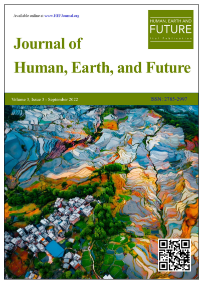
Available online at www.HEFJournal.org

Journal of
Human, Earth, and Future

ISSN: 2785-2997

Vol. 3, No. 3, September, 2022



Predictive Process Control Framework for Online Quality Control in a Hot Rolling Mill

Mpiyakhe R. D. Mabunda [1*], Able Mashamba [1]

[1] *Department of Mechanical and Industrial Engineering, University of Johannesburg, Johannesburg, 2000, South Africa.*

Received 08 July 2022; Revised 16 August 2022; Accepted 22 August 2022; Published 01 September 2022

Abstract

Recent data science advances in statistical classification techniques, and in particular, machine learning techniques, have resulted in more efficient and robust ways of continuously monitoring and managing processes to achieve continuous quality improvement. With increased automation and process data outputs in industrial processes, traditional univariate Statistical Process Control (SPC) is proving impractical in some aspects and slowly loosing overall relevance. The main aim of this study is to develop a predictive process control framework for online quality control. This framework was validated for efficacy when compared to univariate SPC in a selected metal rolling plant based in South Africa. This predictive process control framework employs data science approaches through machine learning techniques and algorithms from the Python programming language. The research methodology is a single case study. An experiment approach was undertaken at hot roughing and hot finishing processes. The results of the study revealed a marginal 17 percent improvement in the predictive process control defect rate compared to the univariate SPC defect rate. The predictive model was based on the Random Forest algorithm and achieved an AUC of 0.84 compared to a 0.81 AUC for the neural network model. Factors found to have a positive impact on the success and sustainability of predictive process control were compliance with predictive model prescriptions, data science knowledge, senior management commitment, and the Extract, Transform, and Load (ETL) approach. These results contribute to the theory of online quality control and can be used as a guide by rolling mill process engineers and quality practitioners.

Keywords: Predictive Process Control; Quality Control; Online Quality Control; Rolling Mill Quality; Machine Learning; Data Science.

1. Introduction

Over recent years, the metal rolling industry has seen a dramatic boom. Demand for rolled metal is at levels last seen before the 2008 recession, with United States of America (US) production estimated to have grown by 2.8 percent between 2018 to 2021 [1]. With increased demand for fabricated metal comes increased customer expectations that add complexity to the metal rolling process [1]. This complexity requires an appropriate response to ensure the industry is ready to meet the challenge. This response will undoubtedly involve continuous quality improvement of process outputs, with online quality control being central to delivering good-quality outcomes [1]. Univariate statistical process control (SPC) is the most common and widely used quality control technique in metal rolling [2]. SPC is defined as a quality control method that uses statistical techniques to monitor and control process variation in quality with the aim of ensuring that processes are efficient and operate effectively [3].

* Corresponding author: robin@cdi.biz

http://dx.doi.org/10. 28991/HEF-2022-03-03-01

For decades, SPC has been credited with assisting organizations to improve their quality, with the benefits of improvement having an overall impact on waste reduction and yield improvement [4]. However, criticisms of SPC as an effective quality control approach have grown steadily, with Bhote [5] arguing that Japan, the primary proponent of SPC, abandoned SPC in the 1970s due to its ineffectiveness. Furthermore, Gunter [6] argues that modern manufacturing and service processes have transcended the relevance and usefulness of SPC, with Woodall [7] emphasizing the need to move away from traditional univariate SPC to newer approaches that include multivariate methods.

According to Singh & Gilbreath [8], multivariate methods are relevant in the digital age of big data because they provide alternatives that incorporate the interdependence and volume of process characteristics that were previously restricted. Multivariate analysis has recently migrated to the machine learning and artificial intelligence spaces, where it can be used to predict future performance. Multivariate process control using machine learning techniques is a new field of study with very limited framework references available. There is a knowledge gap in academia and industry about the structure that best delivers predictive process control using machine learning or artificial intelligence, but increasing data availability and computing power are allowing these structures to be tested and formulated [9]. With SPC becoming less relevant and data science (via machine learning and artificial intelligence) gaining relevance [10], the purpose of this study was to develop a predictive process control framework based on a multivariate approach for online quality control that will be used as a viable alternative to an increasingly irrelevant univariate SPC. The effectiveness of the predictive process control framework was evaluated as part of this study through a single case study experiment at a hot rolling plant known as Company X. The results of this study provide an academic and metal rolling industry reference that is currently lacking. The key questions associated with the study were:

- How can a machine learning or artificial intelligence model be deployed to enable predictive process control framework for online quality control in a rolling mill?
- What is the effect of predictive process control on the defect rate of a rolling mill?
- Which aspects of predictive process control are applicable at different levels of a rolling mill?
- Which factors greatly influence the outcome of the predictive process control deployment in a rolling mill?

The following hypotheses were tested (where the defect rates of predictive process control and univariate SPC were tested):

H_0: There is no significant difference in the defect rate of traditional univariate SPC and predictive process control when defect rates of the two approaches are compared.

H_a: There is significant difference in the defect rate of traditional univariate SPC and predictive process control when defect rates of the two approaches are compared.

2. Literature Review

2.1. Statistical Process Control

Based on traditional quality management literature, quality control is a process that a company uses to ensure that product quality is maintained or improved [11]. Online quality control, on the other hand, focuses on the monitoring and surveillance of a process in order to detect anomalies and analyze and eliminate the causes of variation.

SPC is the most widely used method for online quality control in industry [12]. It forms the foundational construct for predictive process control and this study. Traditional quality control methodologies tend to focus primarily on product quality control which emphasizes defect detection through inspection while SPC is process oriented [11]. Understanding and controlling variation is central to understanding SPC. Variation can be understood as the spread between parameter numbers in a data set and the most common measure of variation is standard deviation [13]. Shewhart recognized that variation is unavoidable in processes and that every process has inherent variation due to what he identified as chance causes (also known as common causes) and assignable causes (also known as special causes) [14].

There are numerous well documented successes, controversies and challenges associated with SPC. Others have even questioned the relevance of SPC in its current form. Gunter [6] argues that modern manufacturing and service processes have transcended the relevance and usefulness of SPC. Gunter [6] further argues that, Shewhart control charts have lost their relevance in the current environment and this will only get worse with the fourth industrial revolution (4.0.IR).

Banks [15] is also very critical of continued SPC research by noting that the time has passed for university research on outdated concepts like SPC because continued research perpetuates the reputation of being out of touch with relevant concepts. Woodall [7] offers a more balanced view in that he notes, SPC's primary objective of understanding, modelling and reducing variability over time remains relevant and important. He further emphasizes the need to expedite the transition from classical methods to some of the newer approaches when appropriate. Montgomery & Woodall [10] highlights how recently developed methodologies that include multivariate methods, variance components, change-point techniques and regression-based methods can greatly increase the usefulness of SPC in some common situations.

2.2. Predictive Process Control

Predictive process control is a term that does not have a universal definition. Micronite [16] defines predictive process control as tool-centric predictions of process patterns and low-risk assurance of specification compliance, whereas Funk [17] defines predictive process control as a method for predicting the performance of a process outcome through separate measurements of selected fundamental properties of each process input parameter. In this study, predictive process control framework should be understood as the structured approach to deploying multivariate machine learning or artificial intelligence models of a process for online quality control. The proposed predictive process control framework is separated into people and data science aspects. The people aspects of the proposed predictive process control focus on: ***Organizational Culture; Leadership; Base Capabilities.***

The data science aspect of the proposed predictive process control focuses on predictive analytics using machine learning and/or artificial intelligence techniques. Data science is an approach or an emerging field that has not been completely defined. Skiena [18] defines data science as a field or approach that lies at the intersection of computer science, statistics and substantive application domains. Machine learning and high-performance computing technologies are categories that sit in the computer science pillar of data science. Applications domain refers to the business, industry or technical know-how to contextualize the data and to develop evaluation standards that assess when problems are adequately addressed. Data science application in manufacturing has several applications. These include predictive maintenance, predictive quality, sales forecasting, KPI forecasting and more.

Despite the fact that data science has seen steady growth in the manufacturing industry over the last ten years, Artificial Intelligence (AI) is not new to the industry. AI software tools such as artificial neural networks, fuzzy logic, genetic algorithms, and support vector machines have been widely used in various manufacturing industries in a piecemeal manner over the last three decades [19]. Figure 1 depicts Alshraideh et al. [20] framework to predictive control of quality, which is limited to a decision tree approach to predicting quality using random forest techniques.

The framework is general in the sense that it can be applied to any production process where profile data are available, and it can easily integrate process specific features based on process operator experience or commonly used transformations of observed signals as additional features. Alshraideh et al. [20] framework summarizes theories for predictive control coming from the following domains:

- Data Engineering;
- Computer Software (machine learning and artificial intelligence);
- Computer Hardware;
- Statistics;
- Maths.

The last ten years have seen a dramatic increase in the awareness of data science, the field has grown considerably in manufacturing. According to a *Mordor Intelligence* [21] report, big data analytics in manufacturing industry market was valued at USD 904.65 million in 2019 and is expected to reach USD 4.55 billion by 2025 at a compound annual growth rate of 30.9 percent over forecast period of 2020 to 2025. In another report, it was stated that the global smart manufacturing market size is estimated to reach USD 395.24 billion by 2025, registering a compound annual growth rate of 10.7% [22]. According to the IFS digital change survey to assess the maturity of digital transformation in a range of sectors, such as manufacturing, oil and gas, aviation, construction and contracting, 46 percent of the companies in all industries are looking to invest in big data and analytics [23]. Even with its growth, data science has challenges in industry, the biggest one being lack of subject matter expertise [24].

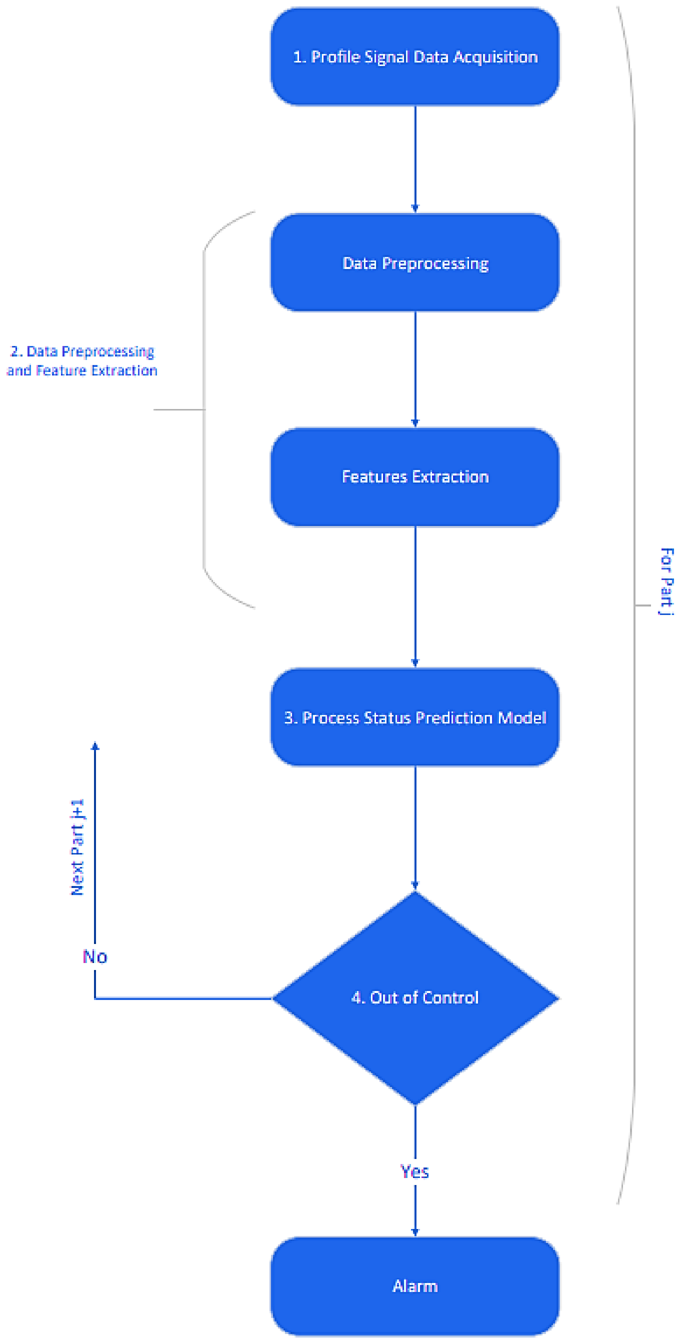


Figure 1. Framework for predictive control model (Reprinted and adapted from Alshraideh et al. [20])

3. Proposed Framework

Univariate SPC has historically been the preferred quality control approach of choice by manufacturers, but criticisms of its effectiveness and future relevance have grown steadily over the last 30 years [5]. Recent advances in statistical classification techniques, and in particular, machine learning techniques, have resulted in more efficient and robust ways of managing online quality control [19]. With very little industry and academia reference framework for effective deployment of machine learning models to predict quality, a predictive process control framework was proposed as part of this study. The proposed predictive process control framework and its elements are discussed in this section.

3.1. Proposed Predictive Process Control Framework

Based on previous research on the impact of variables (both dependent and independent) of variation, SPC, data science, leadership, organizational culture, and base capabilities, it is argued that they have a positive influence on the successful deployment of predictive process control. Figure 2 depicts a proposed predictive process control framework. This framework assumes a discreet manufacturing process with sensors or tags that measure process parameters during production. The proposed framework has ten aspects (with some aspects having multiple steps). The order in which these steps were executed to deliver optimal outcome is presented in section 5.4 of the results section.

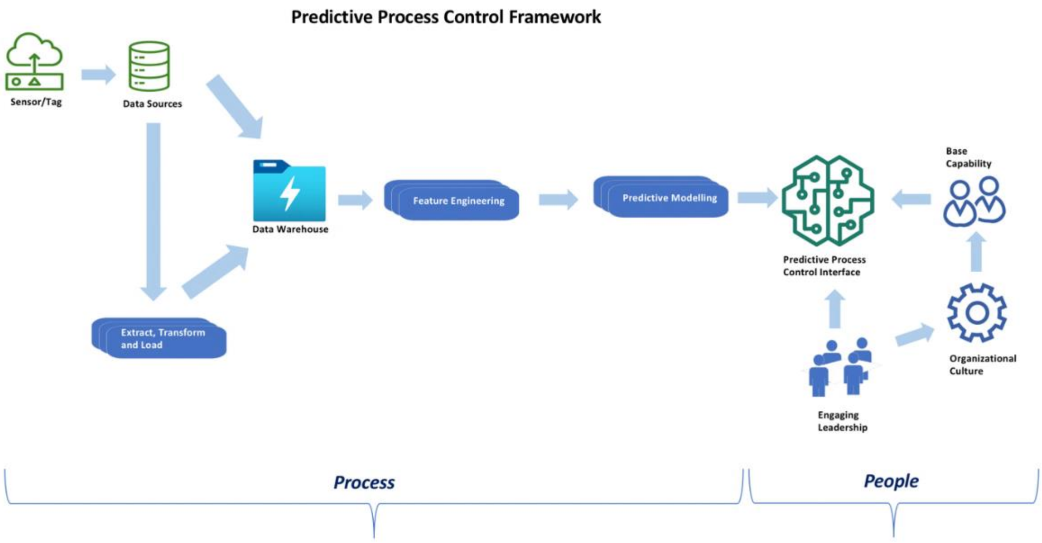


Figure 2. Proposed predictive process control framework for rolling mill process

The ten aspects of the proposed framework are described as follows:

- ***Sensor/Tag Data*** – sensor or tag data refers to raw outputs from an equipment or hardware. These outputs are responses to some type of input from the physical environment.
- ***Data Source*** – a data source refers to the initial location where the structured or unstructured sensor or tag data originates from. Data is typically stored in the form of a data table, a data object, or another format.
- ***Data Engineering*** – according to Black [25], data engineering is the act of designing and constructing pipelines that turn data into a format that is highly useable by the time it reaches data scientists or end users. The first step in any data science or machine learning activity is the requirement that a data signal or data point be available before being fed into a data source that collects big data. The data source converts the signal or data point from sensors or tags into a format that can be used.
- ***Extract, Transform and Load (ETL)*** – is the process of extracting data from source databases, transforming it into a uniform format for specific business objectives, and then putting the reformatted data into storage [26]. There are three steps in ETL. The first step is extracting data from data sources. The second step involves transforming the data into a usable format. The third and last ETL step involves loading the usable data into a data warehouse.
- ***Feature Engineering*** – is the process of leveraging domain expertise to choose and convert the most important variables from raw data [27].
- ***Predictive Modelling*** – is a statistical technique that employs machine learning and data mining to predict and forecast likely future outcomes using current and historical data.
- ***Engaging Leadership*** – this is management at all levels of an organization that can articulate its predictive process control objectives, drive active management routines to support delivery of these predictive process control objectives, ensure resources are available to execute aspects of predictive process control, and ensures these resources are challenged and supported to deliver on the objectives. There are multiple activities involved in this step that include, setting up the vision for predictive process control, setting up, leading routines that drive full adoption of predictive process control and ensuring continuous development of capabilities.
- ***Organizational Culture*** – this is how things are done in a particular environment.
- ***Base Capability*** – refers to the ability of an organization to execute certain basic functions well.
- ***Predictive Process Control Interface*** – this refers to a computer monitor that shows operators the actual performance of the process as it runs, allowing them to maintain good performance or rectify process deviations as they occur.

3.2. Proposed Maturity Assessment Model

To help a rolling mill know how well they're implementing the proposed predictive process control framework, a proposed maturity assessment tool has been developed by the researcher and is presented in Figure 3. The proposed maturity model consists of eight areas of focus. These eight areas are assessed using a score of one to five, with one being the worst and five being best.

Maturity Assessment					
Area of Focus	1 Clean slate	2 Brown slate	3 Functional	4 Connected	5 Financialized
Leadership Routines	- Leadership roles are not clarified - Responsibilities and Routines are non-existent	- Leadership roles are loosely defined - Responsibilities are some what established with adhoc check-in sessions	- Leadership roles are clarified and documented - Leadership responsibilities and set routines are defined	- Link with operational performance is ceated - Contribution to operational performance is measured	- Quanitified financial impact reported annually - There is link with Cost of Quality measure
Vision	- No link with operational excellence strategy and framework	- Some link with operational excellence strategy and framework exists	- Link with operational excellence (OE) strategy is established and documented on the OE framework	- Performance outcomes form part of operational performance routines	- Financial outcomes form part of operational performance routines
Human Resource(s)	- Zero skill set identified internally and externally to drive and support predictive process control - Responsibilities and routines are not documented in process team (s) role profiles - Does not form part of process technician/engineer responsibilities and routines	- Inetrnal candidate(s) is identified and deployed to set-up system but supported by external experts with appropriate skills set - Responsibilities and routines are clearly defined but not documented in process team (s) role profiles - Adhoc activities form part of process technician/engineer responsibilities and routines	- More internal resources are identified and trained centrally and operationally to support predictive process control activities - Responsibilities and routines are clearly defined and documented in process team (s) role profiles - Structured routines are developed to support predictive process control objectives	- Data science consideations are taken into account when employing new recruits - Data science aspects form part of personal development plans	- Development spent and financial improvement are managed so to ensure there is return on investment
Capability	- Knowledge around data science is non-existant	- Self-taught resource(s) experiment without buiness case	- A clear capability plan is deployed and measured by management	- A scorecard for predictive process control exists based on existing capabilities, with performance trends being measured by managers across all levels - There is alignment between improvement in scorecard and operational performance	- There is alignment between improvement in scorecard and improved financial performance
Financial Resource(s)	- Predictive process control is not budgeted for	- Funds for uncoordinated activities are made available	- A budget to fund identified proof of concept(s) is established and forms part of the business budget	- A budget to fund structured activities exist centrally and operationally - Capex considerations are included in Capex budgets	- Budget extends to research and innovation
Business Model	- Traditional univariate model with zero multivariate view	- Uncoordinated univariate and multivariate model	- Coordinated univariate and multivariate model exist though centred around internal organizational capability	- Coordinated univariate and multivariate models exist and centred around customer specification limits	- Models link with downstream customers and upstream suppliers
Data Management	- Disparate and disconnected data sources - Lack of standardization	- There's some established data architecture - Descriptive data exists	- Data architecture in place - Automated collection of data - Standard for data strctures exists	- Data warehouses support multivariate modelling - Data management routines support operational excellence frameworks	- Data management routines support cost optimization initiatives
Technology	- Undefined application portfolio - Undefined architectural governance - Multiple point solutions in use	- Limited application portfolio - Limited architectural governance - Diagnostic insight established	- Managed and standardized. Application portfolio - Architectural governance fully developed - Use of automation and workflow - Predictive insight established	- Enterprise and Business Architecture are vertically & horizontally aligned - Platform based technology - Predictive insights driving quantifiable operational excellence outcomes	- Predictive insight outcomes support financial contribution objectives

Figure 3. Proposed maturity assessment model

4. Research Methodology

The research process applied in this study follows Saunders et al. [28] research onion approach. The research onion used in this study comprises of six layers or stages as shown in Figure 4.

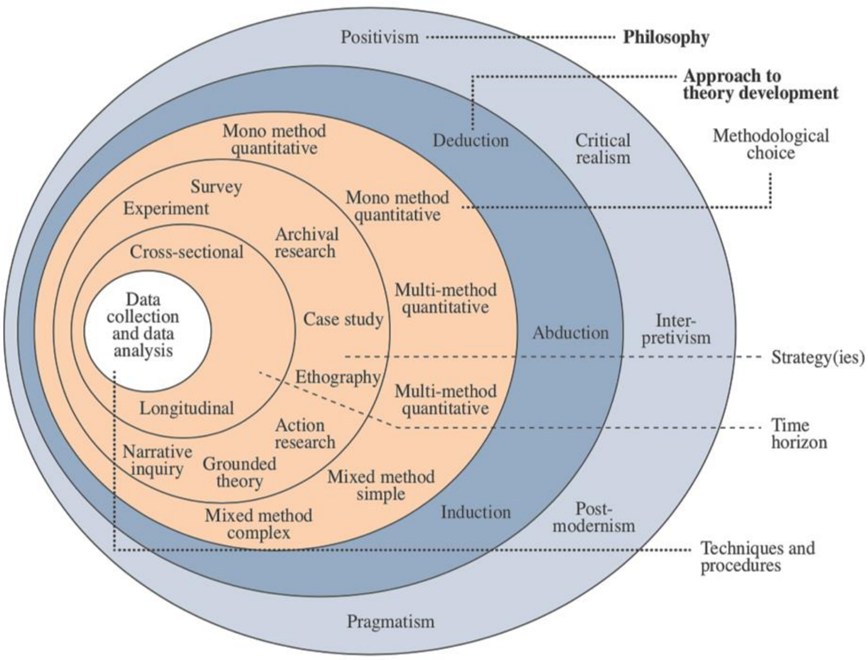


Figure 4. The research onion (Saunders et al. [28])

This study uses a positivist philosophical worldview to answer the research questions. The choice for a positivity paradigm is based on the scientific nature of the methods applied. Because this study is guided by theory and tests for hypotheses, the applied theory development is deductive. Furthermore, this research study used case study research design. The choice of a case-study method was based on the limited information and studies about predictive process control, this necessitating the need for an in-depth study in which careful consideration needed to be given to the development of predictive process control over time. The theory testing case study research method was used, and the research method for validating the proposed predictive process control framework was an experiment. The choice for an experiment is guided by the principle of wanting to manipulate one variable in order to observe a change in another variable, thereby being able to evaluate the relationship between variables.

Validation of the proposed framework was based on single-case study research, focusing on real-life events that demonstrated a single source of evidence at Company X. The theory testing procedure used for this study was divided into two phases: the pilot study (phase 1) and the main experiment (phase 2). Because activities do not always work out as planned, a pilot study was used to test the research methodology, tools for data collection and assumptions to ensure the complete study achieves its objectives [29]. A pilot study into the feasibility of the proposed predictive process control framework with sample data from the Hot Roughing Mill (HRM) and Hot Finishing Mill (HFM) preceded the establishment of the main experiment procedure and approach to validation. 396 tags or sensors for selected HRM and HFM data sets were extracted from the study site and tested using a predicting model that supports the proposed predictive process control framework. A total of 58 metal ingots or metal identifications (ids) were analyzed in the pilot study. The sample size was established using power analysis.

For the purposes of this study, the population is defined as an organization (referred to as Company X). Company X denotes a single company with a single hot rolling plant. Company X is a company based in South Africa. Although the population of one organization may appear to be small, the test and control groups were large enough to test the hypotheses without jeopardizing the experiment's credibility. Quantitative sample design was used in this study, and the specific method used in order to establish sample size is called power analysis. Power analysis methods refer to a group of statistical methods used to determine the appropriate sample size for an experiment. These methods produce a power statistic, which quantifies the likelihood that the planned experiment will successfully detect a meaningful difference between the test and control populations, if one exists [30]. Random sampling is a key assumption in power analysis. The sample on which power analysis is performed is drawn through the random sampling process. The test and control groups in this study's experiment are made of the AA5182 alloy. Power analysis for the AA5182 alloy was performed using Python software, with multiple confidence intervals indicated by alpha, as shown in Table 1. Table 2 provides data that support the sample size calculation.

Table 1. Power analysis table for AA5182 alloy

Statistical Power	Alpha	Experiment sample size (Number of coils)
0.8	0.01	80
0.8	0.05	27
0.8	0.1	23
0.8	0.025	35

Table 2. Input data table for power analysis of AA5182 alloy

	Number of coils evaluated in data modelling using historical data falling within defect free zone	Number of coils evaluated in data modelling using historical data falling outside defect free zone	Total number of coils evaluated in data modelling using historical data
Count	267	2840	3107
Mean	0.288	0.54	0.53
Standard Deviation	0.45	0.5	0.5

Using the above AA5182 alloy power analysis table, to obtain a 95 percent confidence interval (alpha = 0.05) on the evaluation of test results, the number of sample metal ingots or metal ids required for the main experiment of the study to reach a statistically significant conclusion is 27. The initial sample size of 27 ingots or coils was increased by a factor of three to get to 81 ingots or metal ids as the final sample size to account for non-related process drifts. Feature selection was a key driver in the power analysis. The power analysis was carried out based on the significance of each feature in explaining defects in the hot rolling process. The feature importance rating is attributed to that feature, relative to all other features. Below list has examples of features that were included in the model:

- HRM_TGT_EXIT_PROFILE1
- HFM_TGT_GAUGE_TOLPOS1
- HFM_TGT_TRIM_WIDTH1
- BRUSH_ID_RIGHT_BOT
- TEMP_EXIT_TRGT
- DIAM_BR_BOT
- HRM_TGT_EXIT_TEMP1

Prior to data analysis, data preparation was completed. This process entailed verifying the completeness and accuracy of collected data, as well as converting the data into a format that allows for analysis and interpretation. The Hot Mill process and control engineers validated the test metal ingots or coils that were hot rolled using data science approaches for the purposes of this study. The quantitative data was analysed using Minitab (version 21.1.0) statistical software package. Hypotheses testing for this study was performed using the two-samples proportion test. Field [31] argues that prior to testing hypotheses, a parametric or non-parametric test should be performed. For data sets with a normal distribution, a parametric test is performed. For data sets that do not have a normal distribution, a nonparametric test is performed. The data set for this study is binary and thus not normally distributed. The binary nature of the test also guided the choice of two-samples proportion test.

5. Results

The results were divided into the five sections. Section 5.1 reported on the machine learning or artificial intelligence (AI) deployment results, Section 5.2 reported on defect rate results, Section 5.3 reported results on applicable aspects at different management levels, Section 5.4 reported on results of factors that greatly influence the outcome of predictive process control and Section 5.5 presented hypotheses testing results.

5.1. Machine Learning or AI Deployment Result

Data for the experiment's predictive model was extracted from seven processes shown in Figure 5. These processes are metal melting, metal holding, metal casting, metal scalping, metal reheat (pusher furnace), metal hot roughing (HRM) and metal hot finishing (HFM). The structure of the extracted data was such that it needed to support time-series modelling. Time-series modelling is a statistical technique that deals with data that is in a series of a particular time period or intervals. The machine learning process control prediction model was developed to predict HRM and HFM processes. The other five processes provided input process data that assisted with prediction of HRM and HFM processes.

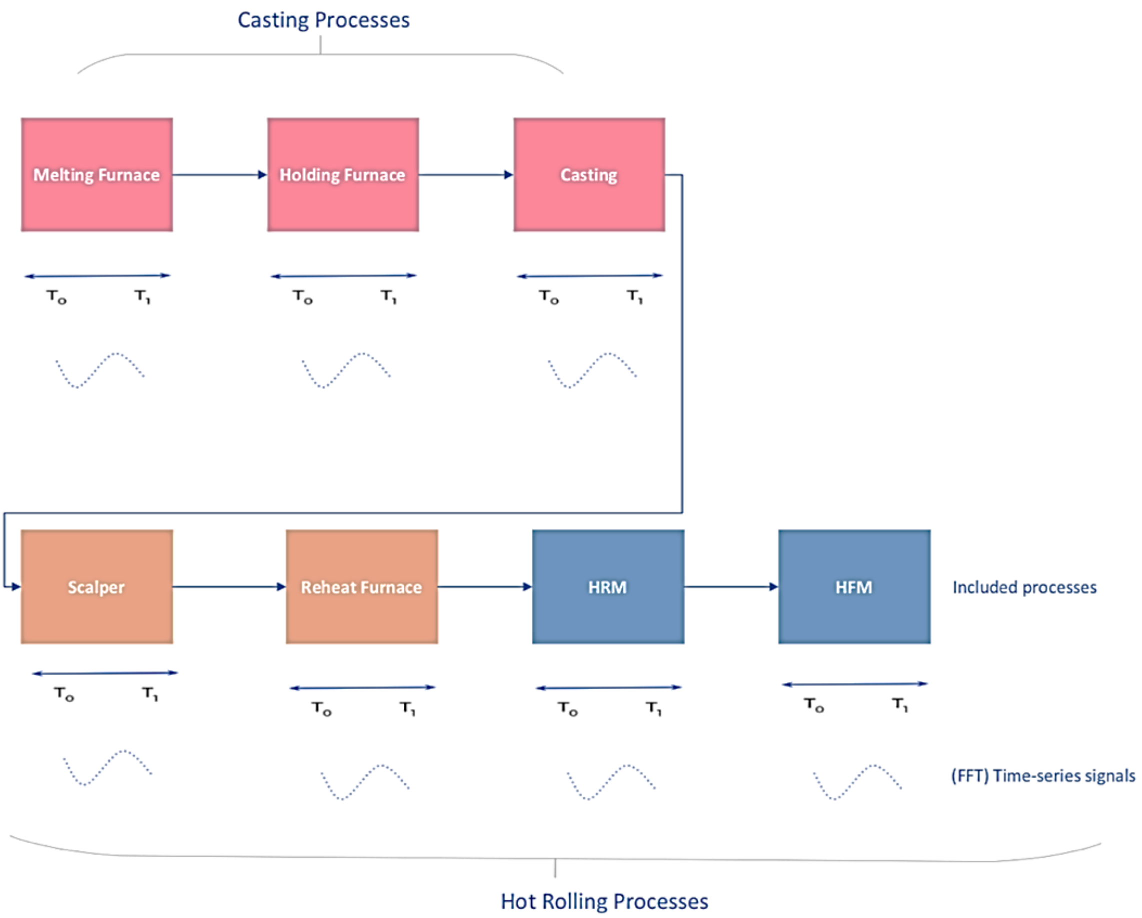


Figure 5. Data provision processes

Figure 6 depicts the predictive model experiment timeline, including planned and actual durations. The only stage that deviated from the plan was ETL. Across all seven processes, the total number of variables extracted as part of ETL for feature engineering was 3100.

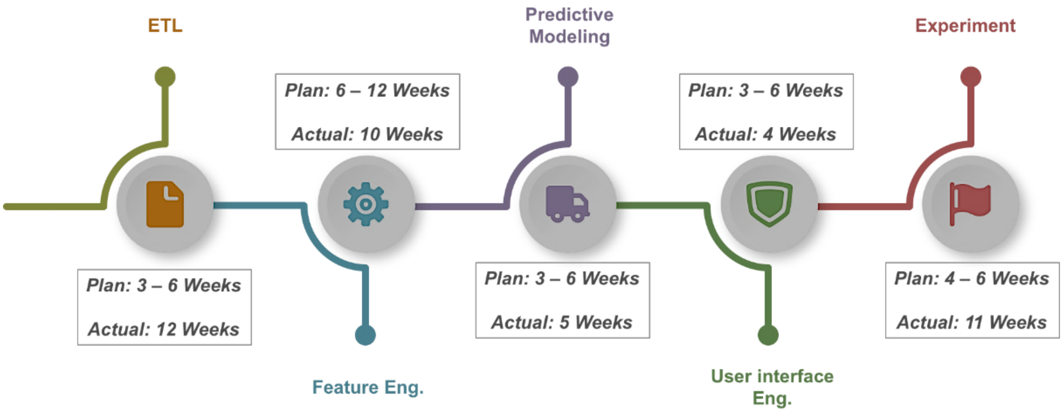


Figure 6. Experiment Timelines

Figure 7 depicts the ETL architecture. The input data span was two years, beginning in 2020. The data was processed in daily batches. Before being loaded into the data warehouse, the extracted data was decoded and dumped to S3 as daily parquet files. There were over 1000 files and 1TB of decoded data, each with approximately 550 unique days.

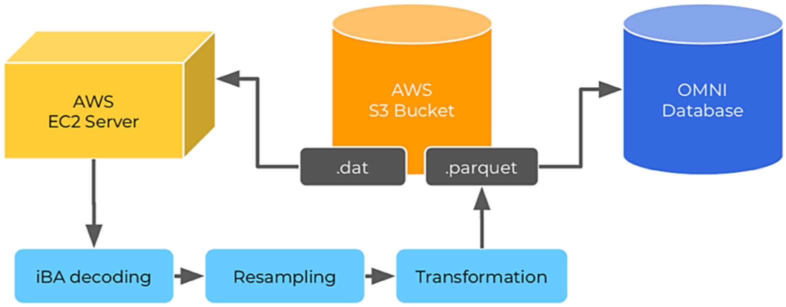


Figure 7. ETL architecture (Bennie Lombard [30])

The principal component analysis (PCA) technique was used to determine features for predictive modelling of this study's experiment. Python language was used for modelling. Variational auto encoders were also used in the feature engineering architecture (VAEs). VAEs generate a low-dimensional representation of a high-dimensional input data set. The input data classes are divided into clusters. The unified process parameters were reduced to two dimensions, providing representations of the state of the process for each metal ingot identification code (metal id) during the relevant production window. Because of the low dimensionality, it is possible to see the state of the relevant processes during the production of each metal id.

Derived features were created using Fast Fourier Transform (FFT) to capture the necessary Fourier components for the model to approximate the input signals. For feature extraction, FFT features of each signal were extracted. This allowed for finding a compressed representation of the time series signals. For the unified view, six FFT components with the largest magnitudes were used. For both HRM and HFM, the upper and lower bound of the body and tail phase of the signals were extracted. Figure 8 depicts an FFT extracted for rolling speed at HFM. To drive the predictive model, 395 features from both HRM and HFM were identified.

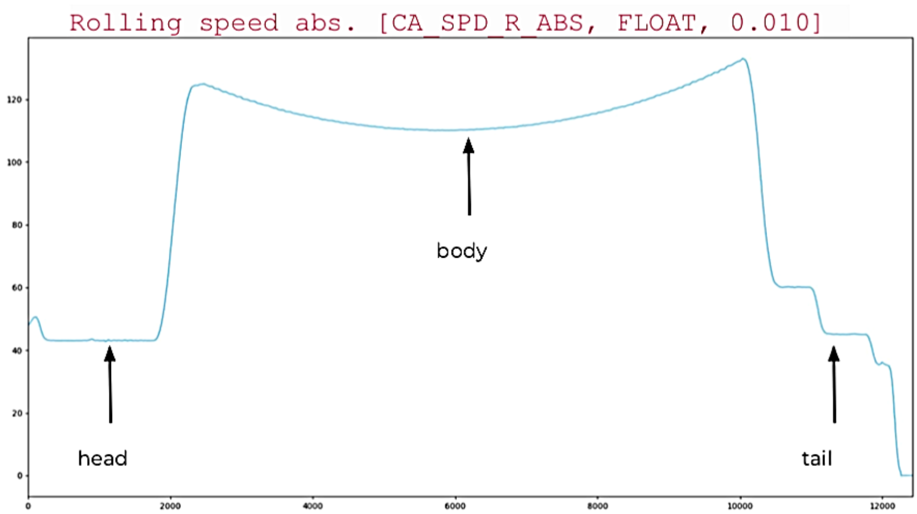


Figure 8. Rolling Speed FFT

The derived features were classified as controllable or non-controllable. Controllable features are those that the rolling mill operator can directly control. Non-controllable features are those directly controlled by the mill control system without involving the operator. Following the establishment of the initial feature list, further investigation was carried out using the random forest algorithm to refine the final list features for accurately predicting quality on both HRM and HFM. Figure 9 depicts the random forest model results for accurate quality prediction. A supervised learning approach was used to apply the random forest model. The model represents the process control problem as a binary class classification problem where the class of interest is the process status being in-control or out-of-control. For accuracy indicator, area under the curve (AUC) is used to evaluate various classifiers in the experiment. The result associated with the model's ability to accurately predict quality is an AUC of 0.84. The random forest predictive accuracy is slightly better compared to an artificial neural network model that was also used in comparison to random forest. The artificial neural network had an AUC of 0.81. Table 3 displays the top ten features, with ranking one having the greatest influence on prediction and ranking ten having the least influence.

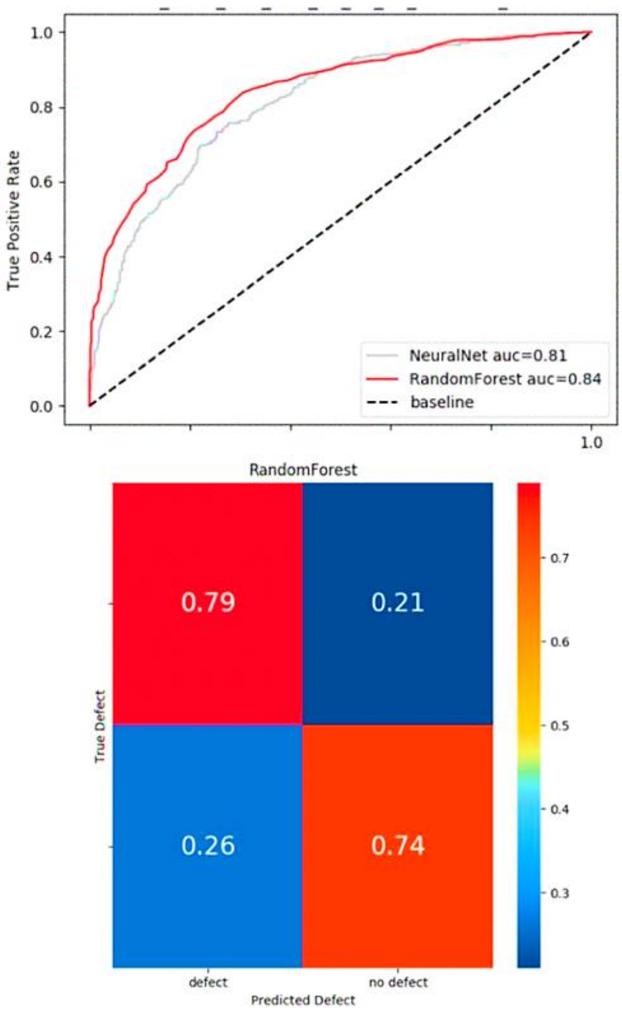


Figure 9. Random forest model for accurate quality prediction

Table 3. HRM and HFM top ten features

Ranking	HRM/HFM Feature name and pass number
1.	HRM HMB Clean
2.	HFM Brush Press Exit OS Pass 3
3.	HRM SPD Strip Exit Pass 2
4.	HFM Brush Press Exit DS Pass 3
5.	HFM Ten Exit Min HD Pass 2
6.	HFM Tilt Force STD HD Pass 2
7.	HFM Soaps
8.	HFM Brush Press Entry DS Pass 2
9.	HRM Acid Split Viscosity
10.	HRM Particle Size (percent<2 micron)

Using the derived final list of features, a predictive machine learning model was created on Python using VAEs and the Keras algorithm to predict quality and determine prescriptions to correct process deviations. Time-series techniques were used in the modelling approach. The time-series data was stitched together into a unified view, making rolling pass performance easier to interpret. The experiment and modelling was limited to AA5182 alloy, 1620 mm ingot width, 2.3 mm gauge and only looking at surface defects of broken surface. The minimum metal ids required to effectively train the model was 1000. The actual number of metal ids used to train the model was 3107. Figure 10 presents the predictive model results in the latent space. The model results shows areas of good quality and uplifted quality.

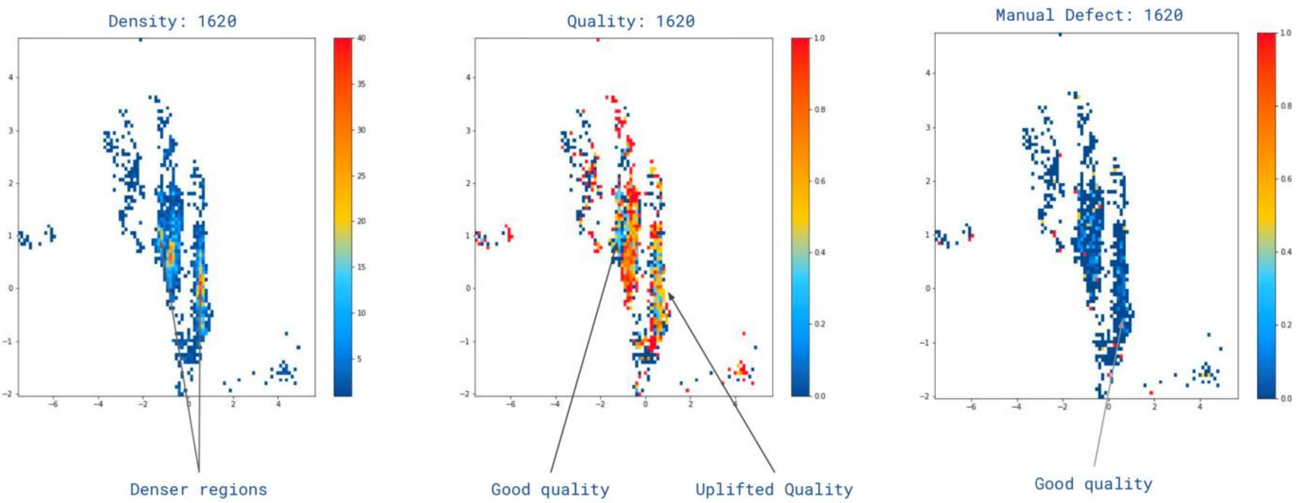


Figure 10. Model Result: 1620mm ingot width areas of good quality and uplifted quality

After establishing the area of uplifted quality, it was important to establish a cluster or region of process stability that gives the best quality outcome. This entailed searching for a region in the model results' latent space with a low defect rate and relatively high density. This implies that there is a common set of process parameters for which good quality is achieved. By targeting the process parameters corresponding to this area, quality defect rate can be reduced to be lower than the global average. Process prescriptions were established for this region. Figure 11 shows the cluster of process stability that gives the best quality for 1620 mm width and 2.3 mm final rolled gauge. The cluster or region is depicted by a black polygon. The cluster was observed to show temporal stability with the region being visited consistently during production runs of 1620 mm width metal ids.

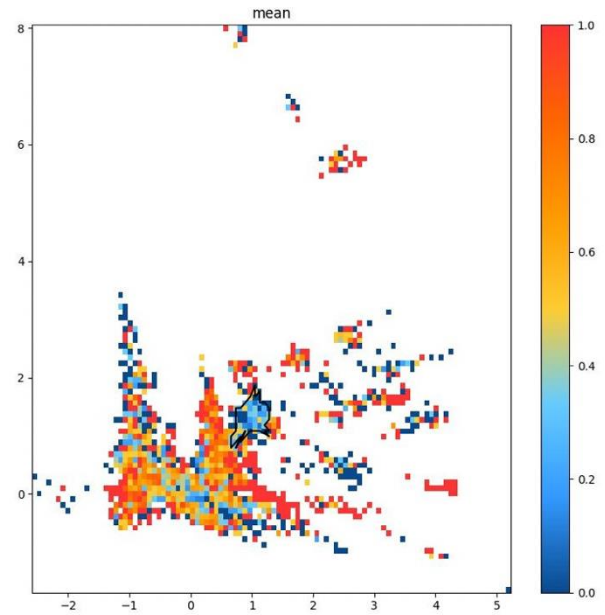


Figure 11. Model Result: 1620 mm ingot width cluster/region of process stability and uplifted quality

The predictive model prescriptions generated from the model were deployed for the 126 metal ids that formed the experiment. Eleven batches of 126 metal ids were rolled. The initial model prescriptions were generated from the predictive model and were pre-loaded for metal ids of the first batch on the control system prior hot rolling of the batch commenced. A user interface solution was developed and designed for follow-up batches to provide real-time prescription ranking to HRM and HFM operators and process engineers. Figure 12 depicts the structure of the real-time user interface prescription report, while Figure 13-a and 13-b depicts the actual report. The user interface undertook the following steps everyone hour:

- Pull all new raw values from data sources.
- Upload all values into the data warehouse.
- Build new unified view entries using the new raw values.
- Generate ranking report on the most recent row of the unified view.
- Serve the report to the HRM and HFM operators and process engineers.

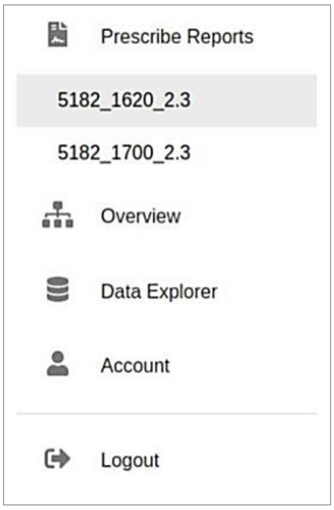
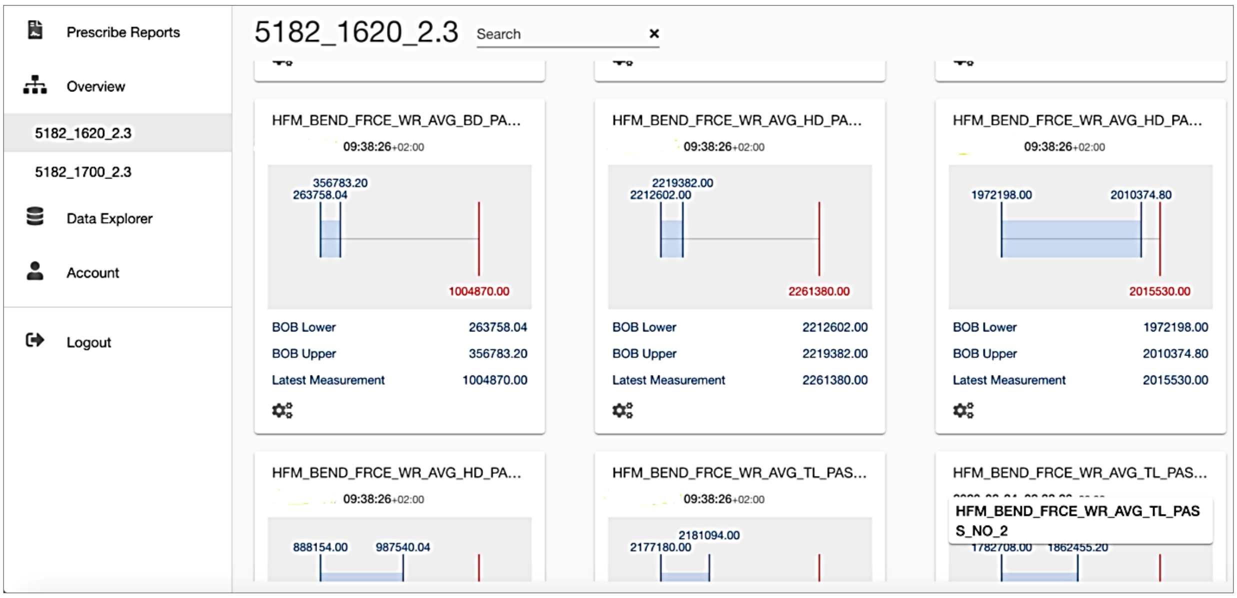


Figure 12. Front-end: User interface prescription report structure

(a)



(b)

Figure 13. a) Front-end: User interface prescription report, b) Front-end: User interface prescription report

5.2. Defect Rate Results

Using power analysis to determine the appropriate sample size for the experiment, the sample size was determined to be 81 metal ids and 126 metals ids were randomly experimented on. Predictive model defect rate result for surface defects of 1620 mm wide ingot is 3.17%. The aforementioned 3.17% was realized against a 45 percent compliance rate to predictive model prescriptions. Compliance is the measure of proportion of controllable parameters that are within their prescribed limits. The combined (historical and during experiment) surface defect rate for univariate SPC is 3.74%. The historical defect rate data was collected over an 18-month period. The aforementioned equates to a 17 percent difference in defect rate between predictive process control and univariate SPC. Figure 14 depicts a box plot with a comparison of surface defect rates.

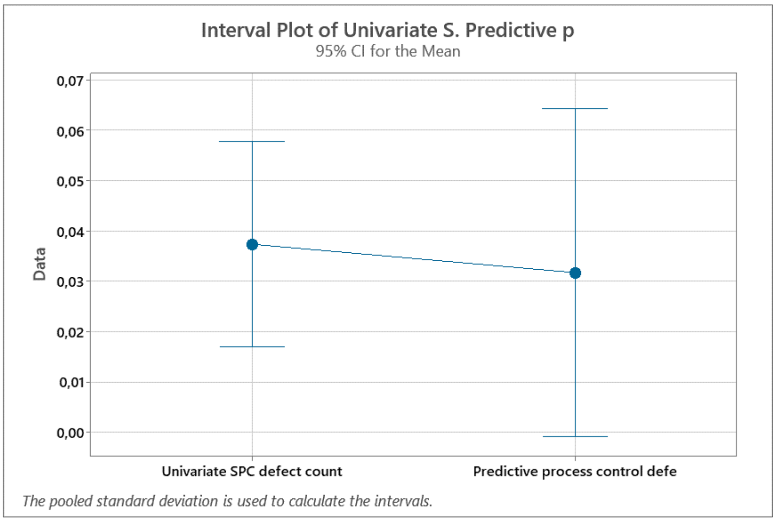


Figure 14. Surface defect rate comparison

5.3. Applicable Aspects at Different Management Levels

Prior to beginning the predictive process control experiment, roles and responsibilities were established. Each aspect of the predictive process control framework was assigned to a role. At senior management, the results show that development of the vision for predictive process control, accountability structures, continuous capability review and built, and driving accountability for predictive process control across all levels were key requirements and drivers for effective deployment of predictive process control. This finding is consistent with Antony's [12] MEST framework theories on management being essential to successful online quality control for SPC. Furthermore, the findings reveal that the technical aspects of the predictive process control framework reside at middle- and first-line management, with shop floor accountable for execution.

The aforementioned results for all rolling mill levels identify the two dimensions of the proposed predictive process control framework, namely process and people. This further validates this study’s contribution in identifying organizational development and leadership gaps.

5.4. Factors that Greatly Influence the Outcome of Predictive Process Control

In the experiment of this study, the steps to deploying predictive process control framework followed the order below:

- Run predictive process control awareness workshop with senior management;
- Develop a vision for predictive process control at senior management;
- Develop predictive process control project charter and get senior management signoff. Charter includes objective, metrics of success, risks and opportunities;
- Develop change management plan;
- Set-up effective structure to enable predictive process control;
- Capability determination;
- Identify resource(s) to close capability gap;
- Run training and coaching sessions to close capability gap;
- Data engineering (including ETL);

- Identify and set-up data warehouse;
- Feature engineering and signoff;
- Predictive modelling;
- Development of user-interface;
- Run predictive process control.

All but one of the aforementioned steps of the deployment process (develop change management plan) required data science training to effectively execute predictive process control. Because of the training requirements, a capability assessment activity was required prior to step 1 to ensure key concepts that enable predictive process control were understood. The single most important factor influencing the outcome was predictive process control base capability. Another factor that influenced the experiment was senior and middle management buy-in to ensure effective structures were in place and that progress was tracked. The buy-in factor related to management physically participating in project activities to enable and guide the intervention.

To understand how predictive process control capability requirements can impact its success, a gap assessment was conducted prior conducting the experiment. The approach towards the gap assessment applied skills requirements comparison. The skill requirements are divided into two categories: process control and rolling knowledge. The following steps were taken in the gap assessment process:

- Standardise Process Control knowledge requirements for univariate SPC and predictive process control. The standard is based on literature review from chapter 2 of this study.
- Standardise rolling knowledge requirements based on aggregated hot rolling skills matrix from Company X.
- Compare Company X hot rolling job profile competency requirements against univariate SPC and predictive process control standard knowledge requirements.

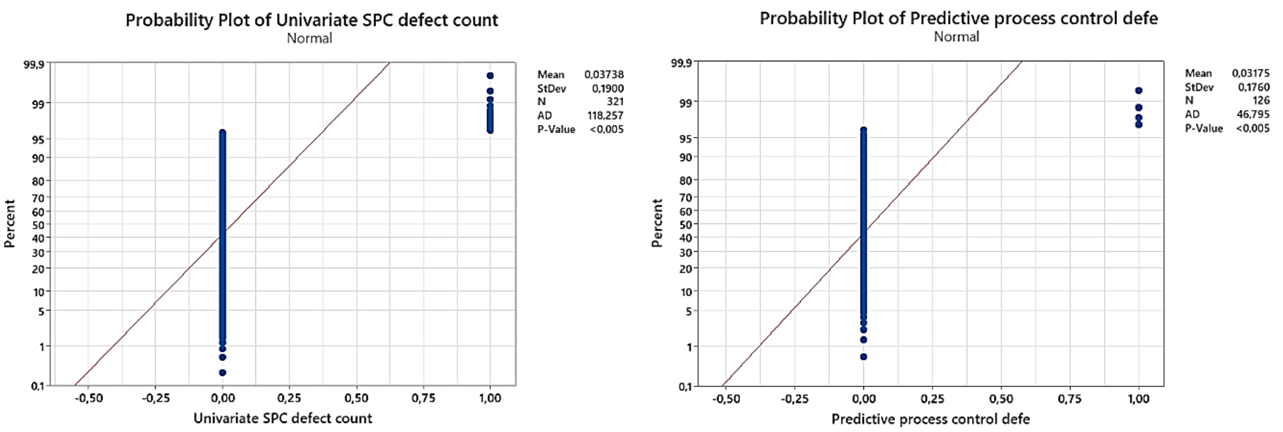
The aggregated comparison results of the baseline capability assessment show a 0 percent gap for univariate SPC and a 25.71 percent gap for predictive process control. This result supports Amruthnath's [24] contention that the most difficult challenge in deploying data science solutions in manufacturing applications is expertise in these techniques and their application in a real-world setting.

5.5. Hypotheses Testing

H*$_0$*: There is no significant difference in the defect rate of traditional univariate SPC and predictive process control when defect rates of the two approaches are compared.

H*$_a$*: There is significant difference in the defect rate of traditional univariate SPC and predictive process control when defect rates of the two approaches are compared.

Although a normality test is not required for binary data, one was performed as a confirmatory step for both univariate SPC and Predictive Process Control. The probability plot for univariate SPC is shown in Figure 15-a. The univariate SPC p-value is less than 0.005, confirming that the data is not normally distributed. The probability plot for predictive process control is shown in Figure 15-b. Similar to univariate SPC, the predictive process control p-value is less than 0.005, confirming that the data is not normally distributed.



(a) (b)

Figure 15. a) Univariate SPC probability Plot, b) Predictive process control Plot

The two-samples proportion test results are outlined in Table 4. The normal approximation may be inaccurate for small samples.

Table 4. Two-samples proportion test results for hypotheses

Method			
Event: 1			
p_1: proportion where Univariate SPC defect count = 1			
p_2: proportion where Predictive process control defect count = 1			
Difference: $p_1 - p_2$			
Descriptive Statistics			
Sample	**N**	**Event**	**Sample p**
Univariate SPC defect count	321	12	0.037383
Predictive process control defect	126	4	0.031746
Estimation for Difference			
Difference		**95 percent CI for Difference**	
0.0056371		(-0.031346. 0.042621)	
CI based on normal approximation			
Test			
Null hypothesis		H_0: $p_1 - p_2 = 0$	
Alternative hypothesis		H_1: $p_1 - p_2 \neq 0$	
Method		**Z-Value**	**P-Value**
Normal approximation		0.30	0.765
Fisher's exact			1.000

At the 5 percent level of significance, H0 is accepted since the p-value is higher than 0.05 and therefore, the conclusion is that defect rates of univariate SPC and predictive process control are not significantly different.

6. Conclusions

In this study, a predictive process control framework for online quality control was proposed and validated in a hot rolling process. The results of the study point out the following themes:

- Data engineering is an important first step in deploying a predictive process control model. ETL was found to be the most complex step in the data engineering stage. This finding corresponds to Buvaneshwaran's [32] eight steps for developing a machine learning model. According to Buvaneshwaran [32], once a problem or objective is defined and understood, the first and foundational step in a machine learning process is data engineering.
- Feature engineering is an important second stage in deploying a predictive process control model.
- At a compliance rate of 45 percent, the metal surface defect rate of predictive process control is marginally better by 17 percent when compared to the univariate SPC defect rate. These initial positive results are consistent with positive results reported for an artificial neural network case study by Cho at POSCO [19].
- Key predictive process control aspects for all the roles of the rolling plant or organization are different across the levels. Senior management has a special role in the initial stages of deploying the predictive process control framework. This role includes the responsibilities for creating the vision for predictive process control, creating an environment for the development of base capabilities, and creating an enabling environment for the successful deployment of predictive process control. Middle- and first-line management is responsible for the technical aspects of predictive process control, while the shop floor is responsible for execution and the feedback loop. This finding is consistent with Antony's [12] MEST framework theories on management being essential to successful online quality control for SPC, which are summarized in chapter 2 of this study.
- The most important factor of the proposed predictive process control framework is base capability, followed by leadership (senior and middle management) commitment. The modelling steps follow once the two aforementioned factors have been executed correctly. This result supports Amruthnath's [24] contention that the most difficult challenge in deploying data science solutions in manufacturing applications is expertise in these techniques and their application in a real-world setting.

The aforementioned results cannot be generalized for all manufacturing processes but only for metal hot rolling processes. Because the study was limited to one case, there is an opportunity to conduct further studies at other hot rolling mills. Furthermore, future research can be conducted to determine the defect rate performance of the proposed predictive process control framework at various compliance rates, replicating the same approach across multiple defect types rather than just surface defects. Beyond metal rolling, further experiments would need to be conducted in other manufacturing industries to understand the effectiveness of the proposed predictive process control framework.

7. Declarations

7.1. Author Contributions

Conceptualization, M.R.D.M. and A.M.; methodology, M.R.D.M. and A.M.; investigation, M.R.D.M. and A.M.; writing—original draft preparation, M.R.D.M. and A.M.; writing—review and editing, M.R.D.M. and A.M. All authors have read and agreed to the published version of the manuscript.

7.2. Data Availability Statement

Data sharing is not applicable to this article.

7.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

7.4. Institutional Review Board Statement

Not applicable.

7.5. Informed Consent Statement

Not applicable.

7.6. Declaration of Competing Interest

The authors declare that there is no conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/or falsification, double publication and/or submission, and redundancies have been completely observed by the authors.

8. References

[1] Rothstein, B. (2019). 2019 Metal industry forecast. MEAS METALS Inc., Shoreview, United States. Available online: https://www.meadmetals.com/blog/2019-metal-industry-forecast (accessed on August 2022).

[2] Knop, K. (2018). Statistical control of the production process of rolled products. Production Engineering Archives, 20(20), 26–31. doi:10.30657/pea.2018.20.06.

[3] Amsden, R. T., Butler, H. E., & Amsden, D. M. (1998). SPC simplified: Practical steps to quality (1stEd.). Productivity Press, New York, United States.

[4] Mason, B., & Antony, J. (2000). Statistical process control: An essential ingredient for improving service and manufacturing quality. Managing Service Quality: An International Journal, 10(4), 233–238. doi:10.1108/09604520010341618.

[5] Bhote, K. (1991). World class quality: using design of experiments to make it happen (1stEd.). AMACOM, New York, United States.

[6] Gunter, B. (1998). Farewell fusillade: An unvarnished opinion on the state of the quality profession. Quality Progress, 31(4), 111.

[7] Woodall, W. H. (2000). Controversies and contradictions in statistical process control. Journal of Quality Technology, 32(4), 341–350. doi:10.1080/00224065.2000.11980013.

[8] Singh, R., & Gilbreath, G. (2002). A real-time information system for multivariate statistical process control. International Journal of Production Economics, 75(1–2), 161–172. doi:10.1016/S0925-5273(01)00189-X.

[9] Schmitt, J., Bönig, J., Borggräfe, T., Beitinger, G., & Deuse, J. (2020). Predictive model-based quality inspection using Machine Learning and Edge Cloud Computing. Advanced Engineering Informatics, 45. doi:10.1016/j.aei.2020.101101.

[10] Montgomery, D. C., & Woodall, W. H. (1997). Discussion on statistically-based process monitoring and control. Journal of Quality Technology, 29(2), 121. doi:10.1080/00224065.1997.11979738.

[11] Kume, H. (1987). Statistical methods for quality improvement (1stEd.). Productivity Press, New York, United States.

[12] Antony, J., & Taner, T. (2003). A conceptual framework for the effective implementation of statistical process control. Business Process Management Journal, 9(4), 473–489. doi:10.1108/14637150310484526.

[13] Illowsky, B. & Dean, S. (2021). Measures of variation. LibreTexts Statistics. Available online: https://stats.libretexts.org/Courses/Las_Positas_College/Math_40%3A_Statistics_and_Probability/03%3A_Data_Description/3.02%3A_Measures_of_Variation (accessed on August 2022).

[14] Leavegood, S. & Reeb, J. E. (1999). Statistical Process Control. Oregon State University, Corvallis Oregon, United States. Available online: https://catalog.extension.oregonstate.edu/sites/catalog/files/project/pdf/em8726.pdf (accessed on May 2022).

[15] Banks, D. (1993). Is industrial statistics out of control? Statistical Science, 8(4), 356–377. doi:10.1214/ss/1177010773.

[16] Micronite. (2015). Predictive process control. Micronite, Deerfield, United States. Available online: http://www.htrmicronite.com/predictive-process-control/ (accessed on August 2022).

[17] Funk, J. E., & Dinger, D. R. (1994). Predictive Process Control of Crowded Particulate Suspensions. Springer, New York, United States. doi:10.1007/978-1-4615-3118-0.

[18] Skiena, S. S. (2017). The Data Science Design Manual. Texts in Computer Science. Springer, Cham, Switzerland. doi:10.1007/978-3-319-55444-0.

[19] Rath, S., Thakur, S. K., Mohapatra, S. K. & Karmakar, D. (2021). Application of machine learning in rolling mills: Case Studies. 6th International Conference on Automation and Information Technology in Steel and Mining industries (AITISM'19), 17-19 October, 2019, Ranchi, India.

[20] Alshraideh, H., Castillo, E. Del, & Gil Del Val, A. (2020). Process control via random forest classification of profile signals: An application to a tapping process. Journal of Manufacturing Processes, 58, 736–748. doi:10.1016/j.jmapro.2020.08.043.

[21] Mordor Intelligence website. (2020). Big data analytics in manufacturing industry market - growth, trends, and forecast (2020 - 2025). Available online: https://www.mordorintelligence.com/industry-reports/category/information-communications-technology/Analytics (Accessed on June 2022).

[22] Shah, G. (2020). Smart manufacturing market size worth $395.24 billion by 2025. Available online: https://marketnewsreports.medium.com/smart-manufacturing-market-d7e324a9ed4c (accessed on May 2022).

[23] IFS. (2017). The IFS digital change survey. Available online: https://www.ifs.com/us/ (accessed on August 2022).

[24] Pasupuleti, M. B., & Adusumalli, H. P. (2018). Digital Transformation of the High-Technology Manufacturing: An Overview of Main Blockades. American Journal of Trade and Policy, 5(3), 139-142. doi:10.18034/ajtp.v5i3.599.

[25] Black, N. (2020). What is data engineering and why is it so important. Quanthub, Birmingham, United States. Available online: https://quanthub.com/what-is-data-engineering/ (accessed on August 2022).

[26] Altexsoft. (2021). Data engineering and its main concepts: Explaining the data pipeline, data warehouse and data engineering role. Available online: https://www.altexsoft.com/blog/datascience/what-is-data-engineering-explaining-data-pipeline-data-warehouse-and-data-engineer-role/ (accessed on June 2022).

[27] HEAVY.AI (2021). Feature Engineering. California, United States. Available online: https://www.omnisci.com/technical-glossary/feature-engineering (accessed on May 2022).

[28] Saunders, M., Lewis, P., & Thornhill, A. (2009). Research methods for business students. Pearson Education, London, United Kingdom.

[29] De Vos, A.S. (2002). Research at grass roots: a primer for the social sciences and human service professions. Van Schaick, Pretoria, South Africa.

[30] Lombard, B. (2020). Prescribe end of the week report. Company X internal document, Cape Town, South Africa.

[31] Field, A., & Miles, J. (2010). Discovering statistics using SAS. SAGE Publication, Newbury Park, United States.

[32] Buvaneshwaran, K. (2021). 8 steps to build a machine learning model. Available online: https://copyassignment.com/8-steps-to-build-a-machine-learning-model/ (accessed on August 2022).