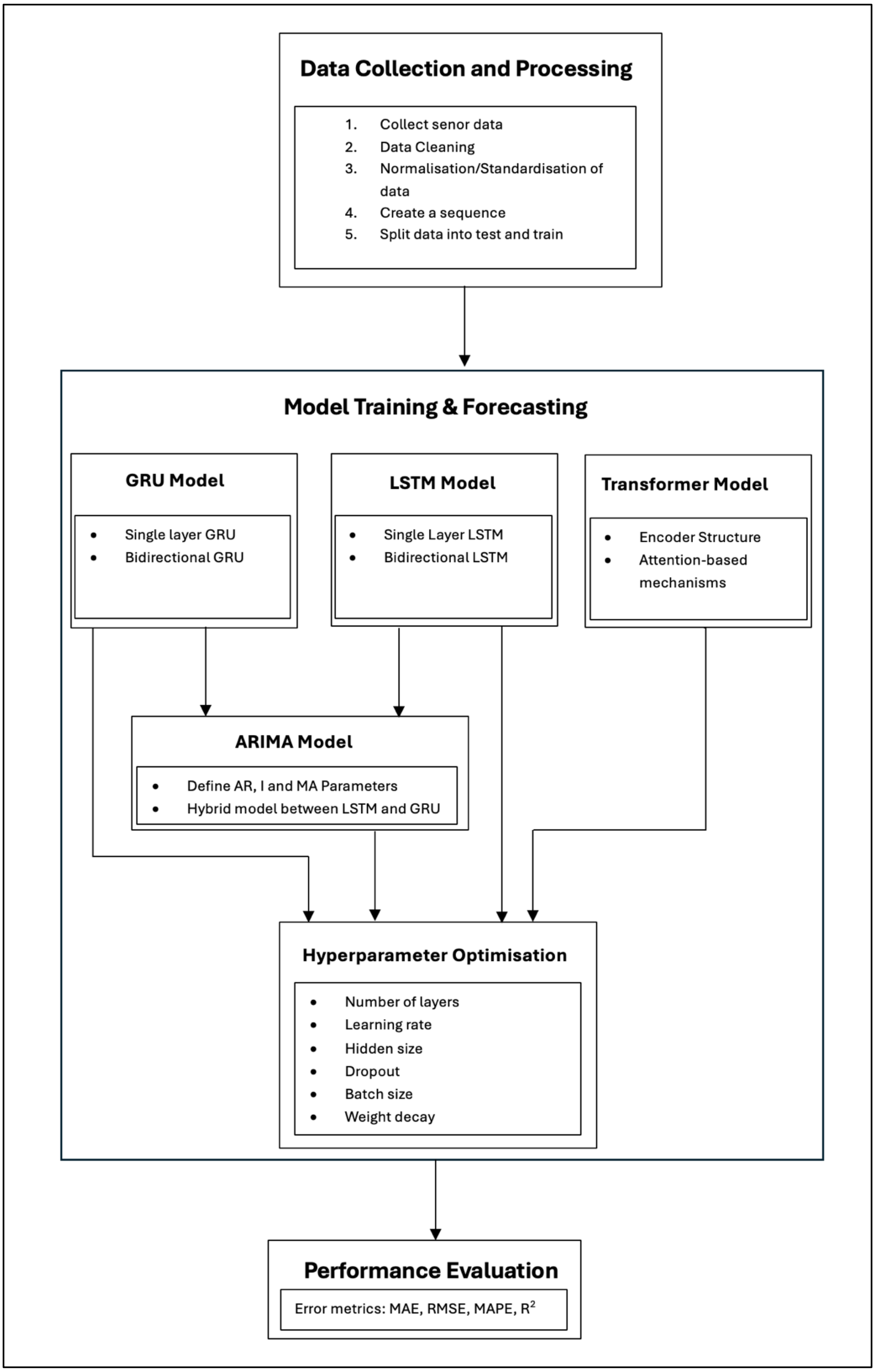

Figure 3.8: Research design, including the model training and forecasting of the Gated Recurrent Unit (GRU), Long Short-term Memory (LSTM), Transformer and Autoregressive Integrated Moving Average (ARIMA) models

3.2.1 Data Collection and Processing

The data collection was achieved by utilising existing instruments to define the behaviour of the mining equipment on site. The physical instruments and sensor outputs from the Supervisory Control and Data Acquisition (SCADA) included:

- Power meters – Mechanical equipment energy data
- Weightometer – Weight of the material within the plant
- Thermometers – Temperature of motor bearings
- Speed switches – Used to measure the speed of the conveyor belt
- Data loggers – Operational hours of the equipment

These sensor outputs are continuously collected and transmitted to the SCADA system, where the information is collated and displayed at a central point. The historical data for these instruments and sensors was available from the Historian. The Historian is a specialised software system used to continuously collect [44], store, and manage the time-series data for the mine. The data was in a raw format, which constituted the unprocessed data. The unprocessed data was processed through exploratory data analysis, by identifying missing values, outliers and anomalies. This process data was utilised for developing the DT.

3.2.1.1 Collecting Sensor Data

The data was collected for a 12-month period from 01/06/2021 to 31/05/2022 for all the crushing circuit parameters, as this was when the plant instruments were fully functional. The data collection period was selected for one year as it captures a complete annual cycle of operational and maintenance variations, which provides a clear indication of the crushing circuit's performance under standard conditions. The plant Historian was able to collect 39 unique datasets for the crushing plant, and the data was in an hourly format. The data was in a time series and was related to the main mechanical equipment in the area, viz., the crusher, apron feeder, and conveyor. The total parameter count of 39 datasets is unpacked in Table 3.1.

Table 3.1: Crushing plant parameters and total counts

Parameters	Parameter counts	Description
Energy and Power	10	The datasets contain the electrical current draw for the crushing mechanical equipment
Crusher Data	11	The specific data relating to crusher, including the auxiliaries and temperature sensors
Conveyor Data	4	Data for the conveyor, including the speed sensors and temperature indication
Apron Feeder Data	6	Data for the apron feeder, including speed and temperature indication
Feed and Production	2	Quality metrics such as particle size distribution and material quality
Run of Mine (ROM) Bin Level Indication	1	Indication of material within the ROM bin
Fraction Data	5	Particle size fraction data within the mill
Total	**39**	

The energy parameters contained mechanical equipment auxiliaries such as heaters and hydraulic units. These datasets were irrelevant to the study, as that equipment was used to supplement the main mechanical equipment. The fraction data (below 150 μm) was excluded from the study, as this was for another process outside of crushing. The reference component 3 is a stationary component with no sensor data and no energy consumption (grizzly screen from Figure 3.2). The remaining parameters were reviewed with the key stakeholders. The critical parameters investigated are shown in Table 3.2. The full list of parameters is presented in Appendix A.

Table 3.2: Critical input and output parameters required for crushing circuit Digital Twins

Critical Parameter	Unit	Reference Component from the Process Flow Diagram (Figure 3.2)	Rational for Selection
Apron Feeder Current	% rated current (A)	2	The apron feeder controls the material flow rate into the grizzly screen and crusher. Monitoring its current ensures that the feeder operates within its rated capacity, preventing overloading and ensuring

			steady material flow into the circuit. Deviations in current can indicate excessive load or mechanical issues requiring attention. The actual motor or equipment current is expressed as a percentage of its rated full-load current. This ensures that operating conditions can be monitored relative to design capacity.
Apron Feeder Drive Speed Indication	% RPM	2	Drive speed directly affects the material feed rate into the grizzly screen. Variations in speed can cause inconsistent feeding, impacting downstream processes like crushing and screening efficiency. Monitoring this parameter allows for proper adjustment of feed rates. The drive or motor speed is expressed as a percentage of its rated speed, providing a normalised measure of speed control.
Jaw Crusher Current	% rated current (A)	4	The jaw crusher consumes significant power during material size reduction. Monitoring the crusher current helps identify operational efficiency, such as excessive loads caused by oversized material. Abnormal current readings can also indicate mechanical wear or blockages. The actual motor or equipment current is expressed as a percentage of its rated full-load current. This ensures that operating conditions can be monitored relative to design capacity.
Stockpile Feeder Conveyor Weight Average Indication	%	6	This parameter measures the weight of material transported to the stockpile. It ensures that the material flow matches production targets and identifies deviations that may signal feed inconsistencies or equipment inefficiencies upstream. The conveyor or feeder weight signal is expressed as a percentage relative to the design or calibrated throughput, indicating how close the system is operating to target values.
Stockpile Feed Conveyor Current	% rated current (A)	5	Monitoring conveyor current ensures that the conveyor is not overloaded and operates within its capacity. Overloading can lead to equipment breakdowns or inefficiencies, while underloading indicates material flow problems upstream. The actual motor or equipment current is

			expressed as a percentage of its rated full-load current. This ensures that operating conditions can be monitored relative to design capacity.
Percentage Feed Material Under 80% of Target Particle Size Distribution (PSD)	% feed material	1	This parameter monitors the PSD of the material being fed into the circuit. Material that does not meet PSD targets may affect downstream crushing performance and energy consumption, requiring corrective measures. It is the proportion of feed material meeting specification, expressed as a percentage of total product mass.
Percentage Product Material Under 80% of Target PSD	% product material	5	This parameter ensures that the product material meets the desired particle size specification after crushing and screening. Achieving consistent PSD is crucial for maximising throughput and downstream milling efficiency. It is the proportion of final product material meeting specification, expressed as a percentage of total product mass.

The key input and output parameters were finalised as part of the development of the DT, as shown in Figure 3.9. The key output parameter (percentage product material under 80% of target) was selected, as this is the key output of the crusher. This effectively determines the crushed material size that comes out of the crusher, and it is a proxy to the plant achieving its production target as per guidance from the board.

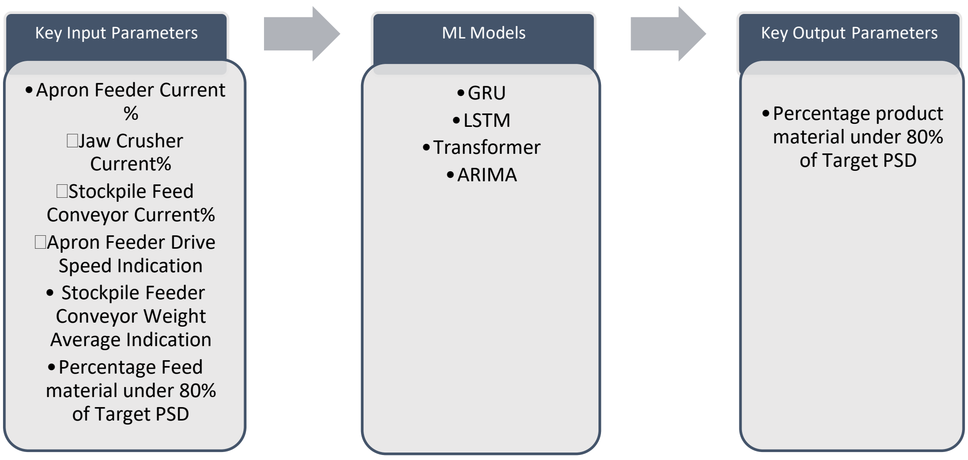

Figure 3.9: Key input and output monitoring parameters for Digital Twin development

3.2.1.2 Data Cleaning

The data cleaning process was required to improve the data quality and help provide a more accurate representation of the mechanical equipment [10]. It was noted from the site team that the crushing plant had a 75% availability. Hence, there is a significant period of time when the target particle size distribution (PSD) was zero, as shown in Figure 3.10. This was removed from the dataset as only occurrences when the crushing is active will form the DT.

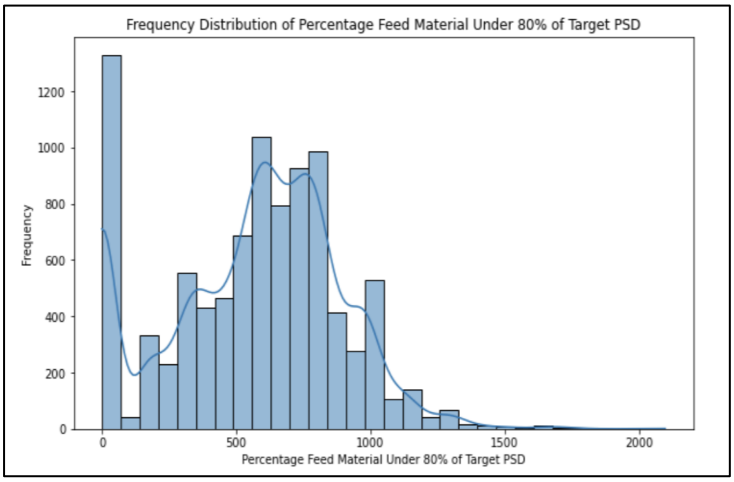

Figure 3.10: Percentage feed material under 80% of target particle size distribution (PSD) (pre-filtered)

The hydraulic pressure is an important indication, as it signals when the crusher is not functional. This is when the pressure is below 120 Nm^2, which would indicate that the crusher has stopped for maintenance or tripped from a fault occurrence. The parameter was used to remove any additional data when the crusher was not functional. The jaw crusher hydraulic indication is shown in Figure 3.11 from 04/01/2022 to 06/01/2022. The hydraulic pressure dropped below 120 Nm^2 at least once a day, so this data needed to be removed.

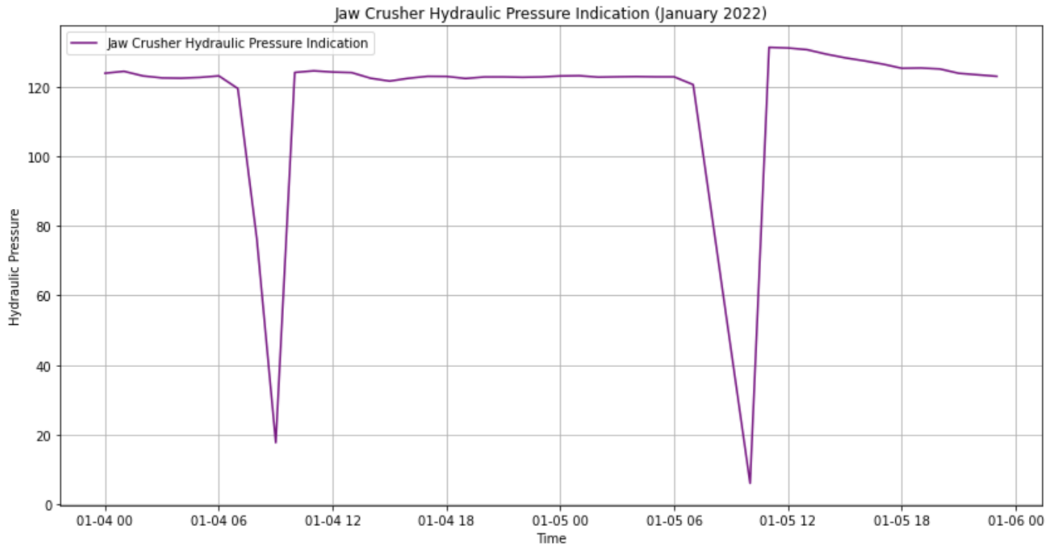

Figure 3.11: Jaw crusher hydraulic pressure indication for January 2022

3.2.1.3 Evaluation of Key Parameter Correlation

To evaluate the strength of correlations of the data, the correlation coefficient is used, typically denoted as r in a Pearson correlation matrix [45]. The strength of the correlation between two variables is measured using the absolute value of *r*. A perfect correlation, $r = 1$, indicates the strongest possible relationship between variables. A strong correlation falls within the range of $0.7 \leq r < 1$, signifying a strong linear relationship [45]. A moderate correlation, with values between $0.5 \leq r < 0.7$, suggests a noticeable but less pronounced relationship [45]. A weak correlation occurs within the range of $0.3 \leq r < 0.5$, indicating a limited linear association. If $r < 0.3$, the correlation is considered very weak or negligible. Finally, when $r = 0$, there is an absence of linearity between the variables, meaning changes in one variable do not predict changes in the other [45].

To refine the analysis, parameters with weak or no correlation were excluded. The remaining parameters, showing moderate to strong correlations, were retained to effectively depict the crushing circuit's behaviour and performance.

3.2.1.4 Data Normalisation and Standardisation

In this study, a data preprocessing strategy was implemented to standardise the dataset before training the model. The process for normalising the data for the ML problem involves transforming each variable using the formula below [46]:

$$x_i = \frac{x_i - \mu_i}{\sigma_i} \tag{3.1}$$

Where μ is the mean, and σ is the standard deviation computed on the training data. The target variable and the input features are scaled separately [46]. This distinction is crucial because it often represents different physical quantities [46]. Standardising the data ensures that all features contribute proportionally during training, which prevents features with higher ranges from skewing the learning process. This step not only enhances numerical stability but also smooths the loss landscape, facilitating faster convergence of gradient-based optimisation algorithms [46].

3.2.1.5 Smoothing the Target Parameter

The rolling window mean was utilised for the target parameter, which calculates the mean for the last observation points [47]. The rolling mean is calculated using the formula:

$$\boldsymbol{Rolling\ Mean_t} = \frac{x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-(n-1)}}{n} \tag{3.2}$$

Where x_t is the target parameter at time t, n is the window size, and t is the current time index. The rolling window reduces the noise and any short-term fluctuations in the datasets, which will assist in improving the model's predictive performance.

3.2.1.6 Create Sequence

The creation of sequences is a critical step in preparing time-series data for forecasting. The selected approach was to generate training samples by sliding a fixed-length window over the data, where each window serves as the input sequence. The subsequent time steps form the corresponding output sequence. The function iterates over the entire dataset and, for each position, extracts a sequence of input features of length equal to the defined sequence length and then assigns the following prediction length time steps as the target. This method transforms a continuous time series into a set of discrete, fixed-size samples, allowing the model to comprehend temporal dependencies and trends within the historical window to predict future values.

This sequence generation process is essential for capturing the inherent structure of the time-series data and ensures that the deep-learning model can exploit patterns over both short- and long-term horizons.

3.2.1.7 Splitting Data into Test and Training

In this study, the time-series data were split chronologically to prevent data leakage and maintain the temporal order crucial for realistic forecasting. Specifically, a cutoff date, where all observations prior to this date were exclusively utilised as the training set, allowing the ML models to capture the inherent temporal dynamics. Data from the cutoff date onward were reserved strictly for testing, ensuring that model evaluation was conducted on unseen, future-like scenarios. This approach aligns with established time-series forecasting practices [48].

To rigorously evaluate the forecasting model while maintaining temporal integrity, the time series splits cross-validation approach from scikit-learn, setting the number of splits to five [49]. This method sequentially divides the dataset into training and validation sets, with each fold expanding the training set forward in time, thereby respecting the chronological order of observations. Each fold uses historical data for training and immediate future data for validation, closely simulating real-world forecasting scenarios [49]. This setup provided a robust validation strategy, ensuring the evaluation captured realistic model performance without data leakage [49].

3.2.2 Model Training and Forecasting

This section details the process of training the GRU, LSTM, ARIMA, and Transformer models. The training phase involves optimising the model's parameters by hyper-tunning it, while carefully managing issues in time-series forecasting, such as vanishing gradient [27]. The aim is to ensure stable convergence and effective learning of the temporal dependencies inherent in the data [27].

3.2.2.1 Gated Recurrent Unit Model

The GRU uses the time series data $(x_1, x_2, \dots., x_T)$, representing the input over T timesteps. The input sequences consist of batched time-series data with a fixed look-back window (sequence length) and multiple features (input size), which are processed by stacked GRU layers that generate hidden states for the entire sequence [28]. From these, the hidden state corresponding to the last time step is selected to capture the learned temporal representation and then passed through a Rectified Linear Unit activation for enhanced non-linearity. This activated output is fed into a fully connected layer that maps it to the required prediction dimension (product of the prediction and length output size), and the resulting vector is reshaped into a tensor with the shape (batch size, prediction length, output size). The final output constitutes the multi-step forecast for the specified prediction horizon. The flowchart of the GRU modelling is shown in Figure 3.12 [28]. In the bidirectional configuration, the GRU processes the sequence in both directions, and the resulting hidden states are concatenated to form a richer representation [28].

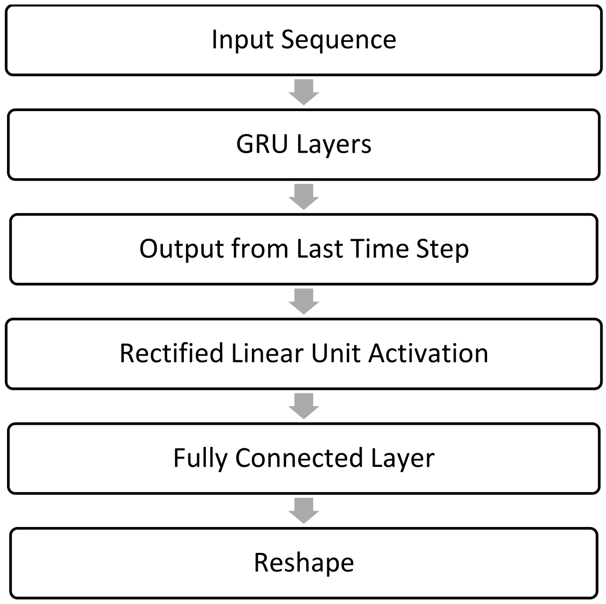

Figure 3.12: Flow chart of the Gated Recurrent Unit (GRU) modelling process for the crushing circuit

3.2.2.2 Long Short-Term Memory Model

The LSTM model processes input sequences of shape (batch size, sequence length, input size) that represent the time series data $(x_1, x_2, \dots, x_T)$. First, the model initialises the hidden state h_0 and cell state c_0 as zero tensors of shape (number of layers, batch size, hidden size) [27]. These states, along with the input, are fed into a multi-layer LSTM that captures the temporal dependencies across the sequence. From the output produced at each time step by the LSTM, the output corresponding to the final time step is selected as it encodes the complete temporal representation of the input [27]. This output is then passed through a Rectified Linear Unit activation to introduce non-linearity, enhancing the model's capacity to comprehend complex trends. Subsequently, a fully connected layer transforms the activated hidden state into the target output dimension prediction, which is the product of the prediction length and output size. Finally, the resulting vector is reshaped into a tensor of shape (batch size, prediction length, output size) that represents the multi-step forecast for each sequence in the batch. The flowchart of the LSTM modelling is shown in Figure 3.13 [27]. In the bidirectional configuration, the LSTM processes the sequence in both directions, and the outputs are concatenated at each time step from both directions to form a more comprehensive temporal representation [27].

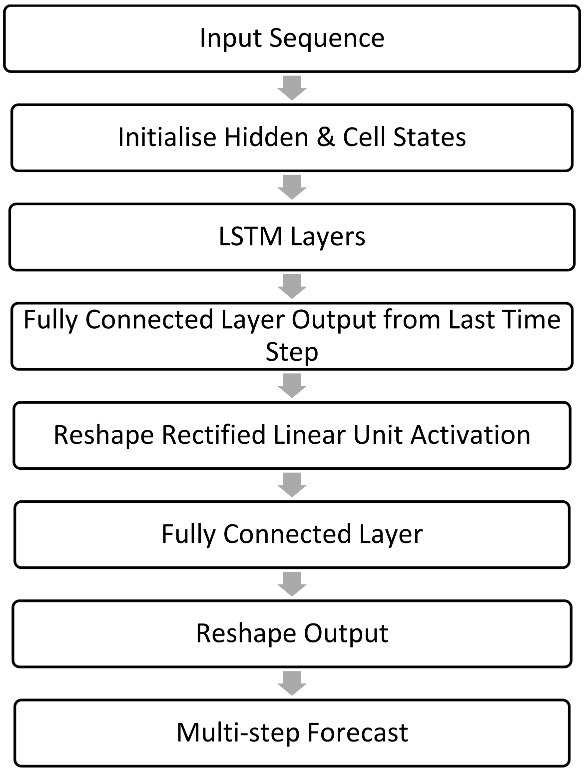

Figure 3.13: Flow chart of the Long Short-Term Memory Model (LSTM) modelling process for the crushing circuit

3.2.2.3 Transformer Model

The Transformer model for time-series forecasting begins by projecting the input sequences of shape [batch size, sequence length, input dimension] into a higher-dimensional space using a linear embedding layer [30]. Learnable positional encodings are then added to these embeddings to capture the temporal order of the data [30]. The enriched sequence is fed into a Transformer encoder composed of multiple layers, where each layer uses multi-head self-attention and feed-forward networks to capture global dependencies and refine the sequence representation [30]. From the encoder's output, the representation corresponding to the last time step is extracted and passed through a dense layer that maps the hidden representation to the required prediction length [30]. Finally, the output is reshaped into a tensor with dimensions (batch size, prediction length, output dimension), resulting in multi-step forecasts for the target feature. The flowchart of the Transformer modelling is shown in Figure 3.14 [30].

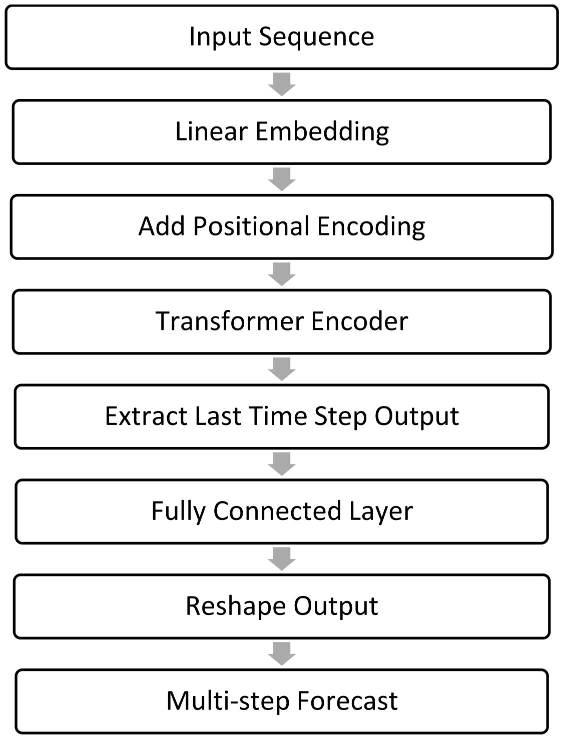

Figure 3.14: Flow chart of the Transformer modelling process for the crushing circuit

3.2.2.4 Autoregressive Integrated Moving Average Model

The parameters (p, d, q) for the ARIMA model involve several key steps. First is to check that the data was stationary using the Augmented Dickey-Fuller test [47]. The next step is the Autocorrelation Function and Partial Autocorrelation Function visuals, which are used to guide initial estimates for the AR and MA order. Ultimately, the best model is the one that minimises Akaike Information Criterion.

The pmdarima's auto ARIMA function automates the selection of optimal ARIMA parameters by systematically exploring various combinations of (p, d, q) and choosing the model that minimises information criteria such as the Akaike Information Criterion [47]. This process yields an ARIMA model that effectively captures the linear dynamics of the target series.

3.2.2.5 Autoregressive Integrated Moving Average and Recurrent Neural Network Hybrid Model

The ARIMA model is fit on the flattened training target series. This model generates forecasts over the test horizon, and its outputs are inverse-transformed to return to the original scale. In parallel, a GRU model is trained on the scaled sequences and produces its own forecasts, which are also inverse-transformed [34]. To create the hybrid forecast, the predictions from the ARIMA model and the GRU model are combined using a weighted average. This hybrid approach uses the linear trend capturing ability of ARIMA along with the non-linear, deep-learning capabilities of the GRU to potentially improve overall accuracy of the predictions. The flowchart of the ARIMA modelling is shown in Figure 3.15 [34].

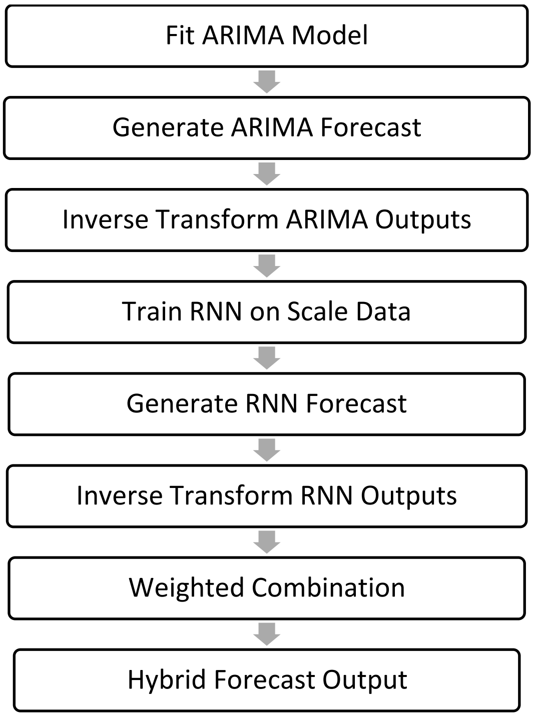

Figure 3.15: Flow chart of the Autoregressive Integrated Moving Average (ARIMA) and Recurrent Neural Network (RNN) modelling process for the crushing circuit

3.2.2.6 Model Optimisation

The optimal hyperparameters for the LSTM, GRU, ARIMA, and Transformer models were tuned using the open-source hyperparameter framework Optuna [50]. Hyperparameters such as the hidden size, number of layers, dropout, learning rate, muti-head attention, and batch size were evaluated to identify the hyperparameters with the largest impact on the coefficient of determination (R^2) score using the Bayesian Optimisation approach [50]. The optimisation problem is solved using the formula below:

$$\boldsymbol{\theta}^* = \boldsymbol{arg}\max_{\boldsymbol{\theta}\in\boldsymbol{\mathcal{H}}} \boldsymbol{\mathcal{M}}(\boldsymbol{\theta}) \tag{3.3}$$

Where θ is the chosen set of hyperparameters, $\mathcal{H}$ is the search space and $\mathcal{M}(\theta)$ is the objective function, which returns the performance metric. The function $\mathcal{M}(\theta)$ can be expressed below for this study:

$$\boldsymbol{\mathcal{M}(\theta)} = \boldsymbol{R^2(y_{true,}\hat{y}_{inv})} \quad (3.4)$$

Where y_{true} is the actual observed values from the dataset, and $\hat{y}_{inv}$ is the model predictions to determine R^2. To achieve the best hyperparameters(θ^*), the following search space ($\mathcal{H}$) was chosen, as shown in Table 3.3.

Table 3.3: The hyperparameters selected, with the indicated notation and the respective search space

Hyperparameter	Variable Notation	Search Space for Model Optimisation
Hidden Size	α	$\alpha \in \{256, 512,1024\}$
Number of Layers	β	$\beta \in \{1, 2, 3\}$
Dropout Rate	γ	$\gamma \in [0.1,0.5]$
Learning Rate	δ	$\delta \in [10^{-4}, 10^{-2}]$
Batch Size	σ	$\sigma \in \{32, 64, 128\}$
Prediction Length (Hours)	φ	$\varphi \in \{96, 144, 192,240\}$
Attention Heads	ω	$\omega \in \{16, 32, 64\}$

The models were trained for 500 epochs using the Huber loss function with gradient clipping to stabilise the training, while the Adam Optimiser was employed to update model parameters, as it uses adaptive moment estimation to efficiently handle sparse gradients and noisy data [50]. Optuna iteratively runs the objective function for a specified number of trials, each time testing different hyperparameter combinations, and then selects the configuration that yields the best performance as measured by the R^2 score.

3.3 Performance Evaluation

The verification process involved ensuring that the DT accurately modelled the physical asset. This was done by comparing the actual model's parameters to those predicted by the DT model and assessing whether the trend in the physical model matched the predicted trend. The validation stage ensured that the DT served its intended purpose to accurately predict and reflect the behaviour of the physical asset [17].

The predicted output of the DT was tested against the actual data using the coefficient of determination (R^2) and root mean square error (RMSE) to measure the accuracy of the model [51]. Lower RMSE values indicate a higher accuracy and better performance for the specific model. A higher R^2 indicates that the model is predicting accurately using the Pearson relationship stated in Section 3.2.1.3. The mean absolute error (MAE) is used to evaluate the absolute difference between the DT's predictions and the actual observed data [51]. A lower MAE indicates that the DT is accurately modelling the actual asset with a smaller error [51]. The accuracy of the DT is measured using the mean absolute percentage error (MAPE) [52]. A lower MAPE indicates a better prediction accuracy [52].

$$\boldsymbol{R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{3.5}$$

$$\boldsymbol{RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}} \tag{3.6}$$

$$\boldsymbol{MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|} \tag{3.7}$$

$$\boldsymbol{MAPE = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i - \hat{y}_i}{y_i}| \times 100} \tag{3.8}$$

Using the R^2, MAE, MAPE, and RMSE gave significant insight into the accuracy of the DT. These four evaluation criteria were sufficient to determine the best ML model for the crushing circuit DT.

The accuracy of the DT is a critical factor in determining its reliability and usefulness for operational decision-making. An example of a study in the South African context was completed for compressed air for a Gold mine [53]. The study utilised a mathematical regression-based model and achieved an R^2 of 0.646 [53]. The simulation identified pressure shortfalls before they occurred, and the implemented solution increased the compressed air pressure by 32%, directly preventing production losses [53]. This indicates that a moderately strong correlation can improve mining production.

3.4 Chapter Summary

This chapter outlined the structured approach used to develop a DT for the crushing circuit of a South African mine. The methodology was designed to ensure that relevant operational data was collected, processed, and used to train various ML models, including LSTM, GRU, ARIMA Hybrid, and Transformer models. The data preprocessing phase ensured that inconsistencies and anomalies were addressed, enabling accurate model training and evaluation.

To determine the most effective model for DT implementation, multiple performance metrics were used, including R^2, RMSE, and MAE, allowing for a comparative analysis of each model's predictive capabilities. Additionally, hyperparameter optimisation was conducted to enhance model efficiency and accuracy.

With the methodology now established, the next chapter presents and discusses the results obtained from the model evaluations. The findings highlight the comparative performance of the selected models and their effectiveness in replicating and predicting the behaviour of the crushing circuit.

4 Chapter 4: Results and Discussion

The results and discussion are presented in this chapter, from the comprehensive evaluation of ML algorithms to support the study. Each model was evaluated using Google Co-lab software due to the high computational demand. The A100 with 40 GB GPU RAM and 83.5 GB System RAM was used for the complex models such as the Transformer model. The remaining models were simulated using the L4 with 22.5 GPU RAM and 53 GB[1]. The sample code hyperlink is shown in Appendix B.

The performance evaluation of the selected ML models (LSTM, GRU, Transformer, and ARIMA) used to develop the DT for the crushing circuit is presented. The models were developed to forecast the percentage product material under 80% of the target PSD.

Evaluation metrics, including RMSE, MAE, and R^2, are utilised to quantify the reliability and accuracy of the models. The analysis focuses on identifying the best-performing model for dynamic system prediction while exploring each model's strengths and limitations. Key findings are contextualised to highlight their practical implications for predictive maintenance and operational optimisation in the mining environment.

4.1 Exploratory Data Analysis for Crushing Parameters

To visualise the relationships between the selected variables and their influence on one another, the correlation map approach was used. To identify the parameters with strong to moderate correlation, those correlations that were greater than 0.5 were selected for the study, as shown in Figure 4.1.

[1] GPU: Graphics Processing Unit; GB: Gigabytes; RAM: Random Access Memory

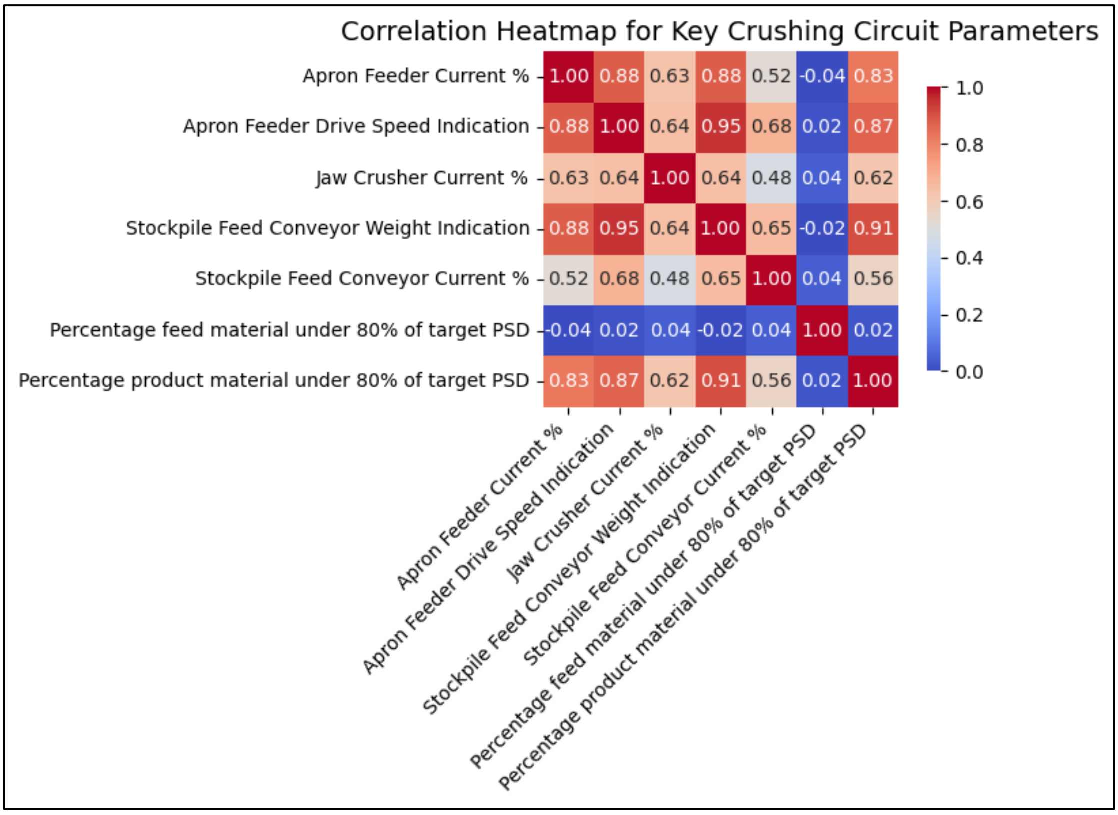

Figure 4.1: Crushing circuit parameter correlation heatmap

The correlation heatmap in Figure 4.1 indicates varying strengths of relationships among the selected parameters. The apron feeder current, apron feeder drive speed, and the stockpile feed conveyor weight indication show a strong positive correlation with values ranging between 0.7 and 0.91 with the target parameter, as seen in the heatmap. The jaw crusher current and stockpile feed conveyor current have a moderate correlation with the target parameter, ranging from 0.56 to 0.62. These achieved correlations reflect a direct relationship among these parameters, implying that changes in one parameter significantly influence the others.

Conversely, the percentage feed material under 80% of the target PSD is weakly correlated with the target parameter, with scores close to zero. This is due to the inconsistent feed provided by mining; hence, there is no fixed feed that is fed into the plant. Therefore, this parameter was excluded from the study.

4.2 Hyperparameter Optimisation

To achieve optimal performance for the LSTM, GRU, ARIMA, and Transformer models, hyperparameter optimisation was conducted using Optuna, an advanced optimisation framework tool, as described in Section 3.2.2.6.

The optimisation results for the GRU, LSTM, ARIMA, and Transformer models highlight the efficiency of hyperparameter tuning in enhancing predictive accuracy for DT applications. The optimal hyperparameters for the LSTM, GRU, ARIMA, and Transformer models are presented in Table 4.1.

Table 4.1: Hyperparameter for the Gated Recurrent Unit (GRU), Long Short-Term Memory (LSTM), Autoregressive Integrated Moving Average (ARIMA), and Transformer models for the crushing circuit

Model	δ	φ	β	α	γ	σ	ω
GRU	$1.5e^{-4}$	192	2	1024	0.2	64	N/A
LSTM	$3e^{-3}$	192	1	1024	0.47	64	N/A
ARIMA	N/A	5	N/A	N/A	N/A	N/A	N/A
Transformer	$9e^{-4}$	192	1	1024	0.16	64	64

The sequence length of 20 days (480 hours) and prediction length (φ) of eight days (192 hours) yielded the highest R^2 for all the models, except for the ARIMA. This optimal point aligns well with the model dynamics to learn and predict the target effectively. The ARIMA model forecasts over a shorter period, i.e., five hours, due to its limitation for long-term nonlinear patterns [29]. The LSTM model had a significantly higher dropout rate (ϒ) compared to the Transformer and GRU models, due to the more internal gates of the LSTM [27].

4.3 Gated Recurrent Unit Model Performance

The target parameter results using the GRU model are shown in Figures 4.2 and 4.3. The results were based on the hyperparameters specified in Section 4.1.

Figure 4.2: The graphs show the predicted vs actual data folds for the percentage product material under 80% particle size distribution (PSD) using the Gated Recurrent Unit (GRU) model

Table 4.2: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) crushing circuit scores of the Gated Recurrent Unit predictions average fold for the percentage material under 80% particle size distribution

Target Parameter	Average R^2	Average RMSE	Average MAE	Average MAPE (%)
Percentage material <80% particle size distribution	0.54	6.39	5.16	2.73

Across the five folds for the GRU model, there is a general, consistent trend of alignment between the actual and predicted data, with the target parameter effectively tracked. The evaluation metrics presented in Table 4.2 revealed that the model moderately tracks the target, with some of the folds showing better performance. The overall results from the folds indicate that the GRU captures the dominant alignment with the target parameter.

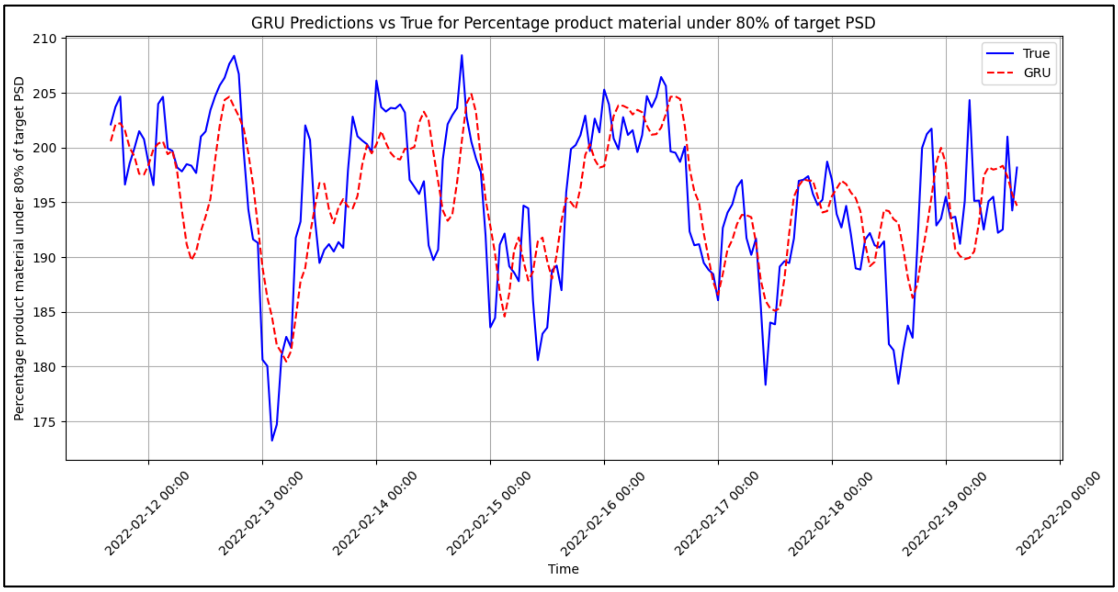

Figure 4.3: The graphs show the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Gated Recurrent Unit (GRU) model

Table 4.3: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) crushing circuit scores for the Gated Recurrent Unit predictions of the percentage material under 80% particle size distribution

Target Parameter	R^2	RMSE	MAE	MAPE (%)
Percentage material <80% particle size distribution	0.61	4.48	3.5	1.8

The GRU model achieved an R^2 of 0.61, as shown in Table 4.3, indicating a moderate correlation between the predicted and actual target parameters. This is complemented by an RMSE of 4.48 and an MAE of 3.5. These metrics indicate acceptable results, as the range of the target is between 150 to 200 mm. The MAPE is 1.8%, which is lower than the recommended threshold of 20% for a robust DT [52]. The visual inspection of the results confirms that the model tracks the trend but occasionally lags or overshoots during rapid change.

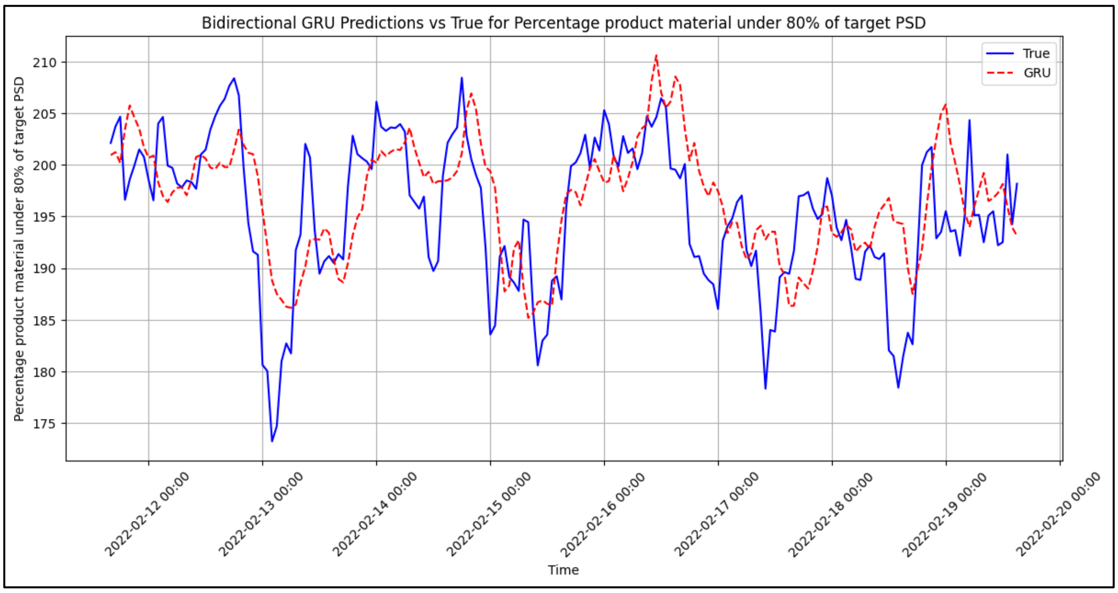

Figure 4.4: The graphs show the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the bidirectional Gated Recurrent Unit model

The results from the bi-GRU predictions are shown in Figure 4.4. The R^2 is 0.50, which is far lower than the standard GRU, shown in Figure 4.3.

4.4 Long Short-Term Memory Model Performance

The target parameter results using the LSTM model are shown in Figures 4.5 and 4.6. The results were based on the hyperparameters specified in Section 4.1.

Figure 4.5: The graphs show the predicted vs actual data folds for the percentage product material under 80% particle size distribution (PSD) using the Long Short-Term Memory model

Table 4.4: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) crushing circuit scores for the Long Short-Term Memory predictions average fold of the percentage material under 80% particle size distribution

Target Parameter	Average R^2	Average RMSE	Average MAE	Average MAPE (%)
Percentage material <80% particle size distribution	0.49	11.70	9.43	4.89

Across the five folds for the LSTM model, there is a general consistent trend of alignment between the actual and predicted data, with the target parameter effectively tracked. The evaluation metrics shown in Table 4.4 revealed that the model moderately tracks the target. However, the first and second fold shows a much weaker correlation, as these folds have the greatest range for the target parameter. The overall results from the folds indicate that the LSTM captures the dominant alignment with the target parameter.

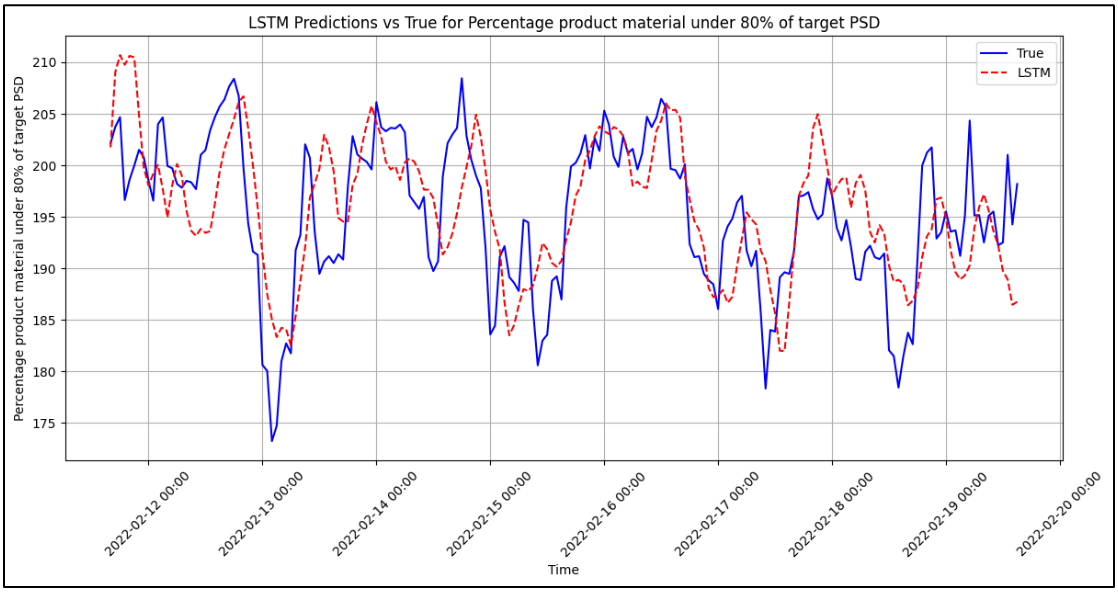

Figure 4.6: The graphs show the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Long Short-Term Memory model

Table 4.5: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) for the Long Short-Term Memory predictions of the percentage material under 80% particle size distribution

Target Parameter	R^2	RMSE	MAE	MAPE (%)
Percentage material <80% particle size distribution	0.55	4.83	3.84	1.98

The LSTM model achieved an R^2 of 0.55, as per Table 4.5, which revealed a moderate correlation between the predicted and actual target parameters. This is complemented by an RMSE of 4.83 and an MAE of 3.84. These metrics indicate acceptable results, as the range of the target is between 150 to 200 mm. The MAPE is 1.98%, which is lower than the recommended threshold of 20% for a robust DT [52]. The visual inspection of the results confirms that the model tracks the trend but occasionally lags or overshoots during rapid change.

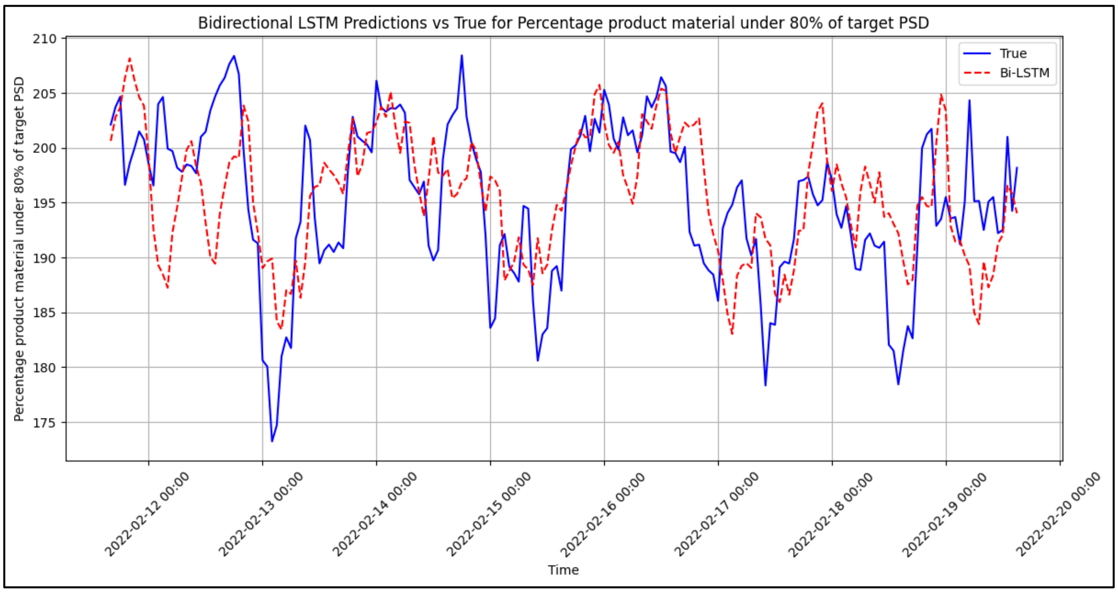

Figure 4.7: The graphs show the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the bidirectional Long Short-Term Memory model

The results from the bidirectional LSTM predictions are shown in Figure 4.7. The R^2 is 0.48, which is far lower than the standard LSTM shown in Figure 4.6.

4.5 The Autoregressive Integrated Moving Average Model

The results from the Augmented Dickey-Fuller test indicate that the ARIMA model is stationary; therefore, the d component would be 0. The Autocorrelation Function in Figure 4.8 displays significant autocorrelation for lags up to six and then tapers off. This pattern indicates that the q component of order six is needed to capture the dependence structure in the residua. The Partial Autocorrelation Function in Figure 4.9 shows a distinct cutoff after the first lag; this implies that only the first lag has a significant partial autocorrelation. This would suggest that a p greater than one would be acceptable. Using the pmdarima's auto_arima function confirmed that the ARIMA components would be (1,0,6).

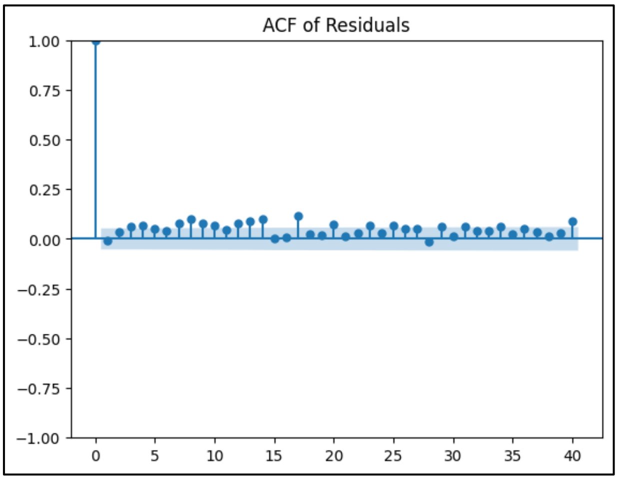

Figure 4.8: The graph shows the Autocorrelation Function (ACF) from the Autoregressive Integrated Moving Average predictions

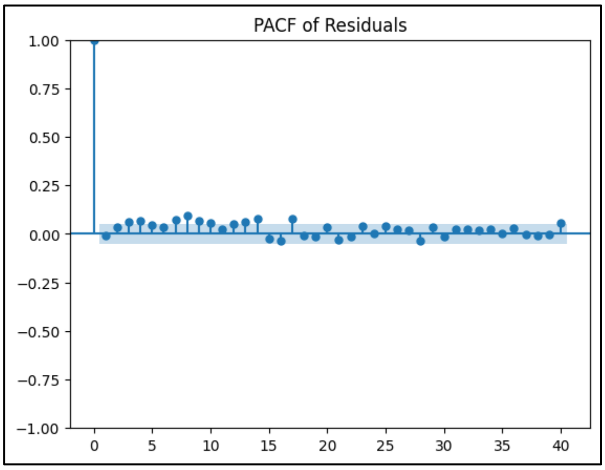

Figure 4.9: The graph shows the Partial Autocorrelation Function (PACF) from the Autoregressive Integrated Moving Average predictions

The predictions of the ARIMA are shown in Figure 4.10. The results show that the ARIMA predictions follow the overall trend. The ARIMA metrics are shown in Table 4.6. The R^2 of 0.52 indicates a moderate correlation, with RMSE (4.66) and MAE (3.62) scores relatively similar, which would indicate no major outliers skewing the model. The MAPE score of 1.87% indicates that the model is predicting accurately.

Table 4.6: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) for the autoregressive integrated moving average (ARIMA) predictions for the percentage material under 80% particle size distribution

Target Parameter	ARIMA R^2	ARIMA RMSE	ARIMA MAE	ARIMA MAPE (%)
Percentage material <80% particle size distribution	0.52	4.66	3.62	1.87

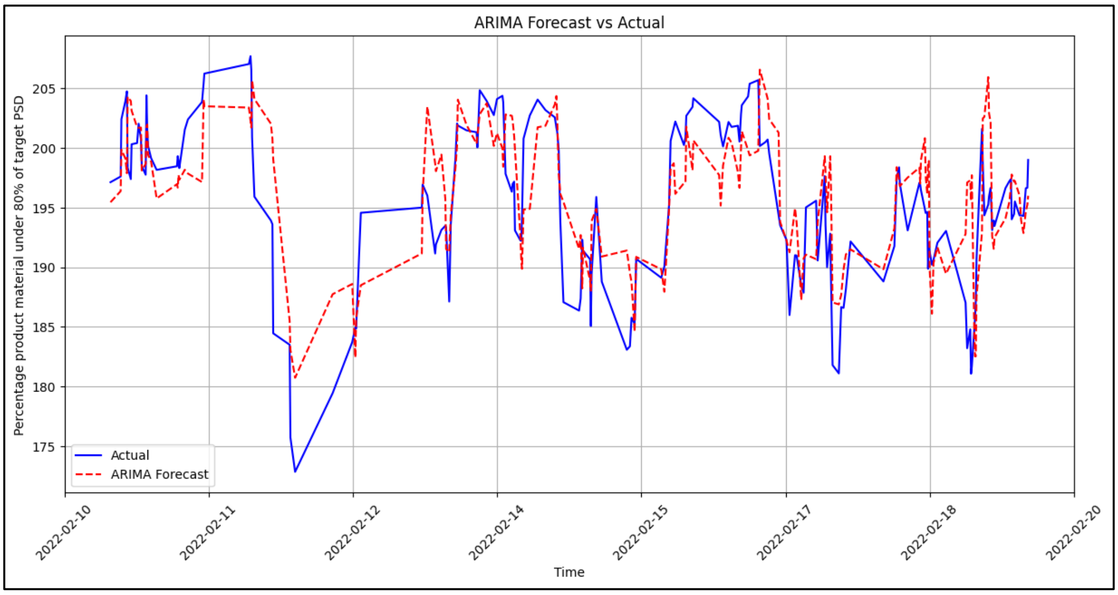

Figure 4.10: The plots of the actual vs prediction for the percentage product material under 80% of target particle size distribution (PSD) using the Autoregressive Integrated Moving Average model

4.6 The Autoregressive Integrated Moving Average and Recurrent Neural Network Hybrid Integration

The ARIMA was paired with both the LSTM and GRU models to determine if the combination would yield better R^2 results. The weighting between the GRU model and ARIMA, and the respective R^2 is shown in Table 4.7.

Table 4.7: The model weighting for the Gate Recurrent Unit (GRU) and Autoregressive Integrated Moving Average (ARIMA) vs the coefficient of determination (R^2)

Model Weighting	**R^2**
GRU 90%/ARIM 10%	0.56
GRU 80%/ARIM 20%	0.55
GRU 70%/ARIM 30%	0.58
GRU 60%/ARIM 40%	0.49
GRU 50%/ARIM 50%	0.42

4.6.1 Hybrid Model Autoregressive Integrated Moving Average and Gated Recurrent Unit

The target parameter results using the GRU and ARIMA models are shown in Figures 4.11. The results were based on the hyperparameters specified in Section 4.1 and the ARIMA components in Section 4.5.

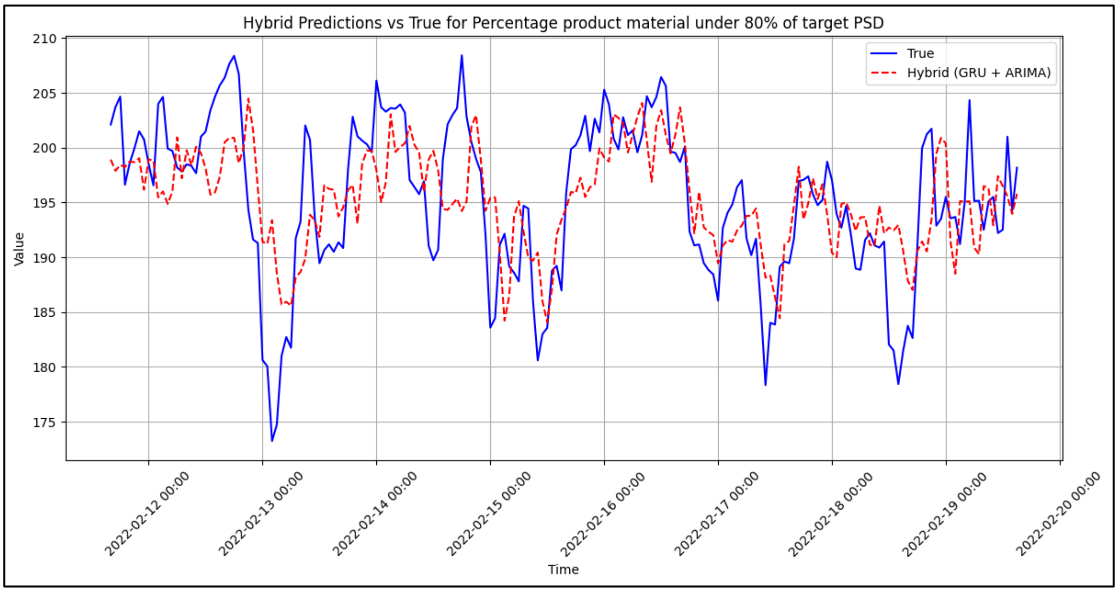

Figure 4.11: The graphs show the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Gated Recurrent Unit (GRU) + Autoregressive Integrated Moving Average (ARIMA) model

Table 4.8: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) for the Gated Recurrent Unit + Autoregressive Integrated Moving Average model predictions for the percentage material under 80% particle size distribution

Target Parameter	**R^2**	**RMSE**	**MAE**	**MAPE (%)**
Percentage material <80% particle size distribution	0.53	6.21	5.45	4.95

The hybrid GRU and ARIMA model achieved an R^2 of 0.53, as shown in Table 4.8, which reveals a moderate correlation between the predicted and actual target parameters. This is complemented by an RMSE of 6.21 and an MAE of 5.45. These metrics indicate acceptable results, as the range of the target is between 150 to 200 mm. The MAPE is 4.95%, which is lower than the recommended threshold of 20% for a robust DT [52]. The visual inspection of the results confirms that the model tracks the trend but occasionally lags or overshoots during rapid change.

4.6.2 Hybrid Model Autoregressive Integrated Moving Average and Long Short-Term Memory

The target parameter results using the LSTM and ARIMA models are presented in Figure 4.12. The results were based on the hyperparameters specified in Section 4.1 and the ARIMA components in Section 4.5.

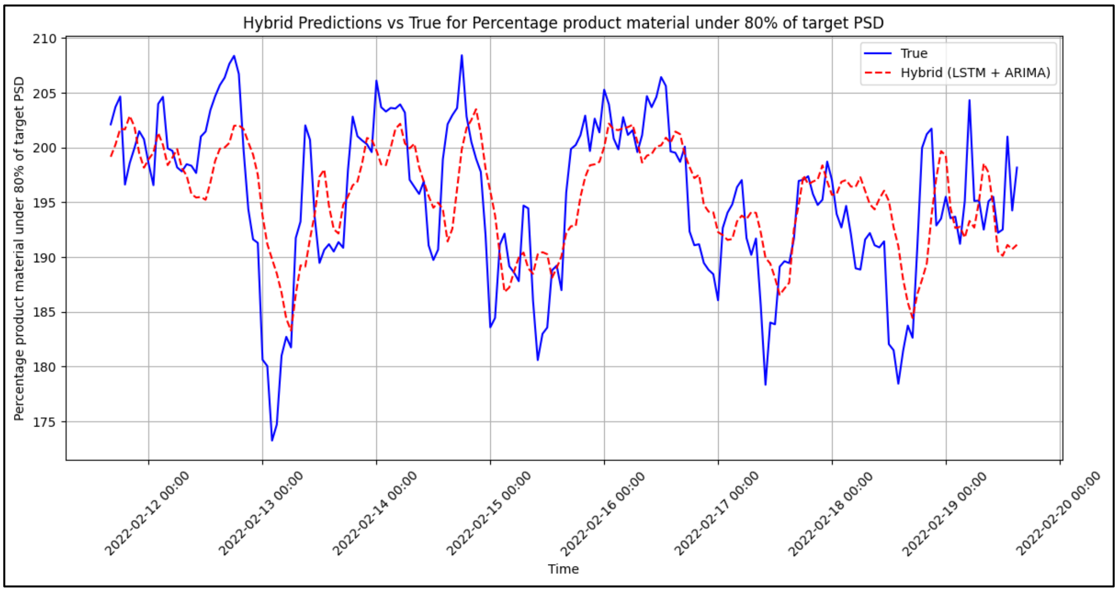

Figure 4.12: The graph shows the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Long Short-Term Memory (LSTM) + Autoregressive Integrated Moving Average (ARIMA) model

Table 4.9: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) for the Long Short-Term Memory + Autoregressive Integrated Moving Average predictions for the percentage material under 80% particle size distribution

Target Parameter	R^2	RMSE	MAE	MAPE (%)
Percentage material <80% particle size distribution	0.52	6.58	5.82	6.78

The hybrid LSTM and ARIMA model achieved an R^2 of 0.52, as shown in Table 4.9, which indicates a moderate correlation between the predicted and actual target parameters. This is complemented by an RMSE of 6.58 and an MAE of 5.82. These metrics indicate acceptable results, as the range of the target is between 150 to 200 mm. The MAPE is

6.78%, which is lower than the recommended threshold of 20% for a robust DT [52]. The visual inspection of the results confirms that the model tracks the trend but occasionally lags or overshoots during rapid change.

4.6.3 Transformer Model Performance

The target parameter results using the Transformer model are shown in Figures 4.13 and 4.14. The results were based on the hyperparameters specified in Section 4.1.

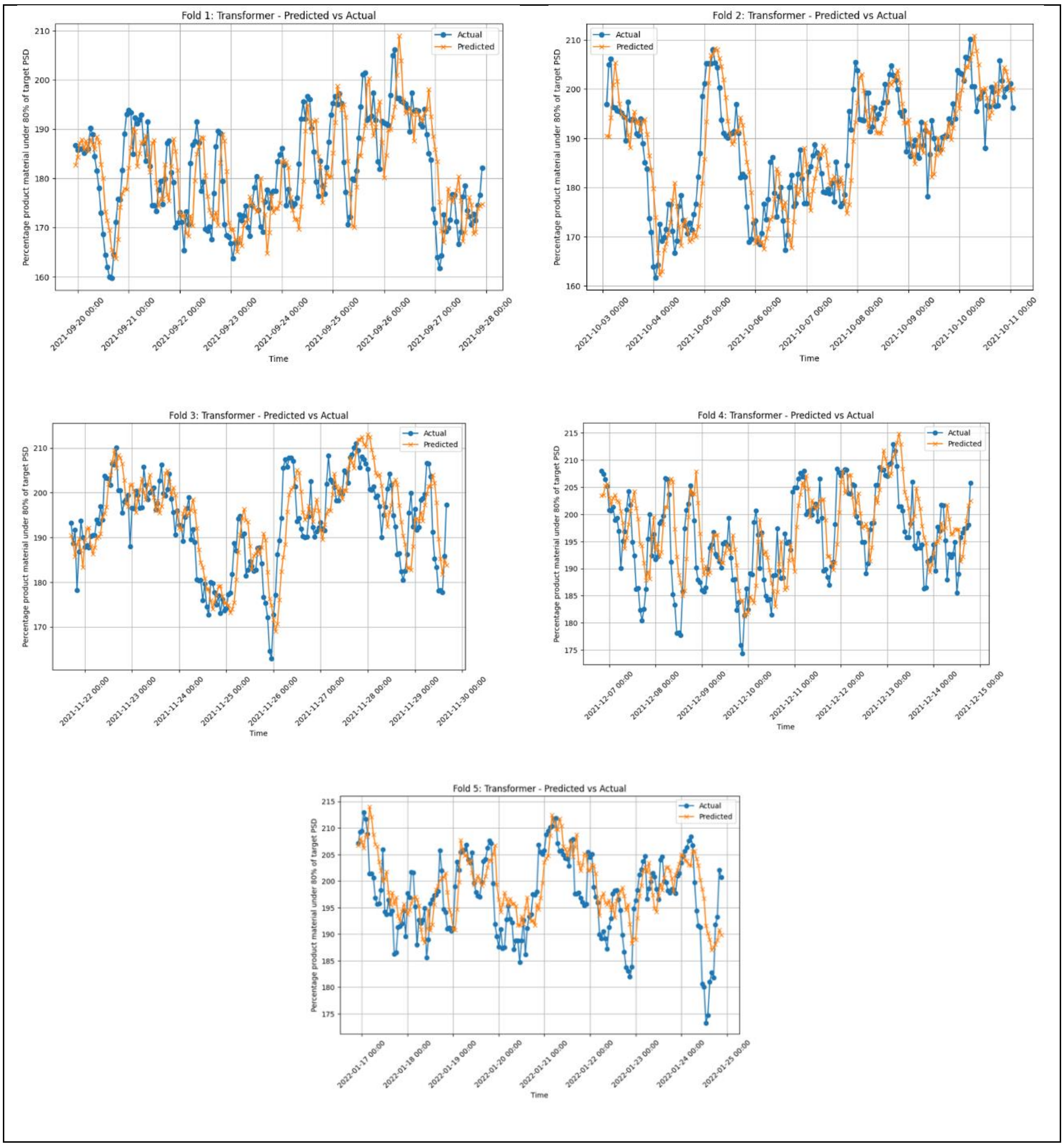

Figure 4.13: The graphs show the folds for the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Transformer model

Table 4.10: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) crushing circuit scores for the Transformer predictions average fold of the percentage material under 80% particle size distribution

Target Parameter	Average R^2	Average RMSE	Average MAE	Average MAPE (%)
Percentage material <80% particle size distribution	0.36	12.17	9.77	5.08

Across the five folds for the Transformer model, there is a general, consistent trend of alignment between the actual and predicted data, with the target parameter effectively tracked. The evaluation metrics presented in Table 4.10 revealed that the model moderately tracks the target, with some of the folds showing better performance. The overall results from the folds indicate that the Transformer model captures the dominant alignment with the target parameter.

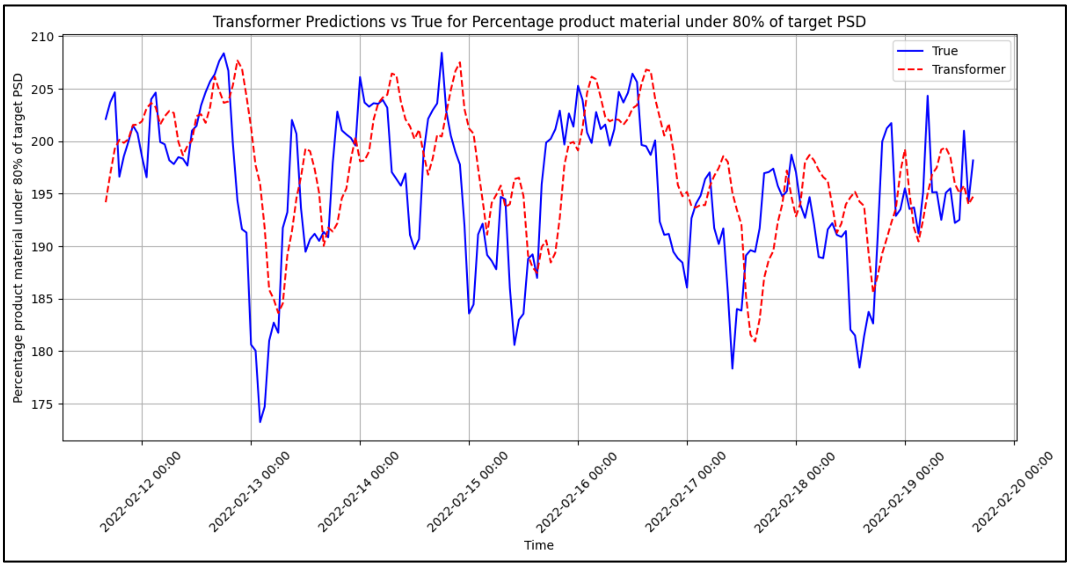

Figure 4.14: The graph shows the predicted vs actual data for the percentage product material under 80% particle size distribution (PSD) using the Transformer model

Table 4.11: The coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE) for the Transformer model predictions of the percentage material under 80% particle size distribution

Target Parameter	R^2	RMSE	MAE	MAPE (%)
Percentage material <80% particle size distribution	0.33	5.92	4.53	2.34

The Transformer model achieved an R^2 of 0.33, as shown in Table 4.11, which indicates a moderate to weak correlation between the predicted and actual target parameters. This is complemented by an RMSE of 5.92 and an MAE of 4.53. These metrics indicate acceptable results, as the range of target is between 150 to 200 mm. The MAPE is 2.34%, which is lower than the recommended threshold of 20% for a robust DT [52]. The visual inspection of the results confirms that the model tracks the trend but regularly lags or overshoots from the true values.

4.7 Chapter Summary

The results of this study provide a comprehensive evaluation of the LSTM, GRU, and Transformer models, both as standalone and hybrid architectures, in predicting the percentage product material under 80% PSD. Notably, the GRU model emerged as the best-performing approach, as it also required the least computational power and completed the simulations the quickest of all the models. The Transformer model performed the worst, as it demonstrated robust capabilities in capturing long-term trends; however, the R^2 is still less than the GRU. The Transformer model generally requires more data, so perhaps two to three years of data might be better suited for this model. The LSTM was also effective in modelling long-term dependencies; however, it could not match the overall performance of the GRU. The bidirectional LSTM and GRU models did not perform as well as the single-layered models.

The ARIMA on its own could not accurately model the target. This is due to the non-linearity of the target variable since the ARIMA only looked at the target variable. The testing of a hybrid approach of the ARIMA and RNN was completed. The results did not show signs of improvement, so the standalone RNN was still better. Overall, these findings suggest that, for predictive modelling in mining operations, the GRU is the most promising approach, followed by the LSTM and then the hybrid models, with the Transformer model performing the worst.

5 Chapter 5: Conclusion

This chapter synthesises the key findings of the study, evaluates their contributions to the research objectives, and highlights the significance of the results in addressing the challenges outlined in the problem statement. By reflecting on the research journey, this chapter provides a comprehensive and forward-looking perspective on the implications and potential of this work.

The value add of this study is that it addresses the gaps of having a structured data collection process and the limited application of DT technologies in South African mines. To overcome these challenges, this research developed and evaluated a DT framework using ML methods, targeting predictive maintenance and process optimisation in the crushing circuit of a South African mine.

5.1 Addressing the Research Questions and Objectives

The research question (Section 1.5) was: What is the best ML approach to build a DT for the crushing circuit as part of the process plant at the mine? This question was answered by developing DT models using different ML techniques and evaluating which is the most accurate model. This model was found to be the GRU model based on Table 5.1. The moderate accuracy of the models is evident for the GRU, LSTM, and ARIMA, with R^2 ranging between 0.5 to 0.7. The Transformer model had a weak R^2, due to the data being limited to a year. The MAPE remained consistently below 20% for all the models, which would indicate acceptable forecasting performance [52]. Based on the range of the datasets and the highly dynamic nature of the crusher circuit, the low MAE and moderately high R^2 confirm the robustness and practical value of the GRU model in the mining industry.

Table 5.1: The summary of the model's performance for the percentage material <80% particle size distribution, showcasing the coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE), and the mean absolute percentage error (MAPE)

Models	R^2	RMSE	MAE	MAPE (%)
GRU	0.61	4.48	3.5	1.8
LSTM	0.55	4.83	3.84	1.98
ARIMA	0.52	4.66	3.62	1.87
Transformer	0.33	5.92	4.53	2.34

The objectives were addressed by collecting time-series data, developing the DT models, and evaluating the models based on the prediction accuracy. The findings revealed that GRU networks outperformed other models, including the LSTM and Transformer models, in predicting key equipment behaviours. The GRU excelled in handling complex temporal patterns within the multivariate time-series dataset. The hybrid integration of ARIMA with the RNN model (LSTM or GRU) did show significant improvement, as the hybrid models did not perform as well as the GRU. The GRU was also the least CPU-intensive compared to the LSTM and Transformer models. This rationale also supports the goal of balancing performance with operational practicality, ensuring that the DT implementation is both effective and deployable within the constraints of a mining environment.

5.2 Implication of the Findings

The research demonstrates the feasibility of using ML to develop a DT for the crushing circuit, providing a structured framework for collecting, preprocessing, and analysing multivariate time-series data. The research addresses the gap in DT applications in South African mining by offering actionable insights for the crushing circuit.

5.3 Limitations of the Study

Despite its achievements, the study is not without limitations. The research focused solely on the crushing circuit, limiting the generalisability of the findings to other mining subsystems. Furthermore, the analysis was limited to one year of operational data, restricting the ability to capture long-term trends or external influences. Future research should explore the extension of DT applications to other stages of the mining value chain, incorporate larger datasets spanning multiple years, and experiment with additional hybrid ML models to further enhance predictive performance.

5.4 Recommendations

Based on the outcomes of this research, the recommendation is to adopt the GRU model into a real-time mining application. The model can be used as part of the mine's digital infrastructure to obtain real-time data for efficient decision making. In addition, plant personal should be upskilled in DT technology to support the successful deployment and maintenance of the DT system.

5.5 Chapter Summary

This study underscores the transformative potential of DT technology in the mining sector, particularly for optimising the crushing circuit, a critical component of the process plant. By integrating advanced ML techniques with real-time and historical data, the research provides a pathway for utilising the best ML methods for a crushing circuit DT. These contributions not only advance the field of digital transformation in mining but also offer practical solutions for enhancing the sustainability and competitiveness of South African mines in the global market.

5.6 Future Work

While this study successfully developed and evaluated a DT for the crushing circuit of a South African mine, significant opportunities exist for extending its scope and deepening its impact. Future research should focus on the practical implementation of the DT and ML models for predictive maintenance and energy optimisation across the mining process.

Future work should also focus on further hyperparameter tuning, particularly refining learning rates and architectural depth and exploring hybrid approaches that integrate Transformer components with LSTM or GRU to leverage their respective strengths.

References

1. G. Rajagopalan, "Industry value chain: Understand its importance and application to the mining industry," Flevy Blog. Accessed Jan. 29, 2025. [Online.] Available: https://flevy.com/blog/industry-value-chain-understand-its-importance-and-application-to-the-mining-industry/?srsltid=AfmBOopaQGBQLKGkEu0zLc1eiPhHuZGw7plp_FriLFC70zXcTmLZEJ0Q
2. Minerals Council South Africa. "Mining matters to South Africa and its people." Minerals Council South Africa. Accessed Jan. 27, 2025. [Online.] Available: https://www.mineralscouncil.org.za/all-categories?task=download.send&id=2251:mining-matters-to-south-africa-file-sheet&catid=3
3. M. R. Enders and N. Hoback. (2019). Dimensions of Digital Twin applications - A literature review. Presented at Twenty-fifth Americas Conference on Information Systems. [Online]. Available: https://www.researchgate.net/profile/Martin-Enders-4/publication/359715537_Dimensions_of_Digital_Twin_Applications_-_A_Literature_Review/links/624ae8d921077329f2f1fb4f/Dimensions-of-Digital-Twin-Applications-A-Literature-Review.pdf
4. SAP. "Value creation in the digital mine." SAP Industry Paper. [Online.] Available: https://illumiti.com/wp-content/uploads/2020/07/value-creation-in-the-digital-mine.pdf
5. IBM. "What is machine learning?" IBM Topics. Accessed Nov. 17, 2024. [Online.] Available: https://www.ibm.com/topics/machine-learning
6. P. G. Ranjith, J. Zhao, R. V. S. De Silva, T. D. Rathnaweera and A. K. M. S. Bandara, "Opportunities and challenges in deep level mining: A brief review," *Eng.*, vol. 3, no. 4, pp. 546–551, Aug. 2017, doi: 10.1016/J.ENG.2017.04.024.
7. M. Grieves, "Digital Twin manufacturing excellence through virtual factory replication." White Paper. 2014. pp. 1–7.
8. M. Singh, E. Fuenmayor, E. P. Hinchy, Y. Qiao, N. Murray and D. Devine, "Digital Twin: Origin to future," *Appl Syst. Innov*., vol. 4, no. 2, pp. 36–55, May 2021, doi: 10.3390/asi4020036.
9. J. Savolainen and M. Urbani, "Maintenance optimization for a multi-unit system with digital twin simulation: Example from the mining industry," *J. Intell. Manuf.*, vol. 32, pp. 1953–1973, Feb. 2021.
10. A. Mussomeli, B. Meeker, S. Shepley and D. Schatsky, "Expecting digital twins," Deloitte Insights. Accessed Jan 29, 2025. [Online.] Available: https://www2.deloitte.com/content/dam/insights/us/articles/3773_Expecting-digital-twins/DI_Expecting-digital-twins.pdf

11. Gartner. "Gartner survey reveals Digital Twins are entering mainstream use." Gartner. Accessed Oct 19, 2024. [Online.] Available: https://www.gartner.com/en/newsroom/press-releases/2019-02-20-gartner-survey-reveals-digital-twins-are-entering-mai
12. A. Gillespie, *Condition-based Maintenance: Theory, Maintenance and Application*. McLean: Science Applications International Corporation, 2015.
13. Q. Min, Y. Lu, Z. Liu, C. Su and B. Wang, "Machine learning based Digital Twin framework for production optimization in petrochemical industry," *IJIM*, vol. 49, pp. 502–519, Dec. 2019, doi:10.1016/j.ijinfomgt.2019.05.020
14. D. Ariansyah, I. F. del Amo, J. A. Erkoyuncu, M. Agha, D. Bulka, J. De La Puante, et al., "Digital Twin development: A step by step guideline," in *TESConf 2020*, Cranfield, UK, 2020.
15. L. Barnewold and B. G. Lottermoser, "Identification of digital technologies and digitalisation trends in the mining industry," *Int. J. Min. Sci. Technol*., vol. 30, no. 6, pp.747–757, Nov. 2020, doi:10.1016/j.ijmst.2020.07.003.
16. A. Gupta and D. S. Yan, *Mineral Processing Design and Operation*. Amsterdam: Elsevier Science, 2006, pp. 99–127.
17. R. G. Sargent, "Verification and validation of simulations models," in *Proc. 2010 Winter Simulation Conf.*, Baltimore, USA, Dec. 5-8, 2010, pp. 166–183.
18. D. Jones, C. Snider, A. Nassehi, J. Yon and B. Hicks, "Characterising the Digital Twin: A systematic literature review," *CIRP-JMST*, vol. 29, pp. 36–52, May 2020, doi:10.1016/j.cirpj.2020.02.002
19. P. Nobahar, C. Xu, P. Dowd and R. Shirani Faradonbeh, "Exploring digital twin systems in mining operations: A review," *Green Smart Min. Eng*., vol. 1, no. 4, pp.474–492, Dec. 2024, doi: 10.1016/j.gsme.2024.09.003
20. A. Johnson, J. Heaton, S. Yule, S. Luke, D. Pocock, A. K. Parlikad, et al., "Informing the information requirements of a digital twin: a rail industry case study," *ICE Smart Infrastruct*., vol. 174, no. 2, pp. 33–45, June 2022, doi: 10.1680/jsmic.20.00017
21. M. Azangoo, L. Sorsamäki, S. A. Sierla, T. Mätäsniemi, M. Rantala, K. Rainio, et al., "Methodology for generating a Digital Twin for process industry: A case study of a fiber processing pilot plant," *IEEE Access*, vol. 10, pp. 58787–58810, May 2022, doi: 10.1109/ACCESS.2022.3178424
22. University of Cambridge. "Gemini principles." University of Cambridge. Accessed Nov. 18, 2022. [Online.] Available: https://www.cdbb.cam.ac.uk/Resources/ResoucePublications/TheGeminiPrinciples.pdf
23. J. Heaton, A. K. Parlikad and J. Schooling, "A building information modelling approach to the alignment of organisational objectives to asset information requirements," *Autom. Constr*., vol. 104, pp. 14–26, Aug. 2019, doi: 10.1016/j.autcon.2019.03.022

24. C. Dilmegani, "Top 15 predictive maintenance tool & selection guide," AIMultiple. Accessed Oct. 15, 2024. [Online.] Available: https://research.aimultiple.com/predictive-maintenance/

25. J. Contreras, R. Espinola, F. J. Nogales and A. J. Conejo, "ARIMA models to predict next-day electricity prices," *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1014–1020, Aug. 2003, doi: 10.1109/TPWRS.2002.804943

26. R. Rana, J. Epps, R. Jurdak, X. Li, R. Goecke, M. Brereton, et al., "Gated Recurrent Unit (GRU) for emotion classification from noisy speech," arXiv:1612.07778, Dec. 2016.

27. G. Van Houdt, C. Mosquera and G. Nápoles, "A review on the long short-term memory model," *Artif. Intell. Rev.*, vol. 53, pp. 5929–5955, May 2020.

28. Y Wang, W. Liao, and Y. Chang, "Gated recurrent unit network-based short-term photovoltaic forecasting," *Energies*, vol. 11, no. 8, 2163, Aug. 2018, doi: 10.3390/en11082163

29. V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, 255, Jul. 2023.

30. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention is all you need," in *NIPS 2017*, Long Beach, CA, USA, Dec. 4-9, 2017.

31. X. Amatriain, A. Sankar, J. Bing, P. K. Bodigutla, T. J. Hazen and M. Kazi, "Transformer models: An introduction and catalog," arXiv:2302.07730, Apr. 2024.

32. B. Peng, E. Alcaide, Q. Anthony, A. Albalak, S. Arcadinho, S. Biderman, et al., "RWKV: Reinventing RNNs for the transformer era," arXiv:2305.13048.

33. P. Kruczek, N. Gomolla, J. Hebda-Sobkowicz, A. Michalak, P. Śliwiński, J. Wodecki, et al., "Predictive maintenance of mining machines using advanced data analysis system based on the cloud technology," in *MPES 2018*, E. Widzyk-Capehart, A. Hekmat and R. Singhal, Eds. 2019 pp. 459–470.

34. J. Yang, Y. Sun, Y. Cao and X. Hu, "Predictive maintenance for switch machine based on Digital Twins," *Information*, vol. 12, no. 4, 485, Nov. 2021, doi: 10.3390/info12110485

35. H. Ding, L. Yang and Z. Yang, "A predictive maintenance method for Shearer key arts based on qualitative and quantitative analysis of monitoring data," *IEEE Access*, vol. 7, pp. 108684–108702, Aug. 2019.

36. Z.-H. Liu, X.-D. Meng, H.-L. Wei, L. Chen, B.-L. Lu, Z.-H. Wang, et al., "A regularized LSTM method for predicting remaining useful life of rolling bearings," *IJAC*, vol. 18, no. 4, pp. 581–593, Mar. 2021.

37. R. Zuo, Y. Xiong, J. Wang and E. J. M. Carranza, "Deep learning and its application in geochemical mapping," *Earth Sci. Rev.*, vol. 192, pp. 1-14, May 2019, doi:10.1016/j.earscirev.2019.02.023

38. M. Perno, L. Hvam and A. Haug, "Implementation of digital twins in the process industry: A systematic literature review of enablers and barriers," Compu. Ind., vol. 132, 103558, Jan. 2022, doi: 10.1016/j.compind.2021.103558

39. A. Hazrathosseini and A. Moradi Afrapoli, "The advent of digital twins in surface mining: Its time has finally arrived," *Resour. Policy*, vol. 80, 103155, Jan. 2023, doi: 10.1016/j.resourpol.2022.103155

40. J. Mwanza, P. Mashumba and A. Telukdarie, "A framework for monitoring stability of tailings dams in real-time using Digital Twin simulation and machine learning," *Procedia Comput. Sci.*, vol. 207, pp. 2279–2288, 2024, doi:10.1016/j.procs.2024.02.047

41. P. de Wilde, "Building performance simulation in the brave new world of artificial intelligence and digital twins: A systematic review," *Energy Build.*, vol. 292, 113171, Aug. 2023, doi: 10.1016/j.enbuild.2023.113171

42. M. S. Mahmood and I. R. Shareef, "Applications of Artificial Intelligence for smart conveyor belt monitoring systems: A comprehensive review," *JESA*, vol. 57, no. 4, pp. 1195–1206, Aug. 2024, doi: 10.18280/jesa.570426

43. A. Borucka and P. Guzanek, "Application of the Digital Twin concept in assessing the readiness of production systems," Eur. Res. Stud., vol. 25, no. 2B, pp. 45–58, 2022.

44. M. Barton, R. Budjac, P. Tanuska, I. Sladek and M. Nemeth, "Advancing small and medium-sized enterprise manufacturing: framework for IoT-based data collection in Industry 4.0 concept. Electron., vol. 13, no. 13, 2485, June 2024, doi: 10.3390/electronics13132485

45. P. Schober, C. Boer, and L. A. Schwarte, “Correlation coefficients: Appropriate use and interpretation,” Anesthesia & Analgesia, vol. 126, no. 5, pp. 1763–1768, May 2018

46. S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, “How does batch normalization help optimization?,” in Advances in Neural Information Processing Systems, vol. 31, NeurIPS, 2018.

47. R. J. Hyndman and G. Athanasopoulos, Forecasting: Principles and Practice, 3rd ed. Melbourne: OTexts, 2021.

48. Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang and G. W. Cottrell, "A dual-stage attention-based recurrent neural network for time series prediction," IJCAI-17, Melbourne, Australia, Aug. 19-25, 2017, pp. 2627–2633

49. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, et al., "Scikit-learn: Machine learning in Python," JMLR, vol. 12, pp.2825–2830, Oct. 2011.

50. J. -P. Lai, Y.-L. Lin, H.-C. Lin, C.-Y. Shih, Y.-P. Wang, and P.-F. Pai, “Tree-based machine learning models with Optuna in predicting impedance values for circuit analysis,” Micromachines, vol. 14, no. 2, p. 265, Jan. 2023.

51. G. James, D. Witten, T. Hastie and R. Tibshirani, An Introduction to Statistical Learning: With Applications in R, 2nd ed. New York: Springer, 2021.

52. R. R. Corsini, A. Costa, S. Fichera, and J. M. Framinan, “Digital twin model with machine learning and optimization for resilient production–distribution systems under disruptions,” Computers & Industrial Engineering, vol. 191, 2024, Art. no. 110145.

53. P. Burger, J. C. Vosloo, and J. H. van Laar, “Future mines’ compressed air planning using digital twin simulations,” South African Journal of Industrial Engineering, vol. 35, no. 3, pp. 96–112, Nov. 2024

Appendix A

Apron Feeder Current %

Apron Feeder Drive Speed Indication

Apron Feeder Hydraulic Pack Air oil Cooler Current %

Apron Feeder Hydraulic Pack Aux Motor Current %

Apron Feeder Hydraulic Pack Heater Current %

Apron Feeder Hydraulic Pack Motor Current %

Apron feeder Hydraulic power pack level Indication

Apron feeder Hydraulic power pack Oil Temperature Indication

Apron feeder Hydraulic power pack Pressure Indication

Apron Feeder Lube Pump Current %

Apron Feeder Operating Hours

Apron Feeder Speed Feedback

Jaw Crusher Current %

Jaw Crusher Hydraulic Pack Current %

Jaw Crusher Hydraulic Pack Heater Current %

Jaw Crusher Hydraulic power pack Oil Temperature Indication

Jaw Crusher Hydraulic Pressure Indication

Jaw Crusher Motor DE Bearing Temperature Indication

Jaw Crusher Motor NDE Bearing Temperature Indication

Jaw Crusher Motor U1 Winding Temperature Indication

Jaw Crusher Motor U2 Winding Temperature Indication

Jaw Crusher Motor V1 Winding Temperature Indication

Jaw Crusher Motor V2 Winding Temperature Indication

Jaw Crusher Motor W1 Winding Temperature Indication

Jaw Crusher Motor W2 Winding Temperature Indication

Jaw Crusher Operating Hours

Rom Bin level Indication

Stockpile Feed Conveyor Current %

Stockpile Feed Conveyor Operating Hours

Stockpile Feed Conveyor Speed Feedback

Stockpile Feed conveyor Speed Indication

Stockpile Feed Conveyor Weight Average Indication

Percentage feed material under 80% of target PSD

Percentage product material under 80% of target PSD

Fraction at 106 um

Fraction at 150 um

Fraction at 25 um

Fraction at 75 um

Fraction at 90 um

Appendix B

The project sample codes are shown in the link below via Github:

https://github.com/Uni0201/1110925_Examination.git