FEDERAL UNIVERSITY OF PERNAMBUCO
TECHNOLOGY AND GEOSCIENCE CENTER
PRODUCTION ENGINEERING DEPARTMENT
GRADUATE PROGRAM IN PRODUCTION ENGINEERING

WALDOMIRO ALVES FERREIRA NETO

MAINTENANCE MODELS TO PROVIDE BETTER PERFORMANCE OF STEELMAKING PRODUCTION LINES THAT MAKE USE OF RECYCLED SCRAP

Recife
2021

WALDOMIRO ALVES FERREIRA NETO

MAINTENANCE MODELS TO PROVIDE BETTER PERFORMANCE OF STEELMAKING PRODUCTION LINES THAT MAKE USE OF RECYCLED SCRAP

Master's Dissertation submitted to the Graduate Program in Management Engineering at the Federal University of Pernambuco, to obtain the master's degree.

Concentration area: Operational Research

Dissertation Advisor: Prof. Cristiano Alexandre Virgínio Cavalcante, PhD

Dissertation Co-Advisor: Phuc Do, PhD

Recife

2021

Catalogação na fonte
Bibliotecário Gabriel Luz, CRB-4 / 2222

F383m Ferreira Neto, Waldomiro Alves.
Maintenance models to provide better performance of steelmaking production lines that make use of recycled scrap / Waldomiro Alves Ferreira Neto – Recife, 2021.
74 f.: figs., tabs., abrev. e siglas.

Orientador: Prof. Dr. Cavalcante, Cristiano Alexandre Virgínio.
Coorientrador: Prof. Dr. Phuc Do.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-Graduação em Engenharia de Produção, 2021.
Inclui referências.

1. Engenharia de Produção. 2. Aprendizado por reforço. 3. Modelos de manutenção. 4. Linha de produção. I. Cavalcante, Cristiano Alexandre Virgíno (Orientador). II. Phuc DO (Coorientador). III. Título.

UFPE

658.5 CDD (22. ed.) BCTG / 2021 - 236

WALDOMIRO ALVES FERREIRA NETO

MAINTENANCE MODELS TO PROVIDE BETTER PERFORMANCE OF STEELMAKING PRODUCTION LINES THAT MAKE USE OF RECYCLED SCRAP

Master's Dissertation presented to the Department of Production Engineering at the Federal University of Pernambuco, as a partial requirement for obtaining the title of Master in Production Engineering.

Approved in: 23 / 02 / 2021.

EXAMINATION BOARD

_____
Prof. Cristiano Alexandre Virgínio Cavalcante, PhD (Advisor)
Federal University of Pernambuco

_____
Prof. Phuc Do, PhD (Co-advisor)
University of Lorraine

_____
Profª. Caroline Maria de Miranda Mota, PhD (Internal Examiner)
Federal University of Pernambuco

_____
Prof. Dr. Philip Anthony Scarf, PhD (External Examiner)
Cardiff University

ACKNOWLEDGEMENTS

First, I would like to thank God for giving me the strength to endure all the trials and tribulations during my journey. To my family, who supported my dreams at all times and in all ways. To my two advisors. Prof Cristiano Alexandre Virgínio Cavalcante for being my mentor, supporting and encouraging me in the development of this work. And Prof. Phuc Do for accepting the challenge of guiding me. To everybody who directly or indirectly contributed to making this dream come true. Finally, to CAPES and CNPq for funding this research.

ABSTRACT

To meet growing market demands and remain competitive, modern production systems are widely adopting technological innovations, such as systems monitoring and machine connectivity, which leads a huge amount of data available about the health of the system. In this scenario, condition-based maintenance can be a powerful tool for industry competitiveness due to its ability to intervene in the system in real-time by its condition monitoring, enhancing the system availability, reliability, and cost when compared with time-based maintenance policy. However, the large amount, variety, and dimensionality of the data that comes from a production line create a problem with a large space of states, which is intractable with traditional maintenance models. To overcome this challenge, emerging tools and methodologies of the areas of Artificial Intelligence and Machine Learning are being used in the maintenance planning. Which Deep Reinforcement Learning (DRL) proved to be efficient for maintenance decision making based on multiple component conditions of a production line. Therefore, this work proposes two maintenance models: an opportunistic maintenance model considering production data to anticipate maintenance actions, and a DRL-based model to support the decision-maker in making optimal maintenance decisions in a serial production line based on system monitoring. The environment under study was a steelmaking production line. A simulation model was built to represent and simulate the behavior of the system. In the DRL model, two scenarios regarding distinct aspects of the system were investigated. A DRL framework was constructed for each scenario to learn through interaction between an agent and the simulated environment the optimal maintenance policy. Both models use as a decision criterion the minimization of the expected long-run cost rate. To evaluate the proposed models, a numerical case study was performed. The sensitivity analysis of the models was also performed to observe their behavior in the face of variations in the system parameters. As result, the models behave as expected and the proposed policies show a better result in terms of cost, system availability, and production in comparison with other time-based policies used in the steel context.

Keywords: Deep Reinforcement Learning. Maintenance Models. Steel Production Line.

RESUMO

Para atender às crescentes demandas do mercado e se manterem competitivas, as empresas estão amplamente adotando inovações tecnológicas em seus sistemas produtivos, como sistemas de monitoramento e investindo na conectividade das máquinas, o que leva a um aumento da quantidade de dados disponíveis sobre o estado do sistema. Nesse cenário, a manutenção baseada na condição pode ser uma poderosa ferramenta para a competitividade das empresas devido à sua capacidade de intervir no sistema produtivo em tempo real por meio do monitoramento da condição de seus componentes, aumentando a disponibilidade, confiabilidade e reduzindo o custo operacional em comparação com as políticas de manutenção baseadas no tempo. No entanto, a grande quantidade, variedade e dimensionalidade dos dados provenientes de uma linha de produção criam um problema com um grande espaço de estados, sendo intratável com os tradicionais modelos de manutenção. Para superar esse desafio, ferramentas e metodologias da área da computação estão sendo utilizadas no planejamento da manutenção, das quais o Aprendizado por Reforço Profundo (DRL) provou ser eficiente para a tomada de decisão de manutenção com base nas condições de múltiplos componentes de uma linha de produção. Portanto, este trabalho propõe dois modelos de manutenção: um modelo de manutenção oportunista considerando a condição do sistema para antecipar as ações de manutenção e um modelo usando DRL para dar suporte na tomada de decisão de manutenção em uma linha de produção em série baseado no monitoramento do sistema. O sistema em estudo foi uma indústria siderúrgica. Um modelo de simulação foi construído para representar e simular o comportamento da linha produtiva. No modelo usando DRL, dois cenários relativos a aspectos distintos da linha foram investigados. Uma estrutura de DRL foi construída para cada cenário para aprender, por meio da interação entre um agente e o ambiente simulado, a política de manutenção ideal. Ambos os modelos utilizam como critério de decisão a minimização do custo esperado de manutenção no longo prazo. Para avaliar os modelos propostos, foi realizado um estudo de caso. A análise de sensibilidade dos modelos também foi realizada para observar seu comportamento frente às variações dos parâmetros do sistema. Como resultado, os modelos se comportam conforme o esperado e as políticas propostas apresentam um melhor desempenho em termos de custo, disponibilidade e produtividade em comparação com outras políticas baseadas no tempo adotadas no contexto siderúrgico.

Palavras-chave: Aprendizado por Reforço. Modelos de Manutenção. Linha de Produção.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANN	Artificial Neural Network
CBM	Condition-based Maintenance
CM	Corrective Maintenance
DDQN	Double Deep Q Network
DL	Deep Learning
DQN	Deep Q Network
DRL	Deep Reinforcement Learning
MDP	Markov Decision Process
ML	Machine Learning
MTBF	Mean Time Between Failure
MTTR	Main Time to Repair
PM	Preventive Maintenance
RL	Reinforcement Learning
TBM	Time-based Maintenance

TABLE OF CONTENTS

1 INTRODUCTION

The machinery that makes up the manufacturing systems has an inevitable degradation process over time and usage, which emerges its need for maintenance actions (WANG; WANG; QI, 2014). The deterioration process of the machines is something commonly observed in practice and comes from different causes, such as fatigue and random shocks, directly impacting the machine's productivity and the quality of the final products, resorting to higher production costs (SORO; NOURELFATH; DAOUD, 2010; WANG; WANG; QI, 2014). The main goal of the maintenance is to ensure that the components of the production system maintain their functional capacity and prevent them from operating in an undesirable state (MOUBRAY, 1997).

The costs related to the maintenance activity represent a large portion of the total cost of the industry and adopting an inadequate maintenance strategy causes serious financial impacts (MOBLEY, 1990; EDWARDS; HOLT; HARRIS, 2000; WANG, 2012). For the industry, the main important issue is to reduce maintenance costs and manage risks, and, at the same time, to increase reliability, availability, and security of the assets (ATAMURADOV et al., 2017). Therefore, a widely adopted strategy is preventive maintenance (PM), since its application is essential to guarantee reliable operation of the production system equipment and to reduce the life cycle cost of the assets (ZHANG; SI, 2020).

Briefly, PM can be divided into two categories: maintenance based on time and maintenance based on condition (FITOUHI; NOURELFATH; GERSHWIN, 2017). Time-based maintenance (TBM) schedules the maintenance actions based on the age of the component, while condition-based maintenance (CBM), also called predictive maintenance, takes maintenance decisions by monitoring the system condition, i.e., the maintenance action depends on the system state or the degradation level of the machine (FITOUHI; NOURELFATH; GERSHWIN, 2017; GARRAMIOLA, 2018; ZHANG; SI, 2020). In the last type, the system monitoring occurs through data collected from sensors that record various aspects of the equipment, such as vibration, temperature, fluid pressure, and the lubricant status.

With the advancement of sensor technologies, advanced data collection techniques are being widely used to monitor the condition of the system, providing more information about the state of the system, which significantly stimulates the application and development of CBM (LIU et al., 2019; ZHANG; SI, 2020). With the increase in data availability, adopting CBM can

help planning maintenance activities more efficiently, reducing system downtime, and improving the production flow performance (NGUYEN; MEDJAHER, 2019).

As the information of the systems is becoming even more transparent and detailed, the ideal scenario would be to use all the necessary information of the machine and the system in the maintenance planning (HUANG; CHANG; ARINEZ, 2020). However, the large amount, variety, and dimensionality of the data create a problem with a large space of states, which is intractable with traditional maintenance models (WUEST et al., 2016). This availability of large amounts of data, which is often referred to as Big Data, can be even bigger when involves monitoring a multicomponent system, such as a serial production line, which is the most common type of system in the real life (WANG; WANG; QI, 2014; WUEST et al., 2016; ZHANG; SI, 2020). Besides that, the dynamic behavior of a serial production line makes maintenance management even more complex due to the interdependence between the workstations.

To overcome these challenges, promising results are found by adopting emerging tools and methodologies in the areas of Artificial Intelligence (AI) and Machine Learning (ML) in the development of intelligent systems for decision-making support in production and maintenance management (HUANG; CHANG; ARINEZ, 2020). In this scenario, DRL proved to be an efficient tool for maintenance decision making based on multiple component condition of a production line due to its ability to deal with dynamic environments subject to uncertainty and with a large number of space of states. (WANG; WANG; QI, 2014). DRL is an ML technique that combines Reinforcement Learning (RL), where an agent learns by trial and error how to interact with an environment, and Deep Neural Network to converge for a policy that maximizes a reward (MNIH et al., 2016; SAMMUT; WEBB, 2017). In the maintenance context, the reward could be a system performance indicator such as availability, reliability or cost.

Regarding the existing production lines, due to the growing concern about environmental pollution and waste reduction, together with the steady growth in the global steel demand, the steel production line with the use of recycling has played an important role in global economic activity (ZHOU et al., 2016a). This sort of production line is composed of interconnected workstations that use metal scrap as a raw material for steelmaking. Before being used, the scrap is crushed in a shredding stage, where the shredder machine is the main equipment in this station. This step is crucial for this industry because reduces the line energy consumption,

increases steelmaking efficiency and plays an important role in the quality of the final product by removing impurity and non-metallic material (ZHOU et al., 2016b).

Due to the high efforts involved in the grinding process, the shredder machine has an elevated deteriorate rate and needs frequents maintenance actions (ZHOU et al., 2016a). However, maintaining this equipment is a challenge. Due to its complex robust configuration, the high weight of the components, difficult access, and safety requirements, the maintenance activities on this equipment are time-consuming and require the complete stop of the entire workstation, which interrupts the supply of downstream processes causing loss of production (BRUSA; MORSUT; BOSSO, 2014). So, adopting an efficient maintenance policy regarding this machine brings advantages in the entire steelmaking process, which can be translated into a reduction in the total cost and, hence, gives the industry more competitiveness in the marketplace.

Therefore, this work proposes to develop two different maintenance models, which includes a DRL approach to support the decision-maker in making optimal maintenance decisions, suggesting the best time to perform PM action in a recycling usage steel production line based on system monitoring, and an opportunistic policy that anticipate scheduled maintenance inspections based on system age and the monitored production level. These approaches aim to improve the performance of a whole production line by suggesting the right moment to performing PM action in the shredder machine based on the state monitoring, reducing the expected maintenance long-run cost per unit of time. In the DRL model, two different scenarios were proposed considering distinct aspects of the production line under study. A case study was performed to evaluate the performance of the proposed maintenance models in comparison with time-based maintenance policies that can be found in the literature regarding the steel production line with recycling usage. Also, the sensitivity analysis of the models was performed to observe their behavior with variations of the system parameters.

1.1 PROBLEM DESCRIPTION

The current global increase in steel demand along with the need for reducing environmental pollution and resource-wasting made steel scrap an important resource for steelmaking (ZHOU et al., 2016a). According to the World Steel Association (2019), only in, 2018 the worldwide steel production was 1,808.6 million tons, 4.6% higher than the previous year. With this constant expansion, the steel industry has a major role in global economic

activity. With the recycling appeal and the resource shortage, the usage of steel scrap has increased. Although some countries like China are largely using steel recycling, this process is relatively weak in other nationalities (ZHOU et al., 2016a). The main reason is the overall performance of the process, which can be enhanced by improving the recycling efficiency and decreasing the recycling cost (ZHOU et al., 2016b). For this purpose, adopting an optimal maintenance policy has an important role.

Before being used, scrap must be shredded. The crushing stage is a crucial workstation in the steel production line that uses recycling, where the shredder machine is the main equipment (BRUSA; MORSUT; BOSSO, 2014). The shredder has a set of hammers that crushes the scrap. The objective of grinding is to turn the scrap into small and high-density pieces, so this crushed scrap is now used to meet the demand of the subsequent process (BRUSA; MORSUT; BOSSO, 2014). It turns the removal of non-ferrous material easy and reduces the energy consumption of the rest of the production process. So, the crushing station operations and, consequently, the shredder operation plays a crucial role to ensure good performance in this kind of steel production line, since that it is responsible for the scrap size reduction and material separation, which is an important step to improve the quality of the final product and reduce the cost of steel production (ZHOU et al., 2016a). Moreover, its output is used as input for the steel line and it has a high-throughput process, hence unavailability or performance decreasing of this process promotes an overall impact.

So, according to Zhou et al. (2016a), the shredder is the main equipment of the entire recycling usage steel production line, whose operation directly affects the line efficiency, energy consumption of the entire process and quality of the final product. Due to the high efforts involved in the grinding process, the hammers are progressively worn out, which leads to reduced efficiency in the shredding process and the necessity of a programmed stop to replace it. This PM intervention should be taken frequently to keep the degradation process under control, avoiding the breakdown of the shredder machine, which can stop the downstream process from interrupting the crushed scrap alimentation and, hence, promote several negative consequences.

In the shredder context, the main challenge in its maintenance management is to define an optimal intervention frequency that provides good control of the degradation with the lowest quantity of interruptions. It is quite difficult because the shredder has a complex degradation process due to the high efforts demanding from its process, what require intensive PM actions

to avoid its failure, but its stoppage is undesirable due to the interconnection between the workstations.

Besides that, due to its robust configuration, the elevated weight of the components, difficult access, and security requirement, considerable time is necessary to the complete stop of the machine in order to perform preventive and/or corrective actions. According to Zhou et al. (2016a), during these stops, the whole production process can be interrupted by hours or even days, causing significant production loss, in addition to the cost related to the involvement of maintenance staff and materials. Therefore, applying an inappropriate maintenance policy can lead to a negative impact on the productivity and efficiency of the machine.

1.2 JUSTIFICATION AND RELEVANCE

In the current industrial scenario, the modern production systems not only need to meet the growing requirement of the market to remain competitive, but also, they are demanded in terms of reliability and security, which make these systems increasingly complex and difficult to maintain the continuity of the operational state, especially due to their degradation processes that become stochastic (ASSAF et al., 2018; LIU et al., 2019). Thus, maintenance plays a fundamental role in the company competitiveness based on the cost, quality, and performance of the delivery of a product or service, since it supports to meet the needs for reliability, availability, and quality of equipment and products, as well as reducing costs associated with defects and equipment failures (SWANSON, 1997; PINTELON et al., 2000; AISSANI; BELDJILALI; TRENTESAUX, 2009).

Machine maintenance is a complex issue as it relates to many other aspects of modern industrial practices and affects the economy of a manufacturing system in several ways (HUANG et al., 2019). About 15% to 40% of the total production cost is attributed to maintenance activities in a factory, where it is estimated that about 30% of this value results from inefficiencies in maintenance actions (MOBLEY, 1990; WANG, 2012). Therefore, financial gains can be acquired by optimizing maintenance tasks, which legitimize the current importance of developing methodologies to make maintenance actions more efficient (WANG, 2002; AISSANI; BELDJILALI; TRENTESAUX, 2009).

Adopting an appropriate maintenance policy is essential to ensure uniform and efficient operation, but it is not trivial due to the complex and stochastic nature of modern manufacturing systems, becoming even more complicated when the components of the system have some

interdependence (DINH; DO; IUNG, 2020; HUANG; CHANG; ARINEZ, 2020). Furthermore, with the increasing requirements for reliability, availability, maintainability, and security of systems, traditional maintenance strategies are becoming less effective and obsolete (NGUYEN; MEDJAHER, 2019). To support decision-makers, equipment condition monitoring systems are being widely used, providing the ideal scenario for CBM application and development (ZOU et al., 2018; KUHNLE; JAKUBIK; LANZA, 2018; LIU et al., 2019).

In recent decades, CBM has received increasing attention due to its ability to intervene in the system in real-time, showing advances in preventing system failures and reducing operating costs (LIU et al., 2019). Despite the growing interest and the advancement of CBM techniques, traditional TBM methods are still the most used practices by the industry (HASHEMIAM; BEAN, 2011; HUANG; CHANG; ARINEZ, 2020). This is due to the lack of processes and methodologies for the use of these technologies on the shop floor, making part of the industrial equipment not to benefit from the advantages of CBM (HASHEMIAN; BEAN, 2011; JIN et al., 2016).

Even with the significant increase in data availability, only a small fraction is actually being used in the management and planning of activities linked to production in real-time (ZOU et al., 2018; STRICKER et al., 2018). Using the large amount, variety, and dimensionality of data that comes from the entire manufacturing production process creates a problem with a high number of state spaces, in which traditional TBM models are no longer applicable (WUEST et al., 2016). According to Stricker et al. (2018), a reduction of up to 25% in operating costs could be achieved through better exploration of system data in the planning and control of operational activities, which can help to obtain competitive production performance.

Besides, another huge advantage of using data from sensors is to allow a quick response to environmental changes (ZOU et al., 2018). Response time is an increasingly important strategic performance measure for a company's competitiveness. Therefore, it is necessary to develop new data-driven methodologies based on the available sensor information to, through real-time identification of the system's performance status, facilitate the control and planning of maintenance activities (ZOU et al., 2018). For this propose, ML tools arise as a good solution, as they add flexibility and adaptability to CBM models dealing with dynamic and complex environments with a large number of state spaces (WUEST et al., 2016).

Although they are the most common manufacturing systems in the industrial field, multicomponent systems such as a series production line are not well studied in maintenance

models, and most CBM application studies focus on single-component systems (WANG; WANG; IQ, 2014; ZHANG; SI, 2020). According to Zhang and Si (2020), CBM planning for systems with multiple components or units becomes even more challenging due to the possible interdependencies of the components. The study of these systems has gained a lot of attention, since real-world systems are usually complex and include multiple interacting components where these interdependencies can affect the general availability of the system and, consequently, its performance (ASSAF et al., 2018).

Regarding the recycling usage production line, despite a lot of works recognize the shredder equipment as the most important asset of the entire line and its importance on the overall production cost and system availability, the majority of the works on this field aim to improve the shredding efficiency and reduce the expected maintenance long-run cost through a better understanding of its failure mechanisms and propose some design modification or structure optimization of the shredder hammers (BRUSA; MORSUT; BOSSO, 2014; Zhou et al., 2015; Zhou et al., 2016a; Zhou et al., 2016b). Although adopting an efficiency maintenance policy can play an important role in this scenario, the maintenance schedule in shredder context is not well studied.

Given the above, it is evident the importance of study the maintenance planning into the shredder context, and also developing a PM methodology using DRL to support the decision-maker to make appropriate maintenance decisions in a multicomponent system based on its real-time monitoring information. The studied context was a recycling usage steelmaking production process, due to its importance in the current global scenarios. The inclusion of DRL in maintenance management aims to increase reliability levels of components and the system and reduce the influence of maintenance on the cost of the final product. The proposed maintenance models aim to suggest maintenance policies for the shredder machine to minimize the expected maintenance long-run cost per unit of time.

1.3 OBJECTIVES

This session presents the general and specific objectives of this study.

1.3.1 General objective

The general objective is to develop two models to provide better performance of steelmaking production lines that make use of scrap resulting from a recycled process.

1.3.2 Specific objectives

To achieve the general objective, the following specific contributions were developed:

- Conduct a literature review about maintenance and their policies, as well as about the DRL algorithms, to investigate the particularities of CBM in systems with multiple interdependent components and on the relevant aspects of maintenance policies using ML tools;
- Build a simulation model to represent the dynamics of the studied environment and develop a DRL framework to suggest preventive maintenance policies for the simulated environment based on its monitoring conditions aiming to reduce the expected long-run cost per unit of time;
- Propose two maintenance models considering different environmental assumptions to offer a more complete and realistic analysis about the problem under study;
- Apply the proposed maintenance models in a case study to validate and optimize the DRL frameworks from interactions with the simulated environment;
- Assess the proposed maintenance models performance by comparisons with other commonly used maintenance policies to measure the benefits that these methodologies can bring for the system when it is adopted;
- Perform sensitivity analysis on the models to understand their behavior when there is some variation in the system parameters.

1.4 METHODOLOGY

Regarding the approach, this work is classified as quantitative and qualitative. It is quantitative, as it answers research questions based on mathematical data and methods, and qualitative because it guides the researcher to determine analytical approaches, data collection, and research focus, enabling the generation of theories based on discoveries and understandings of reality (HABES et al., 2018).

Regarding the objective, the research is classified as exploratory, as it aims to provide greater familiarity with a specific phenomenon or to acquire new perceptions, for example, in new contexts or to formulate more precise relationships (GIL, 2002; SCHOLTEN; BLOK; HAAR, 2015).

According to technical procedures, the work is constituted as bibliographic research and a case study. The bibliography review is a fundamental step for any scientific work, as it provides relevant and sometimes unknown information to the researcher. It implies the study of articles, theses, books, and other publications normally available in indexed databases. Taking into account that the effectiveness of the model in a real context will be verified, the work is also characterized as a case study, defined as an empirical study that investigates the phenomena in their real context (GIL, 2002). Finally, about nature, the research is classified as applied, as it focuses on solving practical problems in the real world (MARCONI; LAKATOS, 2002).

Concerning about the fundamental steps to develop the work, first of all, a literature review was carried out on the evolution of fundamental maintenance concepts and their policies, as well as on DRL algorithms (their characteristics and recent applicability). This stage of the research was essential for the work, because, in addition to acquiring fundamental concepts for the study, it was possible to identify gaps in the literature, which promotes a direction for this work.

In sequence, the context under study was investigated. This step allowed to identify the real characteristics of the studied system to be taken into account and which other assumptions can be inferred about the system.

Subsequently, the behavior of the system was modeled mathematically and a simulated environment was developed to represent the system under study. In this step, two maintenance models considering distinct environment assumptions were proposed. A DRL algorithm was built and trained directly with the simulated environment, in what is called online training, to generate the PM policies. Still at this step, improvements were made to the DRL framework and adjustments to the algorithm and model until the proposed methodology converged to maintenance policies with good performance.

Both the DRL framework and simulation model was built in Python language. The artificial neural network used was provided through the high-level open-source library Keras, widely used for the creation and training of deep neural networks (GULLI; PAL, 2017). It is the most well-known Library, written in Python, for building neural networks and machine learning projects, offering modules such as optimizers, activation functions, neural layers, and cost function.

Finally, a case study was carried out seeking to validate and observe the behavior of the proposed models. A comparison among the proposed models and others time-based commonly used maintenance policies was performed to measure the benefit of their application. The sensitivity analysis of the models was performed to understand how the models behaves when some parameters of the system vary.

1.5 WORK STRUCTURE

This dissertation is composed of five sections: Introduction, Review of Literature and Theoretical Framework, Context under study and the proposed maintenance policies, Case study and Conclusion.

The first section presents an introduction about the context of the development of the work and also describes the problem, highlighting both the justification and the relevance of the work existence, as well as addresses the purpose of the study through a general objective and its specific objectives. Furthermore, the methodology applied in the work is also explained.

The second section brings the review of the literature and theoretical framework of the dissertation, presenting the major concepts used throughout the work as also the state of art.

The third section addresses more details about the context under study. The operation of a recycling usage production line is detailed along with all aspects regarding the shredder process and its maintenance issues. After that, the problem investigated is simplified to be described as a mathematical problem and a simulation model regarding the described problem is built. The maintenance policies previously proposed for the shredder context were discussed and the maintenance models proposed by this work were presented. Then, a DRL framework is built to suggest the maintenance policies for the system.

In the fourth section, the proposed maintenance models are applied to a numerical example. The usage of the DRL approach in the real-life system is explained and the performance of the proposed models is investigated when compared with other traditional time-based policies commonly used in the same context. The sensitivity analysis of the proposed models was also performed to evaluate their behavior vis-à-vis some system parameter variations.

Finally, in the fifth section, both the conclusion and researching finds are presented.

2 REVIEW OF LITERATURE AND THEORETICAL FRAMEWORK

In this section, a review of the literature is provided. The main concepts, methodologies, and tools used in the dissertation are presented in detail for a better understanding of the work.

2.1 REVIEW OF LITERATURE

The technological innovation of Industry 4.0 created an industrial environment with a high degree of connectivity and automation of machines, resulting in greater complexity of machines and production systems (KUHNLE; JAKUBIK; LANZA, 2018). Still according to the authors, in this context, maintenance plays an important role in the efficient use of systems in terms of cost, reliability, and availability.

Regarding the abstraction level, maintenance policies can be categorized into multi-component or single-component policies (HUANG; CHANG; ARINEZ, 2020). Almost all existing CBM studies are directed to single-component systems, which often suggests that maintenance should be performed when the monitored system reaches a certain level of degradation, called the maintenance threshold (DIEULLE et al., 2003; CHEN et al., 2011; CHEN et al., 2015). Usually, the maintenance threshold is optimized along with some other decision variables, for example, the frequency of inspection to minimize the cost or to maximize availability criterion (GRALL; BÉRENGUER; DIEULLE, 2002; LIAO, ELSAYED; CHAN, 2006). More practical issues have recently been considered in CBM planning, for instance, measurement accuracy, variation in the operational costs with the age and state of the system, and the presence of sensor deterioration (JONGE; TEUNTER; TINGA, 2017; LIU et al., 2017; LIU et al., 2019).

Although an optimal solution for single-component systems can be obtained efficiently in the majority of the cases, the modern machinery and manufacturing systems are composed of several interdependent components or subsystems (DINH; DO; IUNG, 2020; ZHANG; SI, 2020). The dependence between components, for instance, economic dependence, stochastic dependence, and structural dependence make CBM planning for systems with multiple components even more challenging. To consider stochastic dependencies, Rasmekomen, and Parlikad (2016) developed a CBM policy for systems with K components where the condition of one machine affects the degradation rates of the others. They have set K maintenance limits to perform preventive replacements. Do et al. (2019) proposed a CBM policy for a two-

component system, establishing two thresholds for preventive replacement and two thresholds for opportunistic maintenance, considering both stochastic and economic dependencies.

In a serial production line, a common practice is to use intermediate buffers to minimize the effect of interdependence between the workstation. In this context, Karamatsoukis and Kyriakidis (2010) proposed a CBM for a two-machine system with an intermediate buffer defining thresholds to trigger maintenance actions. Fitouhi, Nourelfath, and Gershwin (2017) studied the advantages of considering the effects of buffer in threshold planning in a similar two-machine system.

In these works, as in most studies that consider CBM for multi-component systems, multiple maintenance thresholds are established. However, looking for optimal maintenance thresholds is often suitable for problems with a low state space, but becomes a challenge when the state space increases (ZHANG; SI, 2020). Besides, setting the maintenance thresholds is not a trivial task in practice, and may induce the adoption of a non-ideal maintenance policy (NGUYEN; MEDJAHER, 2019).

To deal with environments with a large amount of state space, ML techniques were incorporated into CBM (STRICKER et al., 2018). RL algorithms have shown to be applicable for determining optimal policed for different manufacturing tasks in the flow line, including maintenance activity (WUEST et al., 2016). Aissani, Beldjilali, and Trentesaux (2009) developed a RL approach to schedule maintenance tasks into a petroleum production line. Wang, Wang, and Qi (2014) proposed a RL based maintenance policy for a flow line system with resource constraints that suggest performing maintenance activity based on the products reject rate and buffer level. Stricker et al., (2018) use RL for production planning in the semiconductor industry. Kuhnle, Jakubik, and Lanza (2018) developed a RL maintenance policy for a parallel production line fed by an initial buffer.

RL has shown a lack of scalability when the dimension of the problems gets too high (ZHANG; SI, 2020). Besides, RL training was revealed to be time-consuming and required considerable data and memory to be trained efficiently (HUANG; CHANG; ARINEZ, 2020). To overcome this, Artificial Neural Networks (ANN) have been incorporated in the traditional RL algorithm to solve the scalability problem and reduce the computational effort (MNIH et al., 2016). This variation is called DRL, which has gained so much attention recently due to its greater scalability, an ideal characteristic to act on large-scale realistic problems, characterized by highly dimensional state-action spaces (ROCCHETTA et al., 2019). Huang, Chang and

Arinez (2020) proposed a DRL base maintenance policy to a serial production line considering the buffers level and the operational time of each machine. Zhang and Si (2020) developed a similar maintenance policy for a multicomponent system with risk dependence where a failure of one component means the failure of all system.

Regarding the steel production line that uses scrap from a recycling process, as mentioned before, the maintenance schedule in the shredder context is not well studied. Araújo et al. (2018) analyzed the effect of replacing defective hammers during maintenance actions in a time-based maintenance policy for the shredder. Ferreira Neto et al. (2020a), proposed a time-based inspection policy for the shredder equipment where periodical inspections were suggested, but it would be anticipated when opportunity windows arise based on the operational time and the buffer level.

These papers as well as most maintenance policies used in the shredder system presents time-based policies and do not explore the advantages of using information from system monitoring. Recently, Ferreira Neto et al. (2020b) suggest a maintenance policy based on the system monitoring in the steel production line context. They developed a DRL framework to suggest PM action for the shredder machine. The policies proposed by Ferreira Neto et al. (2020a) and Ferreira Neto et al. (2020b) are directly related to this dissertation. In this work, besides a deeper analysis of the policies proposed by Ferreira Neto et al. (2020a) and Ferreira Neto et al. (2020b), different scenarios covering distinct assumptions of the environment were analyzed, and also more realistic aspects of the maintenance problem were covered, resulting in maintenance policies and analysis closer to reality.

2.2 THEORETICAL FRAMEWORK

The main concepts, methodologies, and tools used in the dissertation are presented here.

2.2.1 Maintenance

Maintenance can be defined as a class of activities that can restore a failed or deteriorated asset to its functional state to carry out the designated function (DHILLON, 2002). Thus, maintenance aims to ensure that the components of the productive system maintain their functional operation capacity, seeking to increase the availability of the system at the lowest possible cost (MOUBRAY, 2000). This class of activities is the key to ensure a highly reliable

operation of modern engineering assets and to reduce the asset's life cycle cost (ZHANG; SI, 2020).

The maintenance function influences the availability of the manufacturing system and its rate of resource utilization (AISSANI; BELDJILALI; TRENTESAUX, 2009). According to the authors, the objective of maintenance management is to avoid components/system failures and maximize the availability of the installation at low maintenance cost. Maintenance activity affects the economy of a manufacturing system in several ways. About 15% to 40% of the total production cost is attributed to maintenance activities in a factory (WANG, 2012). Therefore, a good maintenance policy is fundamental to guarantee a uniform and efficient production operation (HUANG; CHANG; ARINEZ, 2020).

A maintenance action that reacts to a random machine failure is known as corrective maintenance (CM). The consequences of random machine failures are often unpredictable and even catastrophic in some situations (WANG, 2002). To reduce these random failures, PM is applied. In PM, activities are carried out proactively, even if the equipment is not defective, to keep it at the desired level of reliability (HUANG; CHANG; ARINEZ, 2020). However, there is a trade-off when making preventive actions. If actions are not taken on time, the system would be interrupted by random failures more often, which could lead to significant production losses. On the other hand, if the actions were very frequent, the costs caused by them may far outweigh the benefits that preventive maintenance could bring, since some actions would be unnecessary (HUANG; CHANG; ARINEZ, 2020).

The CM policy conditions the execution of an intervention on the device to the occurrence of a failure. This maintenance policy is generally used when the effort to prevent the machine or system failure is higher than its impact.

The PM policy is indicated for equipment whose cost involved in correcting its failure exceeds the costs recorded for preventive actions, such as inspections, repairs or preventive replacements. PM can be defined as any maintenance action that precedes the degradation of the quality of products and equipment (CAVALCANTE; ALMEIDA, 2007). Therefore, its application has the objective of increasing the useful life of the components, providing long-term benefits. Thus, as the systems become more complex, this approach tends to be more appropriate, since it significantly reduces interruptions in the production line, which are high-cost events.

Based on the maintenance decision criterion, PM can be divided into two types: TBM and CBM (ZHANG; SI, 2020). TBM policy monitors the equipment's lifetime, deciding about when is the best time to perform actions based on this parameter. Only aspects inherent to the equipment's reliability are considered in the maintenance schedule, e.g., MTBF and MTTR, disregarding any external aspect that may eventually influence in reducing the component's useful life. The CBM, also called predictive maintenance, presents the proposal to monitor the state of the equipment, making the decision on when to perform its actions based no longer on time, but on the state of the equipment. Some practices, such as continuous monitoring of equipment and/or inspections, are carried out to control, measure, and evaluate some parameters of the system, thus observing its state.

Despite the popular belief that adopting a PM policy is sufficient to safeguard the proper functioning of the machines, this statement is not confirmed and eventually, some CM actions can be necessary (NAKAJIMA, 1989).

With the technological advances of the sensors, more information about the degradation of the system can now be accessed, which provides a great basis for the use of CBM. This technique has evolved over the years from the use of visual inspections, which is the oldest method and still one of the most powerful and widely used, to automated methods that use advanced signal processing techniques based on pattern recognition, including, for example, neural networks and ML (HASHEMIAM; BEAN, 2011).

The idea behind it is that as the equipment starts to fail, it can display signals that can be detected. CBM aims to identify the beginning of degradations and failures of the equipment, hence avoid its failure. This maintenance practice, due to its ability to predict future failures, can avoid unnecessary replacement of equipment, save costs, and improve safety, in addition to reducing downtime and planning effort for maintenance activities (HASHEMIAM; BEAN, 2011; KUHNLE; JAKUBIK; LANZA, 2018).

2.2.2 Machine learning

Due to the increasing data availability, ML algorithms have recently gained a lot of importance. ML is a subset of AI that allows a computer to learn about past experiences and improve its behavior when performing a given task. They are considered algorithms that are not programmed explicitly with an exact deterministic procedure (STRICKER et al., 2018). According to Samuel (1959), its real objective is to allow computers to solve problems without

being specifically programmed to do so. To Alpaydin (2010) the goal of certain ML techniques is to detect certain patterns or regularities that describe relations. ML field leads a variety of different sub-domains, algorithms, theories, and application possibilities. Today, ML is already widely applied in different manufacturing areas, for example, optimization, control, and problem-solving (WUEST et al., 2016).

According to Stricker et al (2018), ML algorithms are called data-based approaches because input training data directly affects their performance. Although it can be classified in different ways depending on the author, the most accepted ML classification is: supervised learning, unsupervised learning and RL (WUEST et al., 2016; SUTTON; BARTO, 2018). This classification distinguished the ML by the kind of feedback, the representation of learned knowledge, and the availability of prior knowledge (STRICKER et al., 2018).

In supervised learning, a set of labeled data is available, both input and output. Thus, the machine receives the data previously selected and categorized so that the machine will learn from the interactions between the expected inputs and outputs. Sutton and Barton (2012) summarize supervised ML as an apprenticeship based on examples provided by an experienced external supervisor. The learning process occurs with the machine checking its output with the correct answer (label) provided by a teacher and making the correct adjustments.

In unsupervised learning, data is not labeled, that is, there is no feedback from an experienced external teacher. The purpose of the machine is to identify clusters in the existing data set. Thus, while supervised learning focuses on classification due to known labels coming from the teacher, unsupervised learning aims to discover unknown classes of items by grouping (WUEST et al., 2016). Basically, in unsupervised learning, the machine tries to learn some patterns without an identified output or feedback.

RL is a method of learning by trial and error. It consists of one or more decision entities called agents interacting with an environment, where they learn to choose ideal actions that maximize rewards or minimize losses for systems (ZHANG; SI, 2020). It is characterized by a sequential decision-making process. The agent chooses actions based on the observed states of the environment, and each action results in a reward, as well as the induction of a next state of the environment (SUTTON; BARTO, 2018). Different from the previous ones, in RL an agent has to find which actions cause the best results in the environment by interaction instead of being told, i.e., without labeled data indicating which action is good or bad (WUEST et al., 2016). Through its learning process, an agent learns an action policy that maximizes the

expected accumulative reward. RL was developed aiming to imitate the learning process of human beings.

2.2.3 Reinforcement Learning

As said before, RL is a learning algorithm where an agent learns with trial and error by interactions with the environment which action should be performed in each state to maximize the future reward, considering the environmental uncertainties (ROCHETTA et al., 2019). In a simple way, the agents perform action into the environment and receive a feedback signal about how good the actions were, so they decided whether or not this action should be added to their repertoire.

The reasons that make this method attractive include: a trial-and-error learning process by interactions with the dynamic environment, convergence to a stationary policy and model-free learning without transition probabilities (WANG; WANG; QI, 2016). Also, its learning technique makes RL appropriate for generating on-line solutions, i.e., give solutions in real-time (WUEST et al., 2016). In addition, due to its adaptability to dynamic systems, RL is appropriated to solve many of the manufacturing problems (WANG; WANG; QI, 2016). Besides that, due to their immanent flexibility and performance in learning strategies for real (or simulated) systems, RL algorithms are suitable to find optimal maintenance schedules in the context of a stochastic production environment (KUHNLE; JAKUBIK; LANZA, 2018).

RL algorithms are robust and accessible, their lack of scalability prevents them from being applied to solve a series of real-world problems with a large scope of action (HUANG; CHANG; ARINEZ, 2020). To overcome this, a DRL method emerges to provide to reinforce RL, through the integration of Deep Learning (DL), scalability, and efficiency to deal with practical decision-making problems (ARULKUMARAN et al., 2017).

2.2.3. Deep Reinforcement Learning

DRL is an integration of RL and DL that provides a powerful approach and representation learning properties, which significantly facilitates computing speed and is, therefore, suitable for problems with high-dimensional state space.

The majority of the RL algorithm relies on an intensive tabular memory (ROCHETTA et al., 2019). This RL algorithms create a state-action value function, which associate each state with a correspondent action. In the learning process, these values are stored in a memory table and with each interaction with the environment, these values are updated until the agent finds

an optimal policy. When the problem has a high dimension space states, the tabular method needs a huge memory and becomes computational burdensome besides needs a lot of data to fulfill efficiently the memory table (HUANG; CHANG; ARINEZ, 2020; ZHANG; SI, 2020). Therefore, most applications of the traditional RL algorithms have been restricted to problems with low-dimensional state spaces, e.g., a single-component system (ZHANG; SI, 2020).

The DRL uses some regression tools available in the DL to replace the tabular representation of the traditional RL algorithms and, hence, adapt this method to deal with large state spaces (ROCHETTA et al., 2019). So, The RL capability to deal with real-problems has been increased with the help of deep neural networks.

DL is a branch of machine learning based on deep neural networks that have shown great success in several applications in recent years, mainly to solve problems involving high-dimensional data (HELBING; RITTER, 2018). In terms of structure, DL is a kind of ANN that process the data through multiple layers toward highly non-linear and complex feature representations (WUEST et al., 2016; WANG; WANG, 2018). DL is mainly used by the computer vision and language processing areas, it offers great potential to also boost data-driven manufacturing applications (WUEST et al, 2016). It uses the same principle of the ANN, but the difference is that ANN is designed by developers and the DL learned from data through a self-learning procedure (WANG; WANG,2018). Therefore, it can offer a more suitable and convenient way of treating the feature extraction problems and requires a large set of data for the training.

So, the DRL uses DL to trains a deep neural network that will replace the tabular matrix used in traditional RL algorithms. The deep neural network conveys an approximate value that would be provided if a tabular method was used but requiring less memory and computing effort.

3 CONTEXT UNDER STUDY AND PROPOSED MAINTENANCE POLICY

In this section, the operation of the context under study will be detailed. A simplification of the system will be presented, along with its mathematical modeling and simulation model. The proposed maintenance models will be presented, then, a DRL framework will be developed to provide a maintenance policy for each case.

3.1 SYSTEM DESCRIPTION

A steel production line that uses a recycling process as an input consists of several interconnected workstations and could be seen as a multi-component system with potential interactions among them. The steelmaking process begins when the metallic material to be fragmented is placed on the shredder's feed chute to be subjected to several collisions. This conveyor belt feeds the shredder with a large variety of shredding material composed predominantly of steel such as cars, motors, and fridges, where they are transformed into small dimension crushed scrap (SANDER; BERNOTAT, 2004). The collisions between the shredder hammers and the scrap create micro-cracks into the material and rupture the scrap by the shearing process. It is also common to press the scrap before the shredder feed in order to reduce its volume, avoiding material jams, thus ensuring that the fragmentation process occurs more effectively. The shredding process occurs until the scrap size is reduced to match a required granulometry to pass through the grid located at the bottom of the equipment (KIRCHNER; TIMMEL; SCHUBERT, 1999).

The resulting fragments are moved through a conveyor belt to the magnetic separators where the non-ferrous material is removed. After that, the crushed scrap is stocked in an intermediate buffer. The Electric Arc Furnace (EAF) pulls material from the buffer to perform the melting process. The EAF is also responsible for the primary refining of the material. The liquid steel resulted in this process is poured into a pan and moved to a treatment plant where the secondary refining will be executed. The last station is responsible for the continuous casting and its output is the final product.

Due to the workstation interdependence, shredder stoppages interrupt the flow of crushed scrap, which can starve the entire plant. When this happens, the industry can feed the EAF with unprocessed scrap which increases substantially the production cost and impact the quality of the final product. To reduce the interdependence between the shredding station and the remaining downstream processes, an intermediate buffer is used. The buffer station can enhance

a serial production line performance by minimizing the effect of workstation stoppages (AMEEN et al., 2018). The buffer can keep supplying the production line when the shredder undergoes maintenance activities until its content has been consumed.

On the other hand, a trade-off emerges between the increase of line efficiency and cost due to the material storage (GROOVER, 2015). According to Gan et al (2013), to maintain a production line operating with a good efficiency level, a large capacity buffer is needed to deal with a high frequency of PM. However, when the maintenance frequency is low, it is also necessary to deal with the possible increase in random failures. Hence, buffer management rise as an important issue to be considered in the maintenance planning in order to minimize the impacts of these interventions on the overall line efficiency (SCHOUTEN; VANNESTE, 1995; CHEUNG; HAUSMANN, 1997; DELLAGI; KHATAB, 2014).

Regarding the shredder, this machine consists of a set of fixed hammers distributed in a horizontal rotor, whose rotation converts kinetic energy to achieve a strong impact and comminute the metallic scrap to be feed into the EAF (BRUSA; MORSUT; BOSSO, 2014; ZHOU et al., 2016b). The operating principle is based on a sharp reduction in the size of scrap through successive impacts between the hammers and the material caused by high rotor rotation speeds and the centrifugal inertial force present in the fragmentation elements. The shredding process demands a complex and huge effort that is mainly supported by the hammers (ZHOU et al., 2016b). The different efforts that rise with the impact between the hammers and the scrap, which is composed of a diversity of metallic materials, make the hammers the most damaged component of the shredder.

According to Brusa, Morsut, and Bosso (2014), the extremely bad conditions that hammers are subjected to during their operations induces their failure quite easily. The repetitive high-speed impact with steel composed material in addition to the possible presence of aggressive elements such as sand, pieces of wood, roots, and stone, highly affects the hammers causing it to wear out until its failure. This continuous friction between the hammers and scrap causes fatigue, thermo-mechanical fatigue, abrasion and other degradations processes in the hammer (BRUSA; MORSUT; BOSSO, 2014). This degradation process can be even more aggressive when the shredder is fed with largely heterogeneous material sources, i.e., metal objects coming from very different items, from a car to a steel pipe, for example (ZHOU et al., 2015).

According to Zhou et al. (2016a), the hammers degradation impact its operational ability to shred the scrap, and hence impact in the total production capacity of the shredder machine. Hammers degradations and failures decrease the shredder facility to shear the scrap leading to a productivity reduction. Consequently, the integrity of the hammers is directly related to the shredder production. For these characteristics, the hammers can be considered as the critical component in shredder operation. Worn hammers can be seen in Figure 1.

Figure 1 - Worn hammer

Source: Zhou et al. (2016a)

Various types of failure coexist in a hammer, such as wear, breakage, and fracture, where the wear failure is more noticeable than other types of failure (ZHOU et al. 2016b). When the wear on the hammer is enough, a breakage occurs. Figure 2 shows failed hammers damaged at various degree. The aspects related to failures and imminent failures can be identified by inspection or monitoring operation. During a machine inspection, the wear and tear of the hammers can be detected by mass reduction, deformation, and change of the edge cutting shape. These anomalies mean that an imminent hammer failure is coming. The hammer containing these anomalies can be characterized as a defective hammer. This characteristic allows the failure process to be described as following the delay time concept, where the component can visit three states: good, defective, and failed (CHRISTER, 1999). As the cutting ability of the hammer depends on its state, a good and defective state hammer remains operational while a failed hammer is unable to carry out its function and impacts the overall equipment production capacity. During the system stoppages, when a hammer containing anomalies is identified, its replacement is performed. Also, other PM actions such as re-tightening, adjustments, and cleaning are also executed during this time.

Figure 2 – Failed hammers with severe wear out process

Source: Zhou et al. (2016b)

In this work, maintenance models that consider the monitoring condition of the system and the level of the buffer are applied to a steel production line to determine the optimal time for PM activity on the shredder machine.

3.2 SYSTEM SIMPLIFICATION AND GENERAL ASSUMPTIONS

By the analysis of the system under study, it can be described as a steel production line composed of two workstations with an intermediate finite buffer (FERREIRA NETO et al., 2020a). The first station is a crushing stage where the shredder machine processes the scrap and transfer the crushed scrap for the buffer with a production rate P. The intermediate buffer has a fixed capacity K. It is assumed that this station is fed by an unimitated scrap stock, i.e., it never starves. The second station regards the remaining necessary processes to transform the crushed scrap into a steel product. This station pulls the material from the buffer with constant demand d.

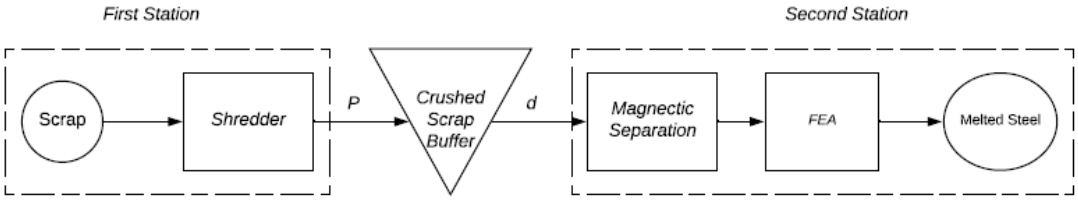
The shredder machine is equipment composed of n identical hammers that operate in parallel and deteriorate with time. The deterioration process of the hammers follows the delay time concept (CHRISTER, 1999), where the hammer can be found in three states: good, defective and failed. It is assumed that each hammer in a good and defective state contribute for the shredder productivity. When a hammer fails the shredder productivity P decreases. Therefore, P is variable over the time and the instantaneous productivity of the shredder at time

t is defined as $P_t = NP_i$, where P_i is the individual productivity of each hammer and N is the quantity of good and defective hammers at that instant t.

Another issue related to the shredder operations is its failure. Although a parallel system fails only if all its components fail, the failure process of the shredder is considered as k out of n, i.e., if the quantity of failed hammers reaches k, the equipment fails. This failure process regards the practical operation of the equipment under study. When the number of failed hammers reaches a certain number, the decrease in productivity is so significant that keeping the system working is not feasible anymore.

After the crushing process, the processed scrap is stocked on the subsequent buffer, which has a finite and known capacity *K*, and the rest of the steel production line is fed by this buffer. The remaining stations of the production line form the second workstation. As stated before, this station needs to be supplied with a constant demand d to keep working at its full capacity. This demand is lower than the initial productivity of the shredder ($P_{t=0} > d$), so it allows the load of the buffer. A simplification of the system under study is illustrated in the Figure 3.

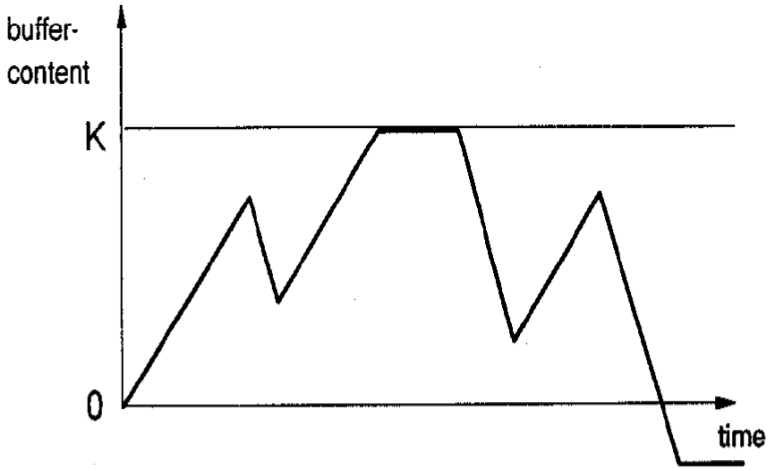
Figure 3 - The System under study



Source: the author (2020)

The loading and unloading of the buffer occur because of the difference between the shredder productivity P and the demand d. The variation of P over time and stations stoppages due to maintenance action or breakouts cause mass variation in the buffer, which can be positive when $P > d$, or negative when $d > P$. The reason for the existence of the buffer is to keep feeding the second station with d when $P < d$, reducing the impact of the shredder stoppages into the production line. Figure 4 illustrates the buffer mass variation over time.

Figure 4 - Buffer level over time



Source: Van der Duyn Schouten and Vanneste (1995)

The instantaneous variation rate of the buffer content is *P*-*d*. During the shredder station stoppages, this rate is $-d$. In Figure 4, when the buffer level is lower than zero means that unmet demand occurred. An unmet demand happens when the current volume of the buffer along with *P* do not meet d over a certain period of time. As a consequence, the second station starves and hence the line ceases its production. Thus, the amount that would be produced when the line was unavailable is the unmet demand.

The following assumptions about the system under study are made in this study:

1) The shredder is composed of n identical components;
2) The shredder fails if k-out-of-n components fail, where $0 \leq k < n$;
3) When shredder fails, the entire station stops its operation and the CM must be conducted;
4) The hammers failure follows the delay time concept;
5) The delay time and the time until the arrival of defect follow probability distributions F_x and F_h, respectively, and are statistically independent and the same for all hammers;
6) The hammers operate and fail independently of each other;
7) Each hammer has the same fixed and constant individual productivity P_i;
8) To reduce random failures, the system can be turned off to receive a PM;
9) The duration of PM is constant and known (T_P);

10) The duration of CM on equipment takes the time T_P plus one extra time that follows a probability distribution F_c.

11) During PM and CM, the failed and defective hammers are replaced;

12) After replacement, the hammer returns to a state as good as new;

13) Both PM and CM are perfect, i.e., do not exist misclassification or defect/fail induction;

14) The demand of the production line d is known and constant;

15) The shredder productivity P varies over the time. At time t, its productivity is P_t, where $0 \leq P_t \leq nP_i$;

16) The buffer has a finite capacity K. The buffer levels are changing with the dynamic of the system. The buffer level at time t is b_t, where $0 \leq b_t \leq K$;

3.3 NOTATION

The notation used in this work is presented in Table 1.

Table 1 - Notation

X, F_X	Component age at defect arrival (non-negative random variable) and its cumulative distribution function
H, F_H	Delay-time (time from the arrival of the defect to the failure) and
f_H, F_H, R_H	Density probability function, cumulative distribution function and reliability function of H
K	Buffer capacity
n	The total amount of hammers
k	The number of failed hammers to the shredder failure
$N_r(t)$	Cumulative quantity of replaced hammers up to time t
$D(t)$	Cumulative amount of unmet demand up to time t
$CM(t), PM(t)$	Total CM and PM activities carried out up to time t
$C(t, \pi)$	Total maintenance cost up to time t
$c_l,\ c_f,\ c_s, c_i, c_r$	Cost of unmet demand, failure, storage, intervention and replacement, respectively
T_c, T_P, TTR	Duration of CM, PM, and the total duration of the maintenance actions, respectively
F_c	Cumulative distribution function for the extra time in the CM

d	Constant demand of the production line
P	Shredder productivity
$b(t)$	Buffer level at time t
P_i	The individual productivity of each hammer
π	PM policy
θ	Parameters for neural network
ε	parameter for -greedy exploration in RL
γ	discount factor
$Q(s,a)$	Q-value
$Q(s,a,\theta)$	Q-value approximated by neural network θ
S_t	System state at time t
R_t	Reward received at time t
a_t	Maintenance action decision taken at time t

Source: This research (2020)

3.4 SYSTEM MODELING

Although the described system in this work really exists in real life, the steel production line system studied here was implemented as a simulation model. A discrete event simulation model was implemented to simulate the operation of the system under study. The shredder was modelled as a multi-component system composed of n identical hammers working in parallel. Due to its operational characteristics, the degradation process of each hammer was assumed to follow the delay time concept (CHRISTER, 1999), in which the hammers can be in three states: good, defective, and failed. The time until the defect arrival is *X*. The hammer stays operating in the defective state until it fails. The time from the arrival of the defect until the failure is called delay time (*H*). These times are assumed to be two statistically independent variables that follow the probability distributions F_x and F_h, respectively.

The simulation model starts running with all hammers as good as new, the production rate $P_{t=0} > d$ and with $b_{t=0} = 0$. The operation of the system causes the load and unload of the buffer. Hence, the system operation has some costs such as cost per unit stored in the buffer (c_s) and cost per unmet unit (c_l), when it happens.

During the maintenance action, the defective and failed hammers are replaced at a replacement cost c_r per hammer. Any shredder stoppage incurs a fixed cost c_i related to the maintenance staff and necessary materials. The duration of maintenance action (*TTR*) varies

depending on the level of the damage. Because it is a planned activity, the duration of a PM action is assumed to be constant and equal to T_P. When *k*-out-of-*n* hammers fail, then the shredder also fails and the CM should start immediately. The duration of the CM (T_c) varies depending on the damage degree caused. It is assumed that the CM activity takes the time of PM activity plus the extra time that follows F_c, where $T_c > T_P$.

Each interaction of the simulation model counts one-unit time regard to the system operation or *TTR* when some maintenance action happens. To simulate the system, the simulation model runs during *T* interactions, where the system behavior is observed. The equations to calculate the operation cost at each operation time *t* and maintenance cost can be found below.

$$Operational\ cost\ _t = c_s b_t + c_l D_t \tag{3.1}$$

$$Unmet\ demand\ D_t = \begin{cases} d - P_t - b_t, & if\ d > P_t \\ 0, & if\ d \leq P_t \end{cases} \tag{3.2}$$

$$PM\ cost = c_i + c_s \int_o^{T_p} b(t)\, dt + c_l \int_0^{T_p} D(t)\, dt + c_r N_r \tag{3.3}$$

$$CM\ cost = c_i + c_s \int_o^{T_f} b(t)\, dt + c_l \int_0^{T_f} D(t)\, dt + c_r N_r + c_f \tag{3.4}$$

The simulation model developed to simulate the dynamic and behaviors of the system is described in Figure 5.

Figure 5 - Simulation model algorithm

Simulation model:

input parameters: $c_l,\ c_f,\ c_s, c_i, c_r, n, K, d, P_i, F_c, T_P, F_x, F_h, k$
generate values of *X* and *H* for each hammer by *Monte Carlo Method* using F_x and F_h
$b_0 \leftarrow 0$
for $t = 0,1, \ldots, T$ **do**
    check the operation time of each hammer
    evaluate the condition of each hammer
    if $N \geq k$ **do**
        calculate the production rate P_t
        perform buffer mass analysis, i.e., calculate b_t and D_t using Eq. (3.2)
        calculate the operational cost using Eq. (3.1)
        if preventive maintenance is trigger **do**
            generate new values of *X* and *H* for defective and failed hammers
            perform buffer mass analysis during T_P
            calculate the maintenance cost per Eq. (3.3)
        end if
    else
        generate news values of *X* and *H* for defective and failed hammers
        gauge the duration of the corrective maintenance by *Monte Carlo Method* using F_c

19: perform buffer mass analysis during T_c
20: calculate the maintenance cost using Eq. (3.4)
21: **end if**
22: compute the interaction cost
23 store interaction cost
24: **return** cost per unit of time

Source: This research (2020)

3.5 PROPOSED OPPORTUNISTIC MAINTENANCE POLICY

The opportunistic maintenance policy is a kind of a PM policy that allows PM activities, such as preventive inspections and replacements, to be carried out before the scheduled time when an opportunity maintenance window appears. Performing maintenance actions during the opportunity window could bring some benefits such as a lower maintenance activity price and reduction in the maintenance duration, which can lead to a long-term advantage.

Regarding the steel production line, the proposed opportunistic maintenance policy for the shredder machine recommends when the shredder should be inspected based on its operational time and monitored buffer level. The policy suggests periodical inspections to occurs in time T, but it could be anticipated when the system has already worked t_{min} time after the last maintenance action and the buffer level has reached a minimal level b_{min}. Let t be the instantaneous operational time of the shredder and b_t be the instantaneous buffer level at time t, so Ferreira Neto et al. (2020a) suggest to perform the inspection when one of the listed criteria is reached:

i. $t = T$
ii. $t \geq t_{\ min}\ and\ b_{min} \leq b_t < K$
iii. $b_t = K$
iv. $b_t = 0$

The time counting t should be restarted after any maintenance intervention and the beginning of the count coincides with the beginning of the shredder operation. The system is scheduled to be inspected in each T time. However, an opportunity to anticipate the inspection arises when the operating time t exceeds $t_{\ min}$ and the buffer level b_t is between $b_{min} \leq b_t \leq K$. Therefore, besides the periodical inspection time T, the policy has two more decision variables: $t_{\ min}$and b_{min}. Another condition that makes this anticipation possible occurs when the buffer is completely full or completely empty.

Thus, an inspection is plausible to be performed at any time when a criterion is met or when a failure occurs. During the maintenance interventions, both defective and failed hammers are replaced. According to Ferreira Neto et al. (2020a), the idea behind this maintenance strategy is to use this opportunity window to capture the dynamic of the system and to suggest an optimal time for the inspection, providing fewer interruptions, and enhance productivity and efficiency of the rest of the production line.

Although the opportunistic policy appears to be complex, the control variables are simple to monitor, and the production line performance is expected to be improved. The accumulated cost of an operational cycle can be calculated using the equations 3.1, 3.2, 3.3 and 3.4. The policy performance is positively influenced by how well the decision variables were chosen. Figure 6 shows the algorithm that was built to measure the expected long-term cost of a given opportunistic policy represented by a combination of decision variables (T, t_{min}, b_{min}).

Figure 6 - Opportunistic maintenance algorithm

Opportunistic maintenance algorithm:

input parameters: (T, t_{min}, b_{min}) values, stopping criterion and number of simulations
while stopping criterion is not reached **do**
    run the simulation model in Figure 5
    if failure occurs before reach any criteria **do**
        observe the interaction cost
        calculate the C_∞ which is the cost per unit of time
    else
        detect which criterion was reached first
        trigger preventive maintenance in the simulation model at the time the criterion was met
        compute the interaction cost
        calculate the C_∞
    end if
    store C_∞ of this interaction
end while
C_∞ of the evaluated policy is the average of all C_∞ found
return C_∞

Source: This research (2020)

The algorithm described in Figure 6 considers two-stop criteria: one for convergence and the other for the quantity of interaction. In the first interaction, the algorithm considers all hammers as good as new. After that, the state of the hammer at the beginning of the next

interaction will be decided by the dynamic of the system. The convergence criteria compare the C_∞ of two consecutive interactions. If the difference is less than or equal predefined value, the algorithm considers that convergence occurs and assigns the value of C_∞ to this policy. About the number of interactions, a certain amount of interaction is simulated, if there is no convergence, the algorithm considers the last value of C_∞ found to this policy. In order to reduce the effect of the variations between each simulation, the algorithm performs a certain number of simulations chosen by the user and takes the average value of C_∞ as the final result of the evaluated policy.

Besides what was previously discussed, this policy generalized multiple special cases to use, such as: time-based inspection policy ($t_{min} \geq T \; or \; b_{min} \geq K$), corrective maintenance policy (when T and t_{min} go to infinity) and buffer-level-based policy (when T goes to infinity and $t_{min} = 0$). This flexibility of the model allows to suggest the manager to use the opportunities only if they provide better results in terms of costs per unit of time. Some of the most common policies used in the context of the shredder, which are special cases of the proposed opportunistic policy, will be briefly commented below.

3.5.1 Time-based maintenance policy

According to Araújo et al. (2018) and Ferreira Neto et al. (2020a), regarding the shredder machine, the steel industry extensively adopts this sort of maintenance strategy. The machine is subjected to a periodic PM activity in each T time. During this activity, the state of the hammers is not verified and only the failed hammers will be replaced. Therefore, the maintenance decisions are only based on the operational time of the equipment and do not take account the effect of the defective hammers in the shredder productivity in the future.

3.5.2 Time-based inspection policy

A time-based inspection policy is a type of CBM policy where the equipment or system is turned off to be inspected when reaches a certain operational time. During the check procedure, some degradation parameters are measured. The set of maintenance actions to be performed depends on its current degradation states.

Regarding the shredder context, Araújo et al. (2018) suggest to scheduling periodical inspections to verify the state of the hammers. Following the inspections result, the authors recommend to replace both failed and defective hammers. The replacement of defective

components is very frequent in practical contexts, because the defective state indicates an imminent failure, and in the shredder case a reduction of its productivity rate. Araújo et al. (2018) conclude that the replacement recommendation of the defective hammers, and not only the failed components as suggested by the traditional time-based maintenance policy, has positive effects in terms of the expected cost rate in the long run. According to the authors, defect detection tends to reduce the number of failed components in an operational cycle, contributing to maintaining good levels of equipment productivity.

3.6 PROPOSED DRL MAINTENANCE POLICY

This work proposes to develop a DRL approach to find the best maintenance planning based on the system features. In the system under study, the first station processes the scrap stored in the previous stock and produces with a production rate of P the crushed scrap that feeds the intermediate buffer. Due to stoppages, breakouts and decreases in the P of the shredder, the buffer level can be found in the range $0 \leq b \leq K$. When the buffer is full ($b = K$) the first station stays idle until the buffer level decreases enough to make possible its operation. This situation raises some opportunity to perform maintenance actions in this station. When the buffer is empty ($b = 0$) or its level plus P do not meet d ($b + P < d$) the second station is said to be starved. This causes a loss of production in the form of unmet demands. The costs regarding to the unmet demands are a significant portion of the overall maintenance cost.

The focus of this study is to propose PM models for the crushing station to indicate what is the best time to perform PM actions in the shredder machine aiming to minimize the long-run maintenance cost per unit of time. The models do not define actions for the second station, but its costs, e.g., unmet demand, are computed into the analysis.

Let π denotes the PM policy for the shredder. This policy indicates when the shredder should be turned off to receive PM. Let $C(t, \pi)$ indicate all costs related to the maintenance activities until the time t when the system follows the PM π.

$$C(t, \pi) = c_l \int_0^t D(t)\, dt + c_s \int_0^t b(t)\, dt + c_r\, N_r(t) + c_i\, PM(t) + (c_f + c_i) CM(t) \quad (3.5)$$

where $\int_0^t D(t)\, dt$ and $\int_0^t b(t)\, dt$ are respectively the accumulative unmet demand and buffer level up to time t. $N_r(t)$, $PM(t)$, and $CM(t)$ are the number of hammers replaced, PM and CM performed up to time t, respectively. Thus, an optimal preventive policy π^*aims to minimize the long-run maintenance cost rate, i.e., maintenance cost per unit of time.

$$\pi^* = arg\min_{\pi}\{\lim_{t\to\infty}\frac{C(t;\pi)}{t}\} \quad (3.6)$$

In order to find the optimal policy π^*, a DRL algorithm was built. The agent must learn during the interaction with the environment when the PM action must be performed based on the system's monitoring parameters. So, through the simulation model, the agent interacts directly with de production system environment to learn the optimal policy. The learning process happens with the agent taking actions a_t in the environment that leads its current state S_t to a new state S_{t+1} and receive a reward R_{t+1} depending on the desirability of the action. Thus, the agent learns the optimal policy that maximizes the reward function.

Regarding the DRL model, two scenarios are considered. Both scenarios are proposed based on the system modeling and general assumptions describe in sections 3.2 and 3.4 and have the same maintenance goal defined in Eq. 3.6. However, each scenario presents news assumptions covering different aspects of the environment regarding the second station as detailed below.

3.6.1 Scenario 1

Scenario 1 makes the assumption that the second station does not deteriorate over time. In other words, the station that is composed for the remaining process in the steel production line after the shredder stage does not stop its operation neither for maintenance action nor for random failures, only when it is starving. In this case, d always has a fixed and positive value that does not vary over time. That is, the buffer is always pulled for a constant and fixed demand d that does not cease. This assumption was made for the policies proposed for Araújo et al. (2018) and Ferreira Neto et al. (2020a), and it also applies to the opportunistic maintenance policy proposed in this work.

The algorithm of the simulation model presented in Figure 5 already covers the environment assumption assumed in the scenario 1 along with the general assumptions and system modeling described in sections 3.2 and 3.4.

Although the focus of the study is the development of maintenance policies for the shredder and maintenance actions are not suggested for the remaining processes of the production line, the exclusion of any interruption of the second station can lead to inaccurate analysis. Therefore, a second scenario was proposed.

3.6.1 Scenario 2

In scenario 2, the second station deteriorates over time which leads to random stops. This station is composed of several interconnected operational units and any known interruptions in these units leads to a stop of the entire station for maintenance actions. During the stoppages, the station stays idle and the demand d goes to 0. Therefore, the demand d varies over time, but in a binary way. When the second station is operating, the demand d is constant and known, and when the station is inoperative it is zero. Thus, in scenario 2, d is a function of time and represents the operation state of the second station.

The cost of the maintenance actions in the second station is not computed into the model. So, the costs taking into account in the maintenance models are the same and can be computed by the Eq. 3.5. Although the second station idleness eventually leads to an interruption of the steel production, the model only considers an unmet demand when this interruption occurs due to shortages of raw material that come from the buffer. Hence, the unmet demand can be calculated using the Eq. 3.2, considering the instantaneous demand value at time t, i.e., d_t.

The following assumptions about of the system under study are made in this scenario:

1) The second station deterioration process is assumed to follow a known probability distribution F_d;
2) Any stop at the second station is considered constant, known, and equal to T_d;
3) During the stoppages, the demand d is equal to 0.
4) The demand d in the time t is d_t.

These new assumptions require a few modifications in the simulation model described in Figure 5. Let Y be the time when the second station suffers a breakdown. This time following the probability distribution F_d. So, the simulation model algorithm used in scenario 2 is presented in Figure 7.

Figure 7 - Simulation model algorithm for S*scenario 2*

Simulation model:

input parameters: $c_l,\ c_f,\ c_s, c_i, c_r, n, K, d, P_i, F_c, T_P, F_x, F_h, k, T_d, F_d$
generate values of *X* and *H* for each hammer by *Monte Carlo Method* using F_x and F_h
$b_0 \leftarrow 0$
for $t = 0,1, \dots, T$ **do**
    check the operation time of each hammer
    evaluate the condition of each hammer
    every t step, d_t is set with the return of the *Function* $d(t)$ in Figure 8
    if $N \geq k$ **do**

```
        calculate the production rate P_t
        perform buffer mass analysis, i.e., calculate b_t and D_t using Eq. (3.2)
        calculate the operational cost per Eq. (3.1)
        if preventive maintenance is trigger do
            generate new values of X and H for defective and failed hammers
            perform buffer mass analysis during T_P
            calculate the maintenance cost using Eq. (3.3)
        end if
    else
        generate new values of X and H for defective and failed hammers
        gauge the duration of the corrective maintenance by Monte Carlo Method using F_c
        perform buffer mass analysis during T_c
        calculate the maintenance cost using Eq. (3.4)
    end if
    compute the interaction cost
    store interaction cost
return cost per unit of time
```

Source: This research (2020)

The algorithm works closer to the former, but now the instantaneous value of d_t is measured every t time through a function presented in Figure 8.

Figure 8 - Function to calculate demand d

```
Function d ( ) algorithm:
input parameters: T_d, F_d, d, t
generate value of Y by Monte Carlo Method using F_d
if Y < t
    d_t ← d
else
    if Y + T_d < t
        d_t ← 0
    else
        d_t ← d
        update the value of Y
    end if
end if
return d_t
```

Source: This research (2020)

Before applying a DRL algorithm to obtain the π^* PM policy for both scenarios, the problem will first be modelled as a Markov Decision Problem (MDP), which is the most

common framework to RL techniques. Each scenario has its proper MDP formulation that covers its assumptions.

3.7 MDP FORMULATION

MDP is a discrete-time stochastic process which holds the Markov propriety that model the sequential decision making in uncertain environments (PUTERMAN, 2014). In MDP there are five components: system state s_t, decision action a_t, stochastic state transition, reward discount factor γ and instant reward function r_t. In the current study, the stochastic state transition is driven by uncertainties of the environment and by the actions, instead of having a fixed probability transition matrix.

At a discrete-time t, the system state s_t is observed, and following some rule, an action a_t is selected to be performed in the environment. In the context under study, this action is to decide if a PM should be carried out or not. After taking the action, a reward $r_{(s_t,a_t)}$ is received, which reflects how good the action choice a_t was for the state s_t. Let R_t be the accumulative reward, as described below.

$$R_t = r_t + \sum_{i=0}^{\infty} \gamma^i \, r_{(s_{t+i},a_{t+i})} \tag{3.7}$$

The discount factor γ is used to consider both the future reward and the immediate reward. It is also widely used to guarantee the convergence of the reward summation (ZHANG; SI, 2020). So, the goal is to find the π^* that maximize the expected R_t, i.e.

$$\pi^* = \arg\max_{\pi}\{E_\pi[R_t]|s = s_t\} \tag{3.8}$$

In order to formulate the MDP, the system state s_t, action a_t, and reward function $r_{(s_t,a_t)}$ should be defined for each scenario.

3.7.1 State definition

The state definition is an important step in the MDP formulation since through the observation of the state space the agent will learn the environment behavior and choose the optimal action. Thereby, the system state s_t should comprehend both the machine-level and system-level information (HUANG; CHANG; ARINEZ, 2020).

For scenario 1, the instantaneous productivity rate of the shredder P_t and the instantaneous buffer level b_t fully describe the system dynamically. Through the observation of P_t, the agent can understand the shredder's deterioration process, and observing b_t it can

cognize the dynamic between the two stations. For this reason, the state considered for scenario 1(s_t^1) is defined as:

$$s_t^1 = [P_t, b_t] \tag{3.9}$$

For scenario 2, adding the instantaneous demand d_t allows the agent to recognize when the second station is operational or not. Hence, the state considered for scenario 2 (s_t^2) is defined as:

$$s_t^2 = [P_t, b_t, d_t] \tag{3.10}$$

All the selected variable to determine the system states can be directly observed or tracked. P_t and d_t could be measured through a flowmeter placed in the downstream of the shredder and the buffer, respectively, and b_t by a level meter.

3.7.2 Action definition

The action a_t performed in the environment decides whether or not to turn off the shredder and initiate a PM. Consequently, the a_t for each scenario will be the same and is defined as:

$$a_t = \begin{cases} 0, & leave\ the\ shredder\ operating \\ 1, & turn\ off\ the\ shredder\ for\ PM \end{cases} \tag{3.11}$$

When the shredder fails, the CM is immediately triggered. This kind of maintenance activity is not under the control of the agent, as it is only a consequence of random failures.

3.7.3 Reward function definition

Since the goal of the agent is to maximize the expected cumulative reward R_t defined in Eq. 3.7, the reward function r_t should be aligned with the maintenance goal. It should be formulated as a stepwise of the objective function, so the agent can learn a satisfactory π^* to achieve the problem purpose. Thus, the r_t must include maintenance cost and operational cost related to the maintenance activity such as unmet demand and storage cost in the buffer. Because both scenarios compute the same maintenance and operational cost, the r_t is the same for both and is defined as:

$$r_t = -c_l\, D(t) - c_s b(t) - c_r\, N_r(t) - c_i\, PM(t) - (c_f + c_i) CM(t) \tag{3.12}$$

where$c_r\, N_r(t)$, $c_i\, PM(t)$, and $\left(c_f + c_i\right) CM(t)$ are the cost of the hammer's replacement, PM actions and CM actions at time t respectively, and $c_l\, D(t)$ and $c_s b(t)$ are the cost of unmet demand and storage during a time step, respectively. In order to find the optimal policy that

minimizes the overall maintenance cost, the reward function is negative. So, when the accumulative reward R_t is maximized, the long-term maintenance cost will be minimized.

3.8 DRL FRAMEWORK

Although exact approaches can handle with a serial production line with two stations and an intermediate buffer such as Dynamic Programming, there are some points that make a RL approach ideal for the maintenance problem under study. Most RL algorithms are model-free, which means that to converge to an optimal policy its learning process does not necessarily needs the system transitions probabilities (WANG; WANG; QI, 2014; HUANG; CHANG; ARINEZ, 2020). Besides, its agent training process is through the sampling of state and action space transitions coming from either experiments or simulation environments (HUANG; CHANG; ARINEZ, 2020). So, it is appropriated to be used in the simulation models developed in this work. Lastly, RL is well-suited for the determination the optimal time to maintenance action, because of its ability to manage the benefit of the long-term reward (KUHNLE; JAKUBIK; LANZA, 2018). PM increase the short-term cost, but it decreases the long-term cost by reduction of the unavailability and failure probability of the system.

A lot of algorithms have been proposed to obtain the optimal policy π^*, among which Q-learning is the most well-known RL algorithm (AISSANI; BELDJILALI; TRENTESAUX, 2009; SUTTON; BARTO, 2018). The traditional *Q*-learning method, initially formulated by Watkins (1989), is a tabular method that creates a table called *Q*-table where the optimal state-action value function *Q* is initiated, stored and updated (WATKINS, DAYAN, 1992). In this method, the expected reward R_t that an action a_t has taken in state s_t following the policy π is called *Q* -value and is presented below.

$$Q_\pi(s,a) = E_\pi[R_t|s_t = s, a_t = a] \tag{3.13}$$

The Q-values are updated after each epoch following the equation below.

$$Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left(r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a)\right) \tag{3.14}$$

The α is the step length taken to update the estimation of $Q(s,a)$. It can be understood as a learning rate. The goal now is to find the optimal π^* that maximize the *Q*-value, as described in the equation below.

$$Q_{\pi^*}(s,a) = E_{\pi^*}[R_t|s_t = s, a_t = a] \tag{3.15}$$

The *Q*-learning algorithm is shown in Figure 9. The ε-greedy is responsible for a key component of the learning process of the RL that is the balance between the exploration and exploitation phase. It decides which phase will be executed in each time step *t*.

Figure 9 - Q-learning algorithm

Q-learning algorithm:

input parameters: $\gamma,\ \varepsilon, \alpha$
Initialize $Q(s,a)$ table arbitrarily
observe s_0
for $t = 0,1,..,T$
  draw a random number $\xi \sim$ Uniform (0,1)
  if $\xi > \varepsilon$ **do**
    choose the action which has the highest Q-value in the Q-table, i.e., $a_t = \arg\max_a Q(s,a)$
  else
    choose any action a_t at random
  end if
  take action a_t, observe r_t and s_{t+1}
  $Q(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha\left(r_{t+1} + \gamma \max_a Q(s_{t+1},a) - Q(s_t,a)\right)$
return $Q(s,a)$

Source: This research (2020)

This method is widely used due to the simplicity of its formulation, robustness, and the ease in which parameters can be adjusted (SUTTON; BARTO, 2018). However, this tabular method ends up using a large table to record the *Q*-values. The problem is not only its lack of scalability to high-dimensional problems but the time and data required to fill the table accurately (HUANG; CHANG; ARINEZ, 2020; ZHANG; SI, 2020). To overcome these challenges, DRL algorithms have been developed during recent years. The DRL algorithms combine the RL with deep neural network to provide a powerful approximation of the *Q*-value, which significantly reduces computer efforts (ROCHETTA et al., 2019). These algorithms train a neural network θ to approximate the *Q*-values, i.e.

$$Q_{\pi^*}(s,a,\theta) \approx Q_{\pi^*}(s,a) \tag{3.16}$$

In order to obtain the optimal π^*, the neural network parameter θ is updated with the interaction, instead of filling the *Q* table. The DRL variant most used of the *Q*-learning is the Deep *Q* Learning (DQN). The DQN training process uses a single neural network to provide an approximation of the *Q*-value (MNIH et al., 2016). However, it has presented an over-estimation of its approximation. The Double Deep *Q* Learning (DDQN) introduced for Hasselt

et al. (2016) relieve this effect by using two neural networks: the online network θ and the target network θ^-. The θ is used to select the action and the θ^- to evaluate the policy π.

In this work, a DDQN algorithm was used to solve the problem under study by obtaining the optimal policy π^* that suggests whether or not maintenance action should be performed in the shredder at each time by solving the MDP problem for each scenario. The DDQN algorithm used is detailed in Figure 10.

Figure 10 - DDQN algorithm

DDQN Algorithm:

input parameters: $\gamma,\ \varepsilon, C, N_{mem}, interaction\ time, m$
build a neural network θ
create a replay memory M with capacity N_{mem}
for $t = 0,1,..,T$
  every C steps, set $\theta^- \leftarrow \theta$
  while $t < interaction\ time$ **do**
    draw a random number $\xi \sim$ Uniform (0,1)
    if $\xi > \varepsilon$ **do**
      select $a_t = \arg\max_a Q(s, a, \theta)$
    else
      choose any action a_t at random
    end if
    input action a_t into the simulation model
    run the simulation model for one-time step
    observe s_t
    calculate r_t
    observe s_{t+1}
    store transition sample (s_t, a_t, r_t, s_{t+1}) in replay memory M
  end while
  sample a minibatch of size m of transitions (s_j, a_j, r_j, s_{j+1}) from M
  set $y_j = r_j + \gamma Q\left(s_{j+1}, arg\max_{a_{j+1}} Q(s_{j+1}, a, \theta); \theta^-\right)$
  perform a gradient descent step on $\left(y_j - Q(s_j, a_j, \theta)\right)^2$
return θ

Source: This research (2020)

The two important concepts used in DDQN are the experience replay memory M and the target network θ^-. The experience replay memory M is a set of past experiences that works as a data set for the θ training. During the training process, new data of the system are added in M and sample mini-batches with size m are randomly withdrawn for the neural network training.

It helps to stabilize the learning process by reducing correlation among the training data and enhances the training performance by enabling a batch training (ZHANG; SI, 2020). About the target network, while the neural network θ is updated with gradient descent, the target network θ^- stays fixed and after *C* steps is updated as a copy of the θ (HASSELT et al., 2016). Let θ_j be the neural network θ at time *j*, the updated neural network of the next time step θ_{j+1} is obtained through the gradient descent process described below.

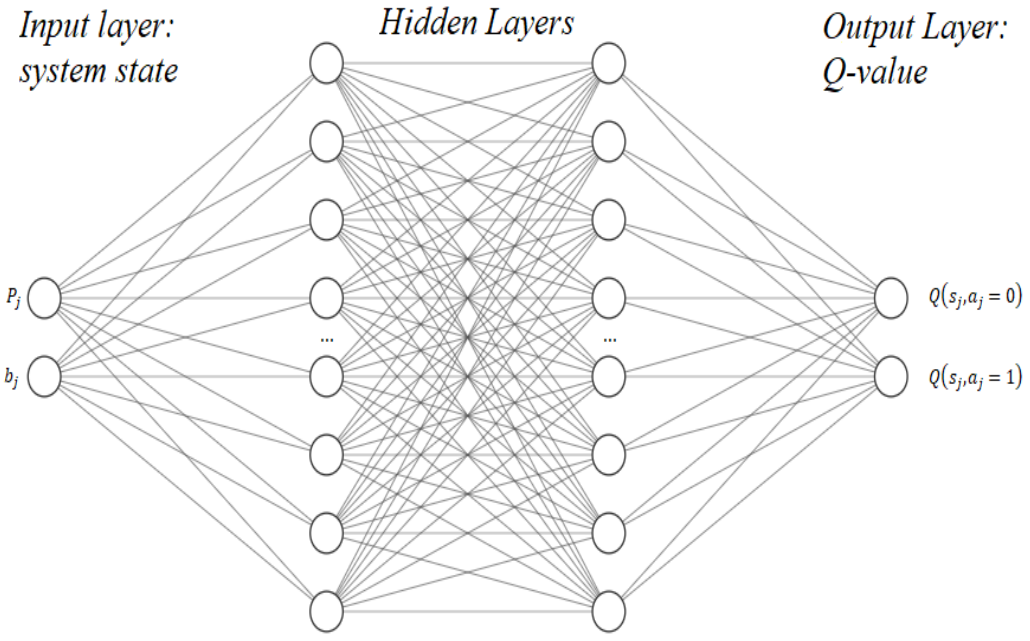
$$\theta_{j+1} = \theta_j + \alpha\left(y_j - Q(s_j, a_j, \theta_j)\right)\nabla_{\theta_j} Q(s_j, a_j, \theta_j) \tag{3.17}$$

The delay between θ and θ^- helps to overcome the overestimation problem (HUANG; CHANG; ARINEZ, 2020). The θ^- is also used to generate the targets y_j that going to be used to update the θ at the time *j*.

$$y_j = Q(s_j, a_j) + \gamma Q\left(s_{j+1}, arg\max_{a_{j+1}} Q(s_{j+1}, a_{j+1}, \theta_j); \theta_j^-\right) \tag{3.18}$$

The neural network architecture used is composed of several fully connected layers as illustrated in Figure 11. The input layer receives the system state and the output layer convey the approximate *Q*-value of each maintenance action.

Figure 11 - Neural Network architecture



Source: This research (2020)

4 CASE STUDY

Intending to analyse the performance and behavior of the proposed PM policies, a numerical case study was conducted. Two steps are required, that is to obtain the value of the parameters of the system and the hyperparameters of the neural network used in the DDQN algorithm.

The operational and maintenance parameters of the system are shown in Table 2. Despite the PM policies be inspired in a real Brazilian steel production line, the system parameters described in this section are only illustrative. However, the chosen values keep the proportion found in the real life.

Regarding the degradation process of the hammers, a Weibull distribution with scale and shape parameter $\eta_1 = 120$ and $\beta_1 = 3$ was used to describe the time until the defect arrival probability distribution. Another Weibull distribution with scale and shape parameters $\eta_2 = 30$ and $\beta_2 = 1$ was used to describe the delay time probability distribution. To indicate when an interruption of the second station occurs, an Exponential distribution with parameter $\lambda = 1/168$ was considered, that is a mean of 168 hours, i.e., one week. The costs have been defined without a monetary term, but using a unit of reference (un.). Regarding to the distribution probability of extra time of the CM duration, a Weibull distribution with scale and shape parameter $\eta_3 = 7$ and $\beta_1 = 3$ were defined.

Table 2 - System parameters

System parameters		
Parameter	Value	Unit
n	20	component
k	10	component
K	1000	ton
P_i	1	ton/h
d	15	ton/h
c_i	50	un.
c_f	3000	un.
c_l	100	un. /ton
c_r	230	un. /component
c_s	0.01	un. /ton.h
T_P	10	hours
T_d	12	hours

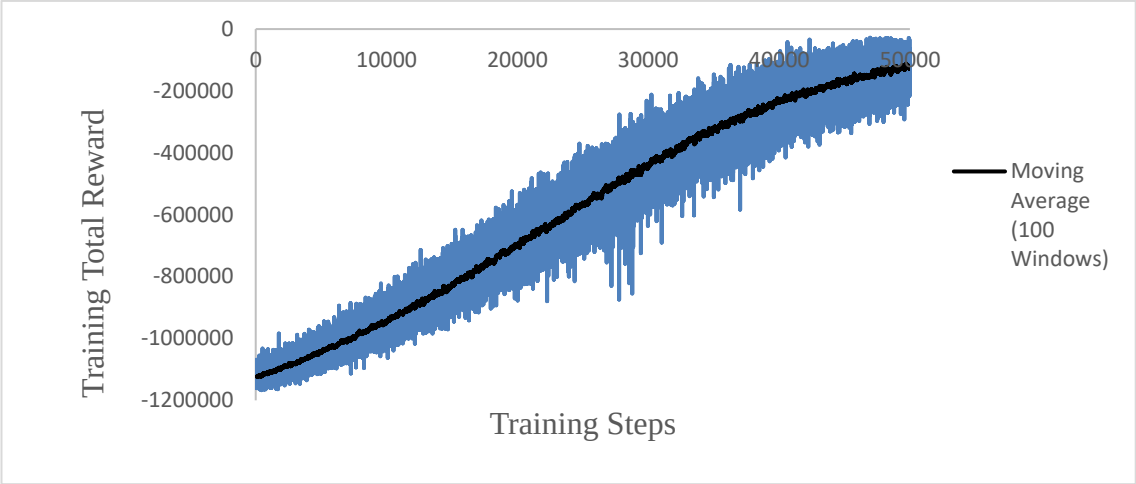
Source: This research (2020)

The architecture of the neural network was almost the same for both scenarios analysed. It has three fully connected hidden layers, the first has 1200 neurons, the second 300 neurons and the third has 100 neurons. The size of the input layer and the output layer are equal to the size of the system states and actions state, respectively. For scenario 1, it has 2 neurons in both the input and output layer, and in scenario 2 it has 3 neurons in the input layer and 2 neurons in the output layer. The replay memory used has a capacity $N_{mem} = 2000$ interacitons, and batch size used to train the neural network is $m = 150$. The target neural network θ^{-} is updated with a copy of the θ after $C = 10$ steps, and the discount factor used was γ =0.9. The ε-greedy starts set to be 1 but is reduced through a decay rate of 0.99994 until 0.05 where it stays fixed until the end of the training process. Regarding the gradient descent, the optimizer Adam (KINGMA et al., 2014) was used. Both the simulation models and the DDQN algorithm was implemented in Python. The neural network was implemented using the open-source library *Keras* (GULLI; PAL, 2017).

4.1 OBTAINING THE PROPOSED DRL MAINTENANCE POLICY

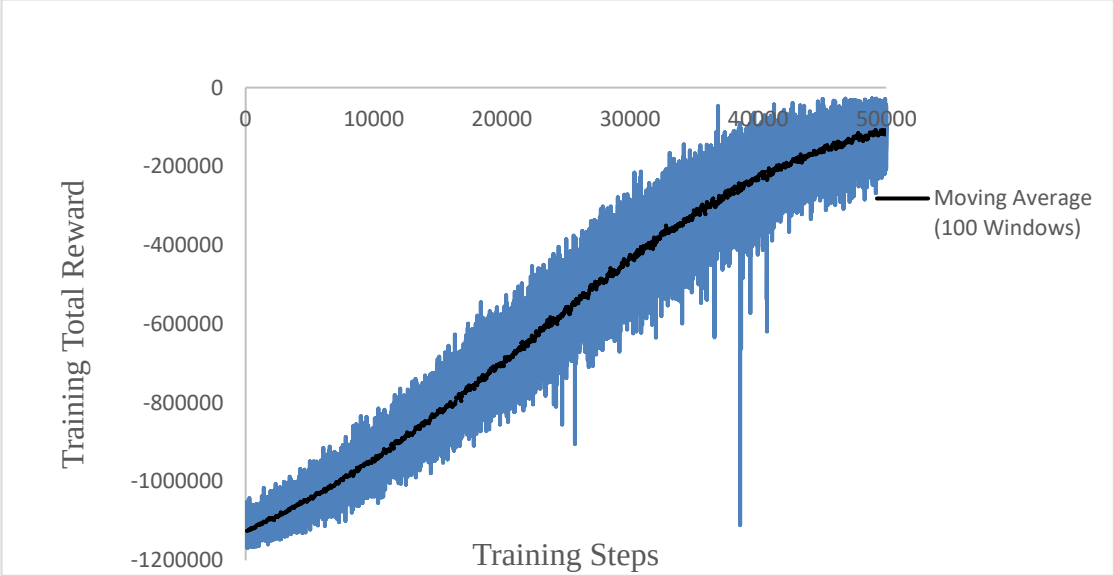
In order to obtain one PM policy, first, the appropriate simulation model of the steel production line was used to perform an offline training of the neural network parameters θ_1, for scenario 1, and θ_2, for scenario 2, using the DDQN algorithm described in Figure 10. The training process of both scenarios ran 50000 interactions, each interaction contained 1000-time steps, and lasted approximately 40 hours each. The behavior of the cumulative reward collected by the agent during each interaction in the training process for both scenarios are shown in Figure 12 and Figure 13.

Figure 12- Training accumulative reward of scenario 1



Source: This research (2020)

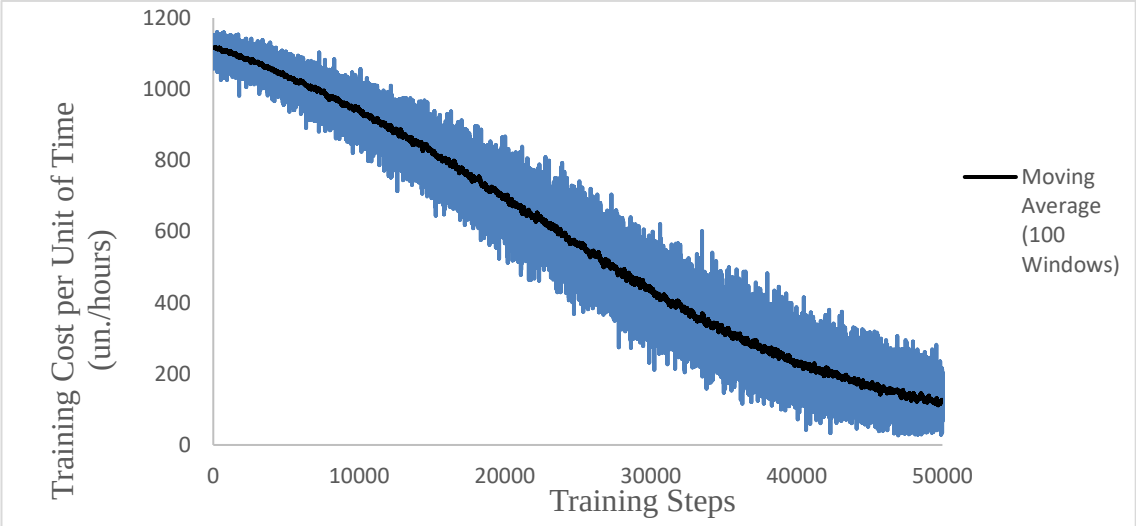
Figure 13- Training accumulative reward of scenario 2



Source: This research (2020)

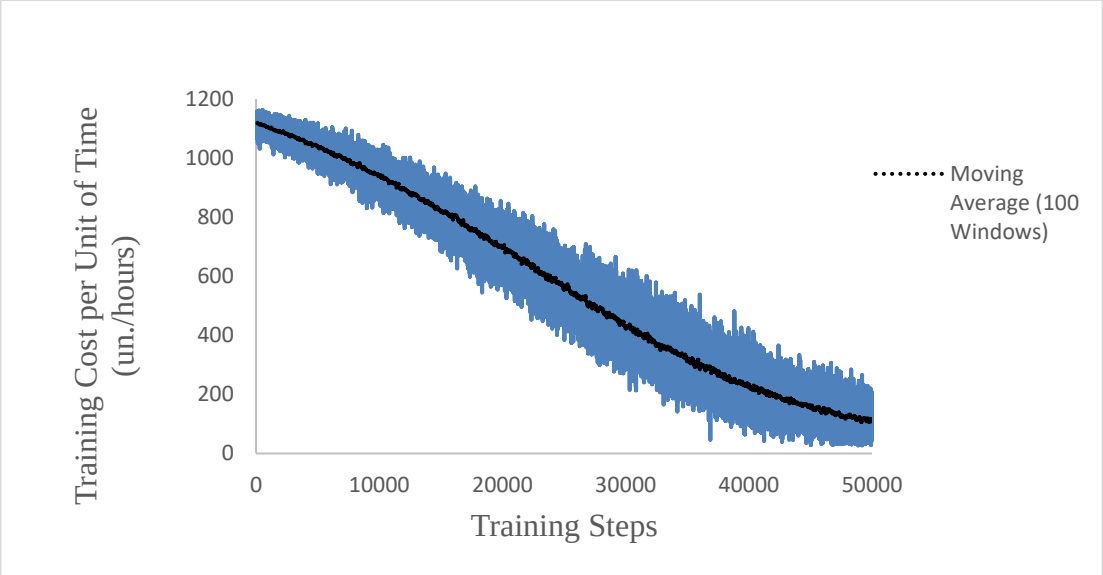
Although the training reward has some peaks and valleys, the accumulative reward shows some convergence, which means that the agent is learning an optimal PM policy. The cost per unit of time during each interaction of the training process is shown in Figure 14 and 15.

Figure 14 - Cost per unit of time during the training process of scenario 1



Source: This research (2020)

Figure 15- Cost per unit of time during the training process of scenario 2



Source: This research (2020)

The DRL-based PM policy for the shredder is obtained through the trained neural networks θ which are ready to be used in online maintenance decision making, i.e., to be applied in the real shredder system. This trained neural network θ is actually the DRL maintenance policy, and the procedure for its usage in real life is detailed in Fig 16.

Figure 16 - Real life procedure

Real Life Procedure:

1: **input parameters:** real-time states s_t
2: observe s_t
3: run a forward propagation in neural network θ to get $Q(s_t, a; \theta)$
4: find the optimal $a_t = \arg\max_a Q(s_t, a, \theta)$
5: **return** PM decision a_t

Source: The Author (2020)

Note that now there is no fixed time to perform PM action. Instead, the maintenance decision will be taken according to the real-time condition monitoring of the steel production line. At each time *t*, the neural network θ will receive as input the system state which represents its condition monitoring. Based on its values, the maintenance decision is taken. Thus, do not have a scheduled time for performing PM action, and the decision of intervene in the system is made in real-time.

Although the training process of the neural network θ is quite burdensome computation, the decision-making process through the proposed PM policies uses only a forward signal propagation in the neural network θ. Hence, it could be applied in real life context.

4.2 RESULTS AND ANALYSIS

In order to evaluate its performance, the proposed maintenance models were compared with three different policies, including the time-based maintenance policy, and time-based inspection policy, all presented in section 3.5. The third policy used in the comparison was a CM policy where the system runs until its failure. No PM actions are carried out, only CM actions. During the CM activity, both defective and failed hammers must be replaced. This policy is used as a baseline to measure the benefits of applying any PM policy in the system.

To optimize the decision variables of the opportunistic policy and the PM policies used in the comparison and to find the maintenance policies that lead to a good performance of the line, the simulation model described in Figure 5 was defined as an objective function and the minimization method Differential Evolution (STORN; PRICE, 1997) was applied. Specially for the opportunistic policy, the minimization method was applied into the algorithm described in Figure 6. The maintenance policies obtained through the optimization are shown in Table 3.

Table 3 - Optimal maintenance policies

Optimal Maintenance Policies			
	Time-based maintenance policy	Time-based inspection policy	Opportunistic Inspections Policy
Decision variables	T = 49.53 hours	T = 34.61 hours	T = 46.15 hours, t_{min} = 12.30 hours, b_{min}= 227.24 ton.

Source: This research (2020)

To measure the performance of the proposed DRL-based PM policy, the trained neural network θ was applied in the real-life procedure described in Figure 16, considering the assumptions of scenario 1 and using the simulation model to simulate the behavior and uncertainties of the environment in study. Different from the previous policies, the proposed DRL-based policy does not have a fixed/scheduled time for PM action. Instead, the trained neural network receives the monitored states of the system as inputs and takes the maintenance decisions in real-time, intervening in the system based on its monitoring conditions. The results are shown in Table 4.

Table 4 - Policies comparison

Policies comparison					
	Time-based maintenance policy	Time-based inspection policy	Opportunistic Policy	CM Policy	Proposed DRL Policy Scenario 1
Cost per unit of time	90.9963 un. /hours	33.9732 un. /hours	31.2646 un. /hours	106.7412 un. /hours	29.5718 un. /hours
Unmet demand per unit of time	0.6314 ton. /hours	0.0166 ton. /hours	0.0012 ton. /hours	0.5545 ton. /hours	0.0 ton. /hours
Shredder availability	0.8261	0.7758	0.7895	0.8646	0.8202
CM per unit of time	0.00090 CM/hours	0.0 CM/hours	0.0 CM/hours	0.0080 CM/hours	0.0 CM/hours

Source: This research (2020)

It is important to keep in mind that both the CM policy and the DRL policy do not have a time T as a decision variable. In corrective maintenance strategy, the maintenance action is performed only when the system fails, which means that the decision-maker keeps the system running until it breaks and then performs CM actions. And in DRL policy, maintenance decision-making is an online decision. The variables that represent the condition of the system are tracked by online monitoring and based on their instantaneous values the policy suggests that the PM's action should be performed or not.

Before doing a numerical analysis of the benefit of adopting the DRL policy, it is important to highlight another advantage. Unlike the other analyzed approaches that use the system data only to optimize the maintenance policy, the DRL method is constantly observing the data and making decisions based on it. This gives the policy more flexibility to deal with changes in the environment or modification of processes. This is aligned with the recent paradigm of industry 4.0, where to meet the trend of mass customization frequent changes on the operation parameters are likely to occur. Since the DRL policy learns about system dynamics rather than simple prior planning, when a change occurs, the agent can handle it by making an online decision and minimizing any impact on overall system performance and productivity.

Regarding the numerical comparison among the policies, the results found show that all four PM policies outperform the CM strategy in terms of the expected long-term maintenance cost per unit of time, which means that they are all effective PM policies. Looking into the comparison between the time-based policies, some important points can be observed. First, time-based inspection policy shows a better performance than time-based maintenance policy, confirming that including the replacement of the defective hammers into the maintenance activities, which can increase the short-term cost, has a positive effect on the expected long-term cost rate. The reason is that identifying and replacing defective hammers tends to reduce the average number of hammers that fails during shredder operations, which helps to keep productivity at a good level, reducing unmet demands and preventing random failures.

It is an important finding because time-based maintenance policy is the current practice of the company that inspired this study. So, the replacement of the defectives hammers can reduce approximately 62.3% of the expected maintenance cost rate, along with a reduction in the expected unmet demand per unit of time and the shredder failure rate. Therefore, this

highlights that the maintenance activities to be performed during the intervention are just as important as when the maintenance activities are scheduled to be performed.

Looking at the suggested time for the execution of the maintenance actions, the time-based inspection policy anticipates the PM intervention compared to the time-based maintenance policy. This is expected because the system will be inspected earlier to identify defective hammers and therefore to prevent failures.

The opportunistic policy works better than the time-based inspection policy. Although the policy has a scheduled time to perform PM actions, this can be anticipated based on the monitored variable. The monitored variable chosen was the buffer level, so the policy monitors it online and, based on its instantaneous value, chooses whether the inspection should be anticipated or not. Thus, taking the buffer level into consideration to anticipate inspections, the policy can capture some aspects of the dynamic of the system, which leads to a better performance than policies only based in operational time. When the policy suggests that the inspection should be anticipated, this event is called an opportunity window and represents an opportunity to gain more benefits if the decision maker chooses to anticipate the inspection.

About the scheduled time for the periodic inspection in the opportunistic policy, it is closer to the time-based maintenance policy. Due to the opportunity window that allows anticipation of the maintenance activities when recognizing a better time to do so, the periodic inspections do not need to be taken early as happens in the time-based inspection policy. This characteristic leads to a reduction of 7.98% in comparison with time-based inspection policy, and 65.64% in comparison with time-based maintenance policy.

Looking at the proposed DRL-based maintenance policy applied into the scenario 1, it has shown the best results of all policies analyzed. By the system monitoring, the agent can learn about the behavior of the system and intervene in it in real-time, giving suggestions about the best time to act, aiming to achieve the lowest long-run maintenance cost. As result, a reduction in the expected maintenance cost per unit of time in comparison of all other PM policies was acquired, and the found values were: 67.5% in comparison with time-based maintenance policy, the current company policy; 13% in comparison with time-based inspection policy, and 5.41% in comparison with the opportunistic policy.

As the unmet demand and the failure event are highly expensive, the proposed PM policy makes a huge effort to avoid this situation. The unmet demand is strongly related to the production line availability. In scenario 1, when an unmet demand occurs, it means the second

station stays idle. It leads to a shortage of final products for the clients, bringing several negatives impacts for the company.

Still discussing the unmet demand, an interesting point can be observed. Although there is a reduction in cost, using a CM strategy shows better performance than time-based maintenance policy in terms of unmet demand. It demonstrates that a maintenance policy should be adopted regarding the system performance indicator that wants to be enhanced and emphasizes the importance of choosing the right maintenance strategy to be adopted.

The shredder availability means the percentage of the operational time that the equipment has worked. Although the time-based inspection policy and the opportunistic strategy has shown significant cost savings compared to the CM strategy and the company's current strategy (time-based maintenance policy), these strategies have shown a reduction in the equipment availability. It means that the cost-saving was acquired by increasing PM actions in order to identify the defects in an initial stage and perform preventive substitution of the defective and failed hammers. In the DRL model, the cost-saving was achieved without the need for too much PM activities. It is the consequence of the policy's ability to intervene in real-time. By the system monitoring, the policy can obtain a better use of the equipment without letting it fail.

In order to observe the behavior of the DRL-based maintenance policy applied in scenario l for different cases in which there is variation in some input parameters, a sensitivity analysis was performed. The results are presented in Table 5, in which case 1 represents the reference case with the parameter values adopted in the case study. In the other rows, the results for the other cases are presented and the parameters with values altered are highlighted in grey.

Table 5 - Sensitivity analysis for the DRL-based policy applied in the scenario 1

	Input parameters														**Proposed DRL-based policy**
Case	T_p *(h)*	η_1 *(h)*	η_2 *(h))*	η_3 *(h)*	*d (ton/h)*	P_i *(ton/h)*	*n*	*k*	*K (un.)*	C_l *(un./ton)*	C_f *(un.)*	C_i *(un.)*	C_r *(un./ component)*	C_s *(un./to n.h)*	C_∞ *(un./h)*
1	10	120	30	7	15	1	20	10	1000	100	3000	50	230	0.01	29.57180
2	10	120	30	7	15	1	20	10	1000	100	3000	50	230	0.001	25.50350
3	10	120	30	7	15	1	20	10	1000	100	3000	50	230	0.1	37.26747
4	10	120	30	7	15	1	20	10	1000	100	3000	50	130	0.01	14.64471
5	10	120	30	7	15	1	20	10	1000	100	3000	50	330	0.01	35.28628
6	10	120	30	7	15	1	20	10	1000	100	3000	25	230	0.01	26.22503
7	10	120	30	7	15	1	20	10	1000	100	3000	100	230	0.01	30.47137

8	10	120	30	7	15	1	20	10	1000	10	3000	50	230	0.01	29.32967
9	10	120	30	7	15	1	20	10	1000	1000	3000	50	230	0.01	30.01399
10	10	120	30	7	13	1	20	10	1000	100	3000	50	230	0.01	24.29561
11	10	120	30	7	17	1	20	10	1000	100	3000	50	230	0.01	132.86285
12	10	120	30	4	15	1	20	10	1000	100	3000	50	230	0.01	29.11888
13	10	120	30	10	15	1	20	10	1000	100	3000	50	230	0.01	29.79620
14	8	120	30	7	15	1	20	10	1000	100	3000	50	230	0.01	27.27031
15	12	120	30	7	15	1	20	10	1000	100	3000	50	230	0.01	30.14386

Source: This research (2020)

In each row, the neural network θ_1 was trained using the new parameters of the system, keeping the neural network architecture and the hyperparameters used in the reference case unaltered. Analyzing the results obtained from the sensitivity analysis, it is found that DRL-based policy behaves as expected.

From the sensitivity analysis, it could be noticed that variations in C_s generate impacts on the maintenance policy performance. Reductions on C_s make the scrap storage more appealing. The agent can store more crushed scrap without additional cost, which helps to supply the production line efficiently when the shredder productivity is reduced or during the system shutdowns. Hence, the PM actions do not need to be taken often, reducing the downtime and contributing to a smaller C_∞ (case 2). On the contrary, increasing the C_s makes scrap storage very expensive. Thus, storing less scrap becomes more advantageous. Therefore, the system is interrupted more frequently for PM actions, resulting in a higher C_∞ (case 3).

During the maintenance interventions, both defected and failed hammers are replaced. When C_r is decreased, the replacement of the hammers can be intensified, avoiding drawbacks such as shredder productivity reduction and failures with no significant extra cost associated. The intensification of the hammer's substitutions can lead to an identification of the anomalies in the hammer in their initial stage, bringing, even more, the benefit of the preventive strategies. This contributes to preventing the occurrence of failures and unmet demands, reducing the C_∞ (case 4). In the opposite sense, when C_r is increased, the better use of the component lifetime is more attractive because its replacement is more expensive. For that, the frequency of the PM halts needs to be reduced. To bear with this, more scrap should be stored. As result, C_∞rises (case 5).

Variations in C_i do not alter dramatically the agent behavior. As a unit cost that is taken into account once the system is turned off, small variations in C_i are not able to make interventions more attractive but it impacts the expected long-term cost (cases 6-7).

The unmet demand and the failure event are very critical for the system, promoting a high influence on the value of C_∞. Thus, these scenarios are immensely unwanted and, consequently, the agent learns to refrain from them, making sure that they will not occur. Hence, fluctuation on C_l (cases 8-9) and η_3 (cases 12-13), which is strictly related to the average duration of CM actions, did not exert a notable effect on policy performance.

As expected, the difference between the shredder production rate and the demand dictates the behavior of the agent. It means the buffer storage rate ($P - d$). Regarding the system, the higher this rate is, the faster the buffer fills and the better the production line performs. As a result, C_∞ is lower (case 10). And the inverse occurs when this rate decreases (case 11). Analyzing the variation of the storage rate by assuming different values of d, when d increases (case 11), filling the buffer becomes more time-consuming. Also, more scrap is required to keep the production line working when the system is shutdown. Besides that, a small number of failed hammers are necessary to reduce the shredder productivity below d. As a consequence, the possibility of unmet demands gets higher, and the production line performance decline.

However, when d decreases (case 10), filling the buffer becomes easier while there is a lower demand to supply. In this scenario, the agent can perform PM actions more frequently, which increases the reliability of the shredder, stores less scrap, and reduces storage costs, reducing C_∞.

Finally, reducing T_P (case 14) means an enhancement of maintenance efficiency. The agent can inspect more often and reduces scrap stock, which provides lower C_∞. But when T_P increases (case 15), the agent needs to store more scrap, increasing C_∞.

Table 6 compares the proposed DRL-based policy applied in scenario 1 with other PM policies along all the cases exhibited in the sensitivity analysis. The comparison was made by searching the optimum for each policy in all cases analyzed in Table 5. In this study, an analysis of C_∞ is the most interesting, as it can compare the economic benefits of these competing policies. The far-right column in table 6 shows the expected savings with the adoption of the proposed DRL approach instead of the opportunistic policy, which is the policy that presented the best results among the analyzed PM policies.

Table 6 - Policies performance comparison

	Optimized policy									
	Opportunistic Policy				**Time-based inspection policy**		**Time-based maintenance policy**		**Proposed DRL-based policy**	**Economy**
Case	T *(h)*	t_{min} *(h)*	b_{min} *(ton.)*	C_∞ *(un./h)*	T *(h)*	C_∞ *(un./h)*	T *(h)*	C_∞ *(un./h)*	C_∞ *(un./h)*	*%*
1	46.15	12.23	227.24	31.26	34.60	33.97	49.53	90.25	29.57	5.41%
2	46.25	26.56	278.07	29.83	39.77	30.50	50.80	89.34	25.50	14.51%
3	42.63	11.88	183.44	42.11	33.96	47.31	46.99	98.11	37.27	11.50%
4	42.41	7.88	218.20	18.77	34.64	20.43	48.74	79.59	14.64	21.98%
5	46.44	22.79	236.66	43.48	35.00	46.00	50.17	99.55	35.29	18.85%
6	41.96	3.40	201.04	30.67	33.93	32.61	48.93	89.55	26.23	14.48%
7	46.90	15.77	233.37	32.28	35.44	34.45	50.40	90.46	30.47	5.60%
8	49.68	26.80	170.72	30.83	35.15	31.93	50.28	34.09	29.33	4.85%
9	40.41	1.01	248.23	31.64	34.53	36.95	48.40	654.74	30.01	5.13%
10	41.57	16.14	131.47	28.22	29.34	33.85	36.33	32.53	24.30	13.92%
11	63.58	10.52	679.45	155.81	59.33	157.14	60.44	268.33	132.86	14.73%
12	47.10	6.81	241.88	30.99	35.15	33.18	53.05	85.92	29.12	6.04%
13	45.51	17.29	216.74	31.27	34.97	34.40	49.18	93.53	29.80	4.73%
14	42.96	11.99	153.40	30.42	31.47	33.62	46.16	40.53	27.27	10.35%
15	50.98	16.57	334.19	31.60	47.43	34.93	54.10	133.75	30.14	4.62%

Source: This research (2020)

Note that proposed DRL policy provides smaller C_∞ for all cases analyzed. It outperforms the opportunistic policy, especially when there is variation in the replacement and storage price, and in the dynamic of the environment represented by the storage rate. The saving can reach 21.98%, which is close to the maximum expected cost reduction of 25% stated by Stricker et al. (2018). Due to the ability of the agent to make maintenance decisions in real-time based on the system monitoring conditions.

Another point is that the opportunistic policy performed better than the time-based inspection policy, which in turn outperformed the time-based maintenance policy in all cases analyzed. This can be interpreted as an evolution of the PM strategies available in the context of the shredder, in which the DRL approach is the most advanced.

Regarding scenario 2, considering that the second station stoppages create some opportunities. The stoppages have a constant and known duration, not necessarily generating

unmet demand. In scenario 2, like scenario 1, an unmet demand is considered only when the production rate of the shredder together with the buffer content is not sufficient to fully supply the second demand.

The second station stoppages, considered in scenario 2, represent events that require a decision for the agent. The agent must choose, according to the long-term expected reward, between the two options. The first is to take advantage of the stoppage to schedule a shredder inspection and increase system reliability. And the second is to keep the system running to accumulate more scrap to address the reduced productivity of shredder and shutdowns.

Either way, the performance of the production line can be improved. This was confirmed through the C_∞ obtained in the case study, which was 26,8591 units. / hours. This value represents a 9.17% reduction in comparison to the C_∞ obtained in scenario 1. The improvement in the maintenance policy performance depends on the frequency and duration of the second station stoppages as demonstrated in Table 7.

Table 7 - Sensitivity analysis DRL-based policy for scenario 2

	Proposed DRL-based policy for scenario 2		
case	T_d	λ	C_∞ *(un./h)*
1	12	1/168	26.8591
2	6	1/168	27.9396
3	18	1/168	24.1398
4	12	1/24	19.5794
5	12	1/336	28.6982

Source: This research (2020)

As expected, considering the same costs of scenario 1 and the unmet demands only happen when station one stops supplying station two, the longer the production line stays stopped, the lower the C_∞ (case 2). That is because this downtime can be considered a window of opportunity for the agent. The longer the window, the more advantages can be taken, either by performing more PM actions or filling the buffer, hence the greater the savings (case 2). On the other hand, the shorter the window, the performance of the model in scenario 2 approaches the results obtained when applied in scenario 1 (case 3).

Similar behavior is expected for variations in λ. For smaller values of λ (case 5), second station failures are less frequent. Consequently, the windows of opportunity will be scarce. On the opposite, the station will be stopped very often (case 4).

Another analysis that could be done is to evaluate the performance of the proposed DRL-based policy for scenario 2 with the results found in scenario 1. Again, the comparison between the scenarios assesses the C_∞ for each case analyzed in Table 5. The analysis can be found in Table 8, where the Reduction column shows the savings in a percentage of scenario 2 when compared to scenario 1.

Table 8 - Comparison between scenarios

	Scenario 1	**Scenario 2**	
Case	C_∞ *(un./h)*	C_∞ *(un./h)*	Reduction
1	29.57180	26.85914	9.17%
2	25.50350	22.43941	12.01%
3	37.26747	38.89481	-4.37%
4	14.64471	14.09188	3.77%
5	35.28628	33.98269	3.69%
6	26.22503	23.29694	11.17%
7	30.47137	27.55145	9.58%
8	29.32967	26.80519	8.61%
9	30.01399	26.88811	10.41%
10	24.29561	22.64059	6.81%
11	132.86285	28.67133	78.42%
12	29.11888	26.40559	9.32%
13	29.79620	26.86813	9.83%
14	27.27031	26.49750	2.83%
15	30.14386	26.91908	10.70%

Source: This research (2020)

An interesting point can be observed in case 3, when the storage cost is high. In this case, the DRL-based policy applied in scenario 1 has presented a better performance than policy applied in scenario 2. It happens because during the window of opportunity that emerges when the second station stops, either when performing a PM action or when letting the system work, the buffer is not required to supply any demand. Therefore, the accumulated quantity of crushed

scrap in the long run increases, hence the C_∞ too. Except for this case, in all others, the maintenance model applied in scenario 2 outperforms the model of scenario 1, especially in case 11 where the saving reaches 78.49%. It has shown that consider scenario 2 is more indicated when the difference between the shredder productivity and the demand is small.

5 CONCLUSIONS

With the growing availability of system data that comes from the sensors distributed into the modern production systems, some challenges arise in the management and control of the production. In maintenance management, the question to be answered is how to explore the system data potential to better plan the maintenance activities and enhance the availability and reliability of the system, at the same time reducing the maintenance cost. While the large amount of data represents some opportunities to improve the system performance, it creates a high dimension space state problem that cannot be solved with the traditional maintenance strategies. In this context, emerging AI and ML tools, such as DRL, have proved to be efficient in dealing with dynamic environment subject to uncertainties, such as a serial production line, and have the ability to solve problems with a high dimension that results from these types of environments.

Therefore, this dissertation proposes maintenance policies that, through system monitoring, suggest the best time to perform PM activities to minimize the long-term maintenance cost rate. The context under study was a steel production line. The proposed PM policies focus on the shredder, machine which impacts the entire production line, and monitor some system conditions such as the shredder productivity, the buffer level, and the production line demand. A simulation model was built to simulate the system's dynamic and thus be used in the development of the PM policies. To assess the performance of the proposed policies and their behavior, comparisons were made with other maintenance policies used in the same context, and sensitivity analyses were carried out.

To summarise, the proposed DRL approach outperforms all other PM policies available in the shredder context. It has shown cost reduction without unmet demand and failure event. By understanding the dynamics of the system, the proposed policy intervenes in the system in real-time when necessary based on their real-time states. Hence, the frequency of PM interventions is maximized, resulting in an increase in the equipment availability without failure events and unmet demands along with a reduction in cost. Besides, the proposed DRL policy are easy to be used, since the monitored variables are simple to be observed and tackled.

Regarding the scenarios analyzed, in scenario 1 a DRL methodology in the shredder context was used and its result was compared with the already existed PM policies, using the same assumptions of the latter. As result, the proposed DRL-based policy has shown the best

performance among all evaluated policies. The reduction in the expected long-run maintenance cost per unit of time was up to 67.5%. Besides, other benefits have been observed. The availability of the shredder and the production line, which could be measured through the quantity of the unmet demand, was enhanced.

About scenario 2, it aims to consider more realistic assumptions about the environment under study. Considering the breakdown of the production line, the agent can use this opportunity to either increase the system reliability by performing more PM actions or store more scrap to bear with shredder productivity reduction and shutdown, which helps to increase the system performance.

Thus, the results found in both models validate the use of the DRL methodology in a real-life context, such as production lines, and the proposed opportunistic strategy that considers a system monitoring parameter to anticipate maintenance actions to improve the performance of the system.

In conclusion, the proposed DRL approach has shown cost reduction with an increase in system reliability and availability. Therefore, integrating emergent tools from AI and ML areas, such as DRL, can help maintenance management and provide competitiveness for the company.

5.1 SUGGESTIONS FOR FUTURE RESEARCH

To deepen what was studied by this work, several ideas for future research can be explored. Concerning the system under study, a good point to be analyzed would be to consider the costs associated with the second station. In the present work, the maintenance cost and the unmet demand due to the stoppages at this station, which represent the remaining steel production processes, are not computed in the simulation model. However, to get closer to reality and provide more realistic results for the decision-maker, these aspects must be considered. Besides, considering them in addition to providing more realistic characteristics, it would be interesting to assess how the model would learn to intervene in the system with the new behavior.

About the AI and ML tools, some points appear here. Although the DRL algorithm used has shown good results, there are no instructions in the literature on which ML algorithm to use for a particular need. Thus, testing other algorithms for the under study context and evaluating their performance would be interesting. Another point is to combine DRL with some ML

forecasting tools. The forecasting tool can be used to predict the machine's remaining useful life and reliability in real-time. Integrating this into the proposed PM policy can lead to better results. Finally, the performance of the maintenance policy is strongly influenced by the structure of the neural network and its hypermeters. Different combinations can be tested, instead of using the same for both scenarios.

REFERENCES

AISSANI, N.; BELDJILALI, B.; TRENTESAUX, D. Dynamic scheduling of maintenance tasks in the petroleum industry: a reinforcement approach. Engineering Applications of Artificial Intelligence, v. 22, n. 7, 2009.

ALPAYDIN, E. Introduction to machine learning (2nd ed.). Cambridge, MA: MIT Press, 2010.

ALMEIDA, A. T.; SOUZA, F. M. C. Gestão da Manutenção – Na Direção da Competitividade. Ed. Universitária, UFPE, 2001.

AMEEN, W.; ALKAHTANI, M.; MOHAMMED, MK.; ABDULHAMEED, O.; EL-TAMINI, AM. Investigation of the effect of buffer storage capacity and repair rate on production line efficiency. Journal of Kind Saud University –Engineering Sciences, v. 30, p. 243-249, 2018.

ARAÚJO, L. H. C.; FERREIRA NETO, W. A.; LIMA, H. B. V.; CAVALCANTE, C.A. Modelo de manutenção para um sistema multicomponente baseado no conceito delaytime: UM ESTUDO DE CASO SOBRE SHREDDER. In: XXXVIII ENCONTRO NACIONAL DE ENGENHARIA DE PRODUCAO, 2018, Maceió. Anais [...]. Maceió, 2018. p. 1-12.

ARULKUMARAN, K.; DEISENROTH, M. P.; BRUNDAGE, M.; BHARATH, A. A. Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, v. 34, n. 6, p. 26–38, 2017.

ASSAF, R.; DO, P.; NEFTI-MEZIANI, S.; SCARF, P. Wear rate–state interactions within a multi-component system: a study of a gearbox-accelerated life testing platform. Proceeding of the Intitution of Mechanical Enginieers, Part O: Journal of Risk and Reliability, v.232, n. 4, 2018.

ATAMURADOV, V.; MEDJAHER, K.; DERSIN, P.; LAMOUREUX, B.; ZERHOUNI, N. Prognostics and health management for maintenance practitioners-review, implementation and tools evaluation. International Journal of Prognostics and Health Management, v. 8, n. 060, p. 1-31, 2017.

BRUSA, E; MORSUT, S; BOSSO,N. Dynamic behavior and prevention of the damage of material of the massive hammer of the scrap shreddingmachine. Meccanica, v. 49, p. 575–586, 2014.

CAVALCANTE, C. A. V.; ALMEIDA, A. T. DE. Modelo multicritério de apoio à decisão para o planejamento de manutenção preventiva utilizando PROMETHEE II em situações de incerteza. Pesquisa Operacional, v. 25, n. 2, p. 279–296, 2005.

CHEN, N.; CHEN, Y.; LI, Z.; ZHOU, S.; SIEVENPIPER, C. Optimal variability sensitive conditionbased maintenance with a Cox PH model. Internacional Journal of Prodution Research, v. 49, n.7, p. 2083-2100, 2011.

CHEN, N.; Y, Z.; XIANG, Y.; ZHANG, L. Condition-based maintenance using the inverse

Gaussian degradation model. European Journal of Operational Research, v.243, n.1, p.190–199, 2015.

CHEUNG, KL.; HAUSMANN, HW. Joint determination of preventive maintenance and safety stocks in an unreliable production environment. Naval Research Logistics, v. 44, p. 257-272, 1997.

CHRISTER, A. Developments in delay time analysis for modeling plant maintenance. Journal of the Operational Research Society, v. 50, p.1120–1137, 1999.

DHILLON, B.S. Engineering maintenance: a modern approach. CRC Press, 2002.

DIEULLE, L.; BÉRENGUER, C.; GRALL, A.; ROUSSIGNOL, M. Sequential condition-based maintenance scheduling for a deteriorating system. European Journal of Operational Research, v. 150, n. 2, p. 451-461, 2003.

DINH, D.; DO, P.; IUNG, B. Degradation modeling and reliability assessment for a multi-component system with structural dependence. Computers & Industrial Engineering, n. 144, p.106443, 2020.

DO, P.; ASSAF, R.; SCARF, P.; IUNG, B. Modelling and application of condition-basedmaintenance for a two-component system with stochastic and economic dependencies.Reliability Engineering & System Safety, n. 182, p. 86-97, 2019.

EDWARDS, D. J.; HOLT, G. D.; HARRIS, F. C. A model for predicting plant maintenance costs. Construction Management and Economics, v. 18, p. 65–75, 2000.

FERREIRA NETO, W. A.; CAVALCANTE, C. A. V..; SANTOS, A. C. J.; ALBERTI, A. R. Uma política de inspeção para linhas de produção de aço considerando o nível de estoque e o tempo operacional. In: INnovation for Systems Information and Decision Meeting, 2020, Recife. Anais [...]. Recife, 2020a.

FERREIRA NETO, W. A.; CAVALCANTE, C. A. V..; PAIVA, R. G. N.; TENÓRIO, V. A. S. Política de manutenção baseada em aprendizado por reforço para uma estação de trituração de sucata de uma linha de produção de aço. In: INnovation for Systems Information and Decision Meeting, 2020, Recife. Anais [...]. Recife, 2020b.

FERREIRA NETO, W. A.; CAVALCANTE, C. A. V..; SANTOS, A. C. J.; ALBERTI, A. R. POLÍTICA HÍBRIDA DE MANUTENÇÃO DE DUAS FASES VOLTADA PARA SISTEMAS CRÍTICOS USANDO O CONCEITO DELAY-TIME. In: LII Simpósio Brasileiro de Pesquisa Operacional, 2020, João Pessoa. Anais[...]. João Pessoa, 2020c.

FITOUHI, M.; NOURELFATH, M.; GERSHWIN, S. Performance evaluation of a two-machine line with a finite buffer and condition-based maintenance. Reliability Engineering & System Safety, n. 166, p. 61-72, 2017.

GAN, S.; ZHANG, Z.; ZHOU, Y.; SHI, J. Intermediate buffer analysis for a production system. Applied Mathematical Modelling, v. 37, p. 8785-8795, 2013.

GARRAMIOLA, F.; POZA, J.; MADINA, P.; DEL OLMO, J.; ALMANDOZ, G. A Review

in Fault Diagnosis and Health Assessment for Railway Traction Drives. Appl. Sci. v.8, p.2475, 2018.

GIL, A. C. Como elaborar projetos de pesquisa. São Paulo: Atlas, 2002.

GRALL, A; BÉRENGUER, C.; DIEULLE, L. A condition-based maintenance policy for stochastically deteriorating systems. Reliability Engineering & System Safety, n.76, p.167–180, 2002.

GROOVER, MP. Automation, Production Systems, and Computer – Integrated Manufacturing. New York: Pearson. 2015.

GULLI, A.; PAL, S. Deep Learning with Keras: Implementing deep learning models and neural networks with the power of Python. Packt Publishing, 2017.

HABES, M. et al. The role of modern media technology in improving collaborative learning of students in Jordanian universities. International Journal of Information Technology and Language Studies, v. 2, p. 71–82, 2018.

HASSELT, H.; GUEZ, A.; SILVER, D. Deep Reinforcement Learning with Double Q-learning. In Thirtieth AAAI conference on artificial intelligence, 2016

HASHEMIAN, H. M.; BEAN, W. C. State-of-the-Art Predictive Maintenance Techniques. IEEE Transactions on Instrumentation and Measurement, v. 60, n. 10, 2011.

HELBING, G.; RITTER, M. Deep Learning for fault detection in wind turbines. Renewable and Sustainable Energy Reviews, v. 98, p. 189-198, 2018.

HUANG, J.; CHANG, Q.; ARINEZ, J.; XIAO, G. A Maintenance and Energy Saving Joint Control Scheme for Sustainable Manufacturing Systems. Procedia CIRP, v.80, p.263–268, 2019.

HUANG, J.; CHANG, Q.; ARINEZ, J. Deep Reinforcement Learning based Preventive Maintenance Policy for Serial Production Lines. Expert Systems with Applications: X, p.100034, 2020.

JIN, X.; SIEGEL, D.; WEISS, B. A.; GAMEL, E.; WANG, W.; LEE, J.; NI, J. The present status and future growth of maintenance in US manufacturing: results from a pilot survey. Manufacturing review, v. 3, 2016.

JONGE, B.; TEUNTER, R.; TINGA, T. The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. Reliability Engineering & System Safety, n. 158, p. 21–30, 2017.

KARAMATSOUKIS, C; KYRIAKIDIS, E. Optimal maintenance of two stochastically deteriorating machines with an intermediate buffer. European Journal of Operational Research, v. 207, p. 297–308, 2010.

KIGMAN, D.P.;BA, J. Adam: A Method for Stochastic Optimization. Computer Science-Machine Learning, v.1 , 2014.

KIRCHNER, J.; TIMMEL, G.; SCHUBERT, G. Comminution of metals in shredders with horizontally and vertically mounted rotors - Microprocesses and parameters. Powder Technology, v. 105, p. 274-281, 1999.

KUHNLE, A.; JAKUBIK, J.; LANZA, G. Reinforcement learning for opportunistic maintenance optimization. Prod. Eng. Res. Devel, v.13, p.33–41 , 2019.

LIAO, H.; ELSAYED, E. A.; CHAN, L-Y. Maintenance of continuously monitored degrading systems. European Journal of Operational Research, v. 175, p. 821–835, 2006.

LIU, B.; WU, S.; XIE, M.; KUO, W. A condition-based maintenance policy for degrading systems with age-and state-dependent operating cost. European Journal of Operational Research, v. 263, p. :879–887, 2017.

LIU, B.; DO, P.; IUNG, B. XIE, M. Stochastic filtering approach for condition-based maintenance considering sensor degradation. IEEE Transactions on Automation Science and Engineering, n. 17, p. 177-190, 2019.

MARCONI, M. DE A.; LAKATOS, E. M. Técnicas de Pesquisa •. Vol. 2 ed. São Paulo: [s.n.].

MOBLEY, R. K. An Introduction to Predictive Maintenance. New York: Van Nostrand Reinhold, 1990.

MOUBRAY, J. Reliability-Centered Maintenance. New York: Industrial Press, 1997.

MNIH, V.; KAVUKCUOGLU1, K.; SILVER, D.; RUSU, A.A.; VENESS, J.; BELLEMARE, M. G.; GRAVES, A.; RIEDMILLER, M.; FIDJELAND, A.K.; OSTROVSKI, G.; PETERSEN, S.; BEATTIE, C.; SADIK, A.; ANTONOGLOU, I.; KING, H.; KUMARAN, D.; WIERSTRA, D.; LEGG1, S.; HASSABIS1, D. Human-level control through deep reinforcement learning. Nature, v. 518, p. 529-533, 2016.

NAKAJIMA, S. Introdução ao TPM - Total Productive Maintenance. Trad. Mário Nishimura. São Paulo: IMC Internacional Sistemas Educativos, 1989.

NGUYEN, K.T.P.; MEDJAHER, K. A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering & System Safety, n. 188, p. 251-262, 2019.

PINTELON, L.; GELDERS, L.; VANPUYVELDE, F. Maintenance Management, 2 ed. Leuven: Acco Belgium, 2000.

PUTERMAN, ML. Markov decision processes.: discrete stochastic dynamic programming. John Wiley & Sons; 2014.

RASMEKOMEN, N.; PARLIKAD, A.K. Condition-based maintenance of multi-component systems with degradation state-rate interactions. Reliability Engineering & System Safety, n. 148, p. 1-10, 2016.

REZG, N.; DELLAGI, S.; KHATAB, A. Joint Optimization of Maintenance and Production

Policies. New Jersey: John Wiley & Sons. 2014.

ROCCHETTA, R.; BELLANI, L.; COMPARE, M.; ZIO, E.; PATELLI, E. A reinforcement learning framework for optimal operation and maintenance of power grids. Applied Energy, v. 241, p. 291–301, 2019.

SAMMUT, C.; WEBB, G. I. (Ed.). Encyclopedia of Machine Learning and Data Mining. Springer US, 2017.

SAMUEL, A. Some studies in machine learning using the game of checkers. IBM Journal, v.3, p. 210–229, 1959.

SCHOLTEN, K.; BLOK, C. DE; HAAR, R. How Flexibility Accommodates Demand Variability in a Service Chain: Insights from Exploratory Interviews in the Refugee Supply Chain. [s.l: s.n.].

SCHOUTEN, FA.; VANNESTE, SG. Maintenance optimization of a production system with buffer capacity. European Journal of Operational Research, v. 82, p. 323–338, 1995.

SORO, I. W.; NOURELFATH, M.; DAOUD, A. Performance evaluation of multi-state degraded systems with minimal repairs and imperfect preventive maintenance. Reliability Engineering & System Safety, v. 95, p. 65–69, 2010.

STRICKER, N.; KUHNLE, A.; STURM, R.; FRIESS, S. Reinforcement learning for adaptive order dispatching in the semiconductor industry. CIRP Annals, 2018.

SUTTON, R. S.; BARTO, A. G. Reinforcement learning: An introduction. Cambridge, MA: MIT Press, 2012.

SUTTON, R. S.; BARTO, A. G. Reinforcement Learning: An Introduction. MIT Press, 2018.

SWANSON, L. An empirical study of the relationship between production technology and maintenance management. International Journal of Production Economics, v. 53, p. 191-207, 1997.

VAN DER DUYN SCHOUTEN, FA.; VANNESTE, SG. Maintenance optimization of a production system with buffer capacity. European Journal of Operational Research, v. 82, p. 323–338, 1995.

WANG, H. A survey of maintenance policies of deteriorating systems. European Journal of Operational Research. n. 139, p. 469–489, 2002.

WANG, W. An overview of the recent advances in delay-time-based maintenance modelling. Reliability Engineering & System Safety, n. 106, p.165–178, 2012.

WANG, X.; WANG, H.; QI, C. for a Machine with Multiple Deteriorating Yield Levels ⋆. v. 1, n. 60904075, p. 9–19, 2014.

WANG, X.; WANG, H.; QI, C. Multi-agent reinforcement learning based maintenance policy for a resource constrained flow line system. Journal of Intelligent Manufacturing, p. 325–333,

2016.

WANG, K.; WANG, Y. How AI Affects the Future Predictive Maintenance: A Primer of Deep Learning. Advanced Manufacturing and Automation VII, v. 451, p.1-9, 2018.

WATKINS, C.J.C.H. Learning from delayed rewards. PhD Thesis, University of Cambridge, England, 1989.

WATKINS, C.J.C.H.; DAYAN, P. Q-learning. Machine Learning, v. 8, p. 279-292, 1992.

WORLD STEEL ASSOCIATION. World Crude Steel Production – Summary. 2019.

WUEST, T.; WEIMER, D.; IRGENS, C.; THOBEN, D. K. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research, v.4, n.1, p.23-45, 2016.

ZHANG, N.; SI, W. Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. Reliability Engineering & System Safety, n. 203, p.107094, 2020.

ZHOU, X; HU, Z; XIAO, X; LI, M. Research on shredding process and characteristics of multi-materia plates for recycled cars. Proc IMechE Part B: J Engineering Manufacture 2015, n.230, v.10, p.1834-1844, 2015.

ZHOU, X.; HU, Z.; QIN, X.; TAO, Y.; HUA, L. Study on the stress characteristic and fatigue life of the shredderpin. Engineering Failure Analysis, n.59, p.444-455, 2016a.

ZHOU, X.; HU, Z.; TAO, Y.; QIN, X.; HUA, L. Failure mechanisms and structural optimization of shredder hammer for metal scraps. Chinese Journal of Mechanical Engineering, n. 29, v.4, p. 792-801, 2016b.

SANDER, S.; SCHUBERT, G.; JÄCKEL, HG. The fundamentals of the comminution of metals in shredders of the swing-hammer type. International Journal of Mineral Processing, n.74, p. 385-393, 2004.

ZOU, J.; CHANG, Q.; LEI, Y.; ARINEZ, J. Production System Performance Identification Using Sensor Data, v. 48, n. 2, 2018.