2021

Towards A Computational Intelligence Framework in Steel Product Quality and Cost Control

Fucun Li

Towards A Computational Intelligence Framework in Steel Product Quality and Cost Control

Fucun Li

Principle Supervisor：Associate Prof Jun Shen

(School of Computing and Information Technology, UOW)

Co-Supervisor：Associate Prof Zengxi Pan

(School of Mechanical, Materials, Mechatronic and Biomedical Engineering, UOW)

Associate Supervisor：Prof Fang Dong

(School of Computer Science and Engineering, at Southeast University, China)

Abstract

Steel is a fundamental raw material for all industries. It can be widely used in various fields, including construction, bridges, ships, containers, medical devices and cars. However, the production process of iron and steel is very perplexing, which consists of four processes: ironmaking, steelmaking, continuous casting and rolling. It is also extremely complicated to control the quality of steel during the full manufacturing process. Therefore, the quality control of steel is considered as a huge challenge for the whole steel industry. This thesis studies the quality control, taking the case of Nanjing Iron and Steel Group, and then provides new approaches for quality analysis, management and control of the industry.

At present, Nanjing Iron and Steel Group has established a quality management and control system, which oversees many systems involved in the steel manufacturing. It poses a high statistical requirement for business professionals, resulting in a limited use of the system. A lot of data of quality has been collected in each system. At present, all systems mainly pay attention to the processing and analysis of the data after the manufacturing process, and the quality problems of the products are mainly tested by sampling-experimental method. This method cannot detect product quality or predict in advance the hidden quality issues in a timely manner. In the quality control system, the responsibilities and functions of different information systems involved are intricate. Each information system is merely responsible for storing the data of its corresponding functions. Hence, the data in each information system is relatively isolated, forming a data island. The iron and steel production process belongs to the process industry. The data in multiple information systems can be combined to analyze and predict the quality of products in depth and provide an early warning alert. Therefore, it is necessary to introduce new product quality control methods in the steel industry. With the waves

of industry 4.0 and intelligent manufacturing, intelligent technology has also been introduced in the field of quality control to improve the competitiveness of the iron and steel enterprises in the industry. Applying intelligent technology can generate accurate quality analysis and optimal prediction results based on the data distributed in the factory and determine the online adjustment of the production process. This not only gives rise to the product quality control, but is also beneficial to in the reduction of product costs. Inspired from this, this paper provide in-depth discussion in three chapters: (1) For scrap steel to be used as raw material, how to use artificial intelligence algorithms to evaluate its quality grade is studied in chapter 3; (2) the probability that the longitudinal crack occurs on the surface of continuous casting slab is studied in chapter 4; (3) The prediction of mechanical properties of finished steel plate in chapter 5. All these 3 chapters will serve as the technical support of quality control in iron and steel production.

The main challenges for the evaluation of scrap quality grade are: (1) How to obtain high resolution pictures which can represent the scrap grade of each vehicle without affecting the production process; (2) The criteria of manual labels corresponding to the collected pictures is difficult to define; (3) How to evaluate and infer the grade of the collected scrap pictures efficiently in real-time. This thesis is addressing such challenges through the in-depth investigation of the acceptance and use of scrap. For example, the installation of high-definition cameras at truck loading sites is effective to obtain high-quality scrap pictures. In the intelligent rating process, the problem of matching the pictures with the scrap grade label is solved, using the method of matching the studied scrap with three seed material type ratio quantitative label. Furthermore, for real-time efficient online rating, this thesis proposes a fast, accurate and reliable YOLOv3 target detection algorithm and VGG16 classification algorithm to evaluate the scrap quality grade and realize the transparency, openness and informationization of the whole scrap quality grade evaluation process.

The main challenges for the prediction of longitudinal cracks on the surface of continuous casting slab are: (1) The collection and integration of data during production process; (2) How to construct a suitable machine learning model to train the collected data. In order to address these challenges, a five-layer industrial Internet platform is established firstly to collect, transmit, store and preprocess the data that affects the factors of longitudinal cracks in continuous casting production. Then, a back propagation (BP) neural network model with hidden layer number pre-optimization is applied to model, fit and evaluate the preprocessed data, and it is used to predict whether longitudinal cracks occur on the surface of continuous casting slab in advance.

In light of steel plate mechanical properties prediction, the main challenge is how to construct a machine learning model with strong ability of learning and generalization to train the data with high dimensionality, nonlinearity and strong correlation in the production process. These factors ensure that the model have high accuracy in the reasoning process. The integrated learning method is proposed to model the data, and the basic learner contains linear and nonlinear base models to predict the mechanical properties of steel products, including tensile strength, yield strength, impact energy and elongation. This base model is integrated by models that can process high-dimensional and nonlinear data. In addition, the integrated model can enhance the expression of generalization ability.

In the actual production, the application of robust quality prediction methods to iron and steel production and quality control will greatly reduce the cost of laboratory sampling, manual testing and subsequent reprocessing, as well as the acquisition of steelmaking raw materials. This effort will benefit from cooperation with steelmakers, which means it is feasible to collect and clean raw data, build models and verify the factories online in the future manufacturing.

Acknowledgments

First and foremost, I am extremely grateful to my supervisors, Associate Prof. Jun Shen, Zengxi Pan and Fang Dong for their invaluable advice, support, and patience during the course of my PhD degree. Their knowledge and experience have helped me greatly through all the time. Additionally, I would like to express gratitude to my wife and little daughter, and it is their tremendous understanding that made it possible for me to complete my study. Lastly, I would like to thank my friends, and colleagues Huaming Chen, Ailing Yang and Xiaochen Zhang, for a cherished time spent together and their encouragement and support all through my studies.

List of Publications

1. Li, F., Wu, J., Dong, F., Lin, J., Sun, G., Chen, H., & Shen, J. (2018, December). Ensemble Machine Learning Systems for the Estimation of Steel Quality Control. In 2018 IEEE International Conference on Big Data (Big Data) (pp. 2245-2252).IEEE.

2. Li, F., Yang, A., Chen, H., Sun, G., Wang, F., Xie, Y., ... & Shen, J. (2020, December). Towards Industrial Internet of Things in Steel Manufacturing: A Multiple-Factor-based Detection System of Longitudinal Surface Cracks. In 2020 IEEE International Conference on Big Data (Big Data) (pp. 4627-4635). IEEE.

3. Yong, B., Li, F., Lv, Q., Shen, J., & Zhou, Q. (2019). Derivative-based acceleration of general vector machine. Soft Computing, 23(3), 987-995.

4. Wu, Q., Shen, J., Yong, B., Wu, J., Li, F., Wang, J., & Zhou, Q. (2019). Smart fog based workflow for traffic control networks. Future Generation Computer Systems, 97, 825-835.

5. Yong, B., Shen, J., Liu, X., Li, F., Chen, H., & Zhou, Q. (2020). An intelligent blockchain-based system for safe vaccine supply and supervision. International Journal of Information Management, 52, 102024.

6. Yong, B., Xu, Z., Shen, J., Chen, H., Wu, J., Li, F., & Zhou, Q. (2020). A novel Monte Carlo-based neural network model for electricity load forecasting. International Journal of Embedded Systems, 12(4), 522-533.

7. Yong, B., Huang, L., Li, F., Shen, J., Wang, X., & Zhou, Q. (2020). A research of Monte Carlo optimized neural network for electricity load forecast. The Journal of Supercomputing, 76(8), 6330-6343.

8. Chen, H., Wang, L., Jin, Y., Chi, C. H., Li, F., Chu, H., & Shen, J. (2019, December). Hyperparameter Estimation in SVM with GPU Acceleration for Prediction of Protein-Protein Interactions. In 2019 IEEE International Conference on Big Data (Big Data) (pp. 2197-2204). IEEE.

9. Yong, B., Wei, Y., Shen, J., Li, F., Jiang, X., & Zhou, Q. (2019, July). Combined General Vector Machine for Single Point Electricity Load Forecast. In International Conference on Frontier Computing (pp. 283-291). Springer, Singapore.

10. Yong, B., Wang, C., Shen, J., Li, F., Yin, H., & Zhou, R. (2020). Automatic ventricular nuclear magnetic resonance image processing with deep learning. Multimedia Tools and Applications, 1-17.

11. Yong, B., Huang, L., Li, F., Shen, J., Wang, X., & Zhou, Q. (2018, July). GVM based copy-dynamics model for electricity load forecast. In International Conference on Frontier Computing (pp. 203-211). Springer, Singapore.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the rise of industry 4.0, each country is gradually implementing intelligent manufacturing. This paper uses intelligent technology to study the field of quality management and control in intelligent manufacturing of the iron and steel industry. Therefore, this chapter introduces the production process of iron and steel in Section 1.1, the general situation of total quality management of existing iron and steel enterprises in Section 1.2, the development process of artificial intelligence in Section 1.3, the quality management of Nanjing Iron and Steel Group in Section Section 1.2.5, the research objectives in Section 1.4 and the organizational structure of this thesis in Section 1.5.

1.1 Steel production process

Iron and steel are the basic raw materials for all industries. They can be widely used in various fields such as the construction of buildings, bridges, ships, containers, medical devices and automobiles[1]. Steel is the most widely used and important material in the world. The total annual output of steel in the world is about 1.69 billion tons. But how is this steel produced?

Generally speaking, iron and steel production is divided into four processes: ironmaking, steelmaking, continuous casting and rolling[2]. Ironmaking is the process of reverting back to iron from sinter and block ore. The coke, sinter and block ore, together with a small amount of limestone, are sent into a blast furnace to make liquid pig iron (hot metal) and then sent to the steelmaking plant as raw materials for steelmaking[3]. The process of steelmaking is to remove excessive carbon, sulfur and phosphorus impurities from molten iron and scrap and add an appropriate amount of alloy composition. Continuous casting is the continuous injection of molten steel into the mould cooled by water through the tundish. After condensing into the shell, it is pulled out of the mould at a constant speed, then cooled by water spray. After all solidification, the billet is cut into a specified length. Steel rolling is the continuous feeding of steel ingots and billets in different rolling mills to form a variety of steel products. The steel production process is shown in Figure 1.1.

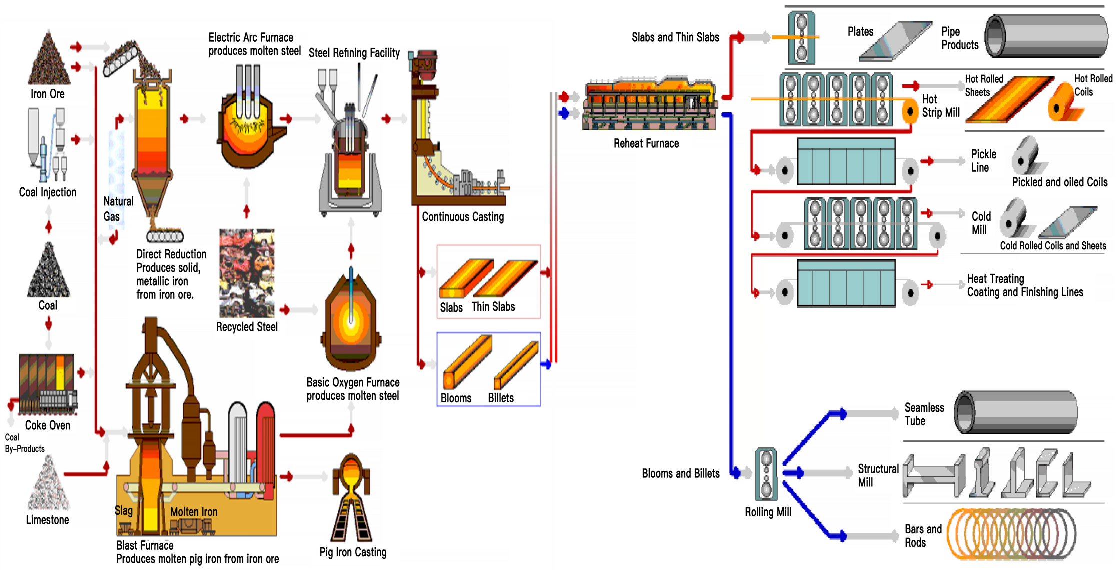


Figure 1.1: Flow Chart of Steel Production.

The functions of ironmaking, steelmaking, continuous casting and rolling in the process of iron and steel production was briefly introduced, and each process is described in more detail below:

The ironmaking process is an important part of the iron and steel industry and a continuous production process to revert iron ore back to pig iron. During this process, solid raw materials such as iron ore, coke and flux are transported to a blast furnace in batches[3]. These batches will be sent into the furnace at the top using a required device, and these materials in the furnace need to be kept at a certain level[4]. Coke and ore are alternately stratified in the furnace. The ore gradually reverts, smelts into iron and slag, accumulates in the furnace, falls and regularly flows from the iron and slag mouth[3, 4].

The steelmaking process is divided into two steelmaking methods: long process steelmaking and short process steelmaking. The long one is iron ore raw materials are produced as hot metal by using a blast furnace after a sintering pellet treatment. After pretreatment, molten iron is converted for steelmaking, refined to qualified molten steel, and different billets are sent through continuous casting process, and finally rolled into various products[5]. The short process for steelmaking is the separation of recycled scrap, direct addition to the electric furnace, the use of electric energy as a heat source for smelting, the attainment of a qualified composition of molten steel through the continuous casting process to create different billets, and finally the billets are rolled into various products[6]. The main tasks of the steelmaking process are decarbonization, deoxidation, heating, degassing and getting rid of the non-metallic inclusions and alloying. In the process of steelmaking, raw materials are the basis of steelmaking, and the quality of raw materials has a direct impact on the steelmaking and also the quality of the finished steel. If the quality of raw materials does not meet the requirements, it will inevitably lead to increased consumption, sometimes poor product quality, and even produce waste products, resulting in an increase in product costs. Practice in China and other countries has proved that the use of refined steel and standardized materials is a prerequisite for the automation of smelting process, and it is also the basis of being able to improve both various technical indicators and economic efficiency.

Continuous casting process is to move the steel ladle equipped with refined molten steel to the turntable. After the turntable rotates to the pouring position, the molten steel will be injected into the tundish, and the tundish will be distributed to each crystallizer by the nozzle[7]. Crystallizer is one of the core equipment of the casting facility, which makes the casting form into the right shape, solidify and crystallize quickly. Tension Leveler acts with the mould oscillating device to pull out the casting in the crystallizer and cut into a slab of a certain length after cooling and electromagnetic stirring[8].

Steel rolling is a metal forming process in which metal stock is passed through one or more pairs of rolls to reduce the thickness, to make the thickness uniform, or to impart a desired mechanical property. The purposes of rolling is the same as other pressure processing. It can help obtain the desired shape and also improve the internal quality of the steel. The rolling method can be divided into hot rolling process and cold rolling process according to different rolling temperature.

Hot rolling process is to first send the continuous casting billet from the steelmaking plant to the heating furnace, and then it will be sent to the finishing mill after repeated rolling by the blooming mill. Rolling steel is a metal pressure processing[9]. In hot rolling process, the billet is heated and softened, and is sent into the mill through the roller table, and finally is rolled into the size required by the user. Also, metals have high plasticity and low deformation resistance, which greatly reduces the energy consumption of metal deformation and therefore reduces the cost[10]. In addition, hot rolling can improve the workability of metals and alloys, that is, the coarse grain is reshaped, and the cracks are obviously reduced. Also, the casting structure is transformed into deformed structure, and the workability of the alloy are improved. The hot rolling process usually uses large slabs and applies large rolling pressure rate, which improves both the production efficiency and the rolling speed and make continuous automatic rolling possible. The main product of hot rolling is the hot rolled coil, which is also used as the raw material in cold rolling[10].

Cold rolling process is to first remove the oxide coating of the coil sent from the hot rolling mill by hydrochloric (sulfuric) acid pickling, and then send the coil to the cold rolling mill[11]. Cold rolling processes usually include uncoiling, rolling, degreasing (acid pickling), annealing (heat treatment), coiling, etc.

From the content above, it is known that there are many processes in the production of iron and steel, and each process is very complex. The control of quality and good management is required in each process. There are already some effective methods in the field, and they will be introduced in Section 1.2.

1.2 Total Quality Management in Iron and Steel Industry

In order to get more profit in the steel market and realize the transformation from a large iron and steel country to a powerful one, it is extremely necessary to improve the quality of steel products and it will also need long-term persistence.

Improving the quality of Chinese iron and steel products requires to strengthen the total quality management of Chinese iron and steel enterprises. The factors affecting quality mainly include people, equipment, raw materials, methods, environment, and each factor can be subdivided into sub-factors. These factors act together and affect each other, which jointly determine the level and improvement of quality. Total quality management is not only reflected in the quality inspection of the products, but also in the overall, full and whole process, including product design, raw material procurement, production, quality inspection, sales to after-sale. In conclusion, the whole process needs to be strictly checked and supervised, realizing the transition from simple inspection to zero defect production under advanced inspection. It can improve product quality and after-sales service quality, also reduce defective rate and production costs.

In addition, the ultimate goal of production is to maximize profits, and the pursuit of profits is the fundamental driving force of enterprise. In order to increase the profit, enterprises need to control the cost and keep the good quality at the same time, but it will be wrong to blindly increase the cost for better quality. Instead, they should take the market demand and consumption structure into consideration and improve the quality of the products under the premise of acceptable cost and guaranteed profit.

For a long time, Chinese iron and steel enterprises have done a lot of research and application in total quality management, and achieved obvious results, but there is still a long way to go. With the fast development of this industry, the quality of the products and services of enterprises are required to be updated. Iron and steel enterprises also asks for higher requirements for further development of total quality management. Only by strengthening the total quality management of iron and steel enterprises, improving the quality and technical content of the products, and reducing the rate of defective products, can we reduce the production cost, increase the efficiency and improve the customer satisfaction. Therefore, strengthening the total quality management of Chinese iron and steel enterprises is a very urgent and necessary task.

1.2.1 Evolution of Total Quality Management

With a long history of evolution, quality management can be dated back to the emergence of commodity production of human beings. Finished product inspection was mainly relied on then. The means and methods for solving quality problems are constantly evolving with the development of social production and the advancement of science and technology. From inspection-based quality management in the traditional handicraft era, to the statistical quality management after the introduction of mathematical statistics methodology, the following total quality management stage, and the modern quality management stage that has only been gradually improved until now, quality management is still developing. For example, a more improved concept of quality management engineering has been introduced currently[12, 13].

Modern quality management[14] started in the 1920s, and has developed into a new discipline over decades of development. From the practice of modern quality management, according to the means and methods for solving quality problems, its evolution may be roughly divided into the following stages:

(1) Quality inspection stage

At the end of the 19th century, with the development of European countries such as Britain, Spain and France after the industrial revolution, capitalist factories gradually replaced decentralized cottage industry workshops, machine production replaced manual labor, and laborers concentrated in one factory for batch production, resulting in inspection-based quality management. Quality inspection is to control and ensure the quality of products transferred to the next process or leaving the factory through strict inspections. Quality inspection is carried out by means of various instruments and meters, in the manner of strict check and 100% coverage.

In the 1870s, according to the needs of production and use, the concept of parts interchange was introduced. At the same time, it was noticed that with guaranteed parts interchange, the machining error in size is allowed of a fluctuation range, and thus the concept of machining tolerance came into being. This initially laid the foundation for the technical theory of quality inspection.

At the beginning of the 20th century, on the basis of systematically summarized past quality management practices and experiences, the idea of scientific management came into being, while a series of scientific management theories and methods were established. Taylor (American) is a representative who proposed this theory and put it into practice. He advocated a reasonable division of labor between managers and operators to separate planning functions from executive functions, and introduction of intermediate inspection. As a result, a functional management system is formed in which design, operation, and inspection are separately assigned with responsible personnel. This is the first time in history that the inspection function is separated from the operation function, and also the first time that inspection personnel are separated from operation personnel. This major change in the division of labor caused by modern mass production had led to rapid improvements in labor productivity, fixed asset utilization, and product quality, thus achieving significant economic effects.

The quality inspection stage is mainly characterized by the separation of three powers, that is, special personnel assigned for developing standards, special personnel assigned for manufacturing, and special personnel assigned for inspecting products in accordance with standards. At this stage, quality management only emphasizes screening of products finished to separate the acceptable from the unacceptable. Quality inspection, intended for controlling purposes, is necessary and effective to ensure that no unacceptable products are released from the factory.

However, such quality management model has many shortcomings. For example, the persons involved in production and management do not participate in quality inspection and management, leading to conflicts; the full inspection method leads to large workloads and long cycles; post-inspection is performed after completion of the product, resulting in great waste and loss. Therefore, this quality management method gradually become unable to adapt to the requirements of economic development then, and should be improved and developed.

(2) Statistical quality control stage

Since there are three problems in quality inspection, a new method is objectively required to solve the problems. Statistical quality control can be dated back to after the 1940s[15], when mass production came into being with the further development of productivity. How to control the quality of large quantities of products became a prominent problem. The United Kingdom and the United States, among others, successively promulgated new tolerance standards, which had played a certain role in ensuring the interchangeability and versatility of mass-produced products. At the same time, some statisticians begun to study the use of statistical methods instead of simple inspection methods to control product quality. In 1924, Shewhart, an American engineer at Bell Laboratories, proposed to use mathematical statistics for quality management, and invented the famous control chart tool, laying the foundation for statistical quality management. After the start of the Second World War, the war imposed new strict requirements on the production quality of munitions, e.g. weapons, ammunition and military ships. The quality of weapons and warships that lack prior control and sabotage inspection guarantees will inevitably affect the progress of the war, which made it urgent to apply new methods of mathematical statistics to quality management. As a result, effective statistical quality management was not only adopted in the national defense and military industry, but also promoted in other sectors, making statistical quality management greatly developed. This method realized the transition from passive check-offs to active prevention in the production process. Compared with the traditional management based on inspection, statistical quality management is an update of concept and inspection functions, and a leap in quality management methods.

Although statistical quality management is scientific and economical, it also has many shortcomings: (l) It aims to meet product standards instead of user needs. (2) It focuses on process management and does not control the entire process of product quality formation. (3) It involves statistical techniques difficult for ordinary people to master. (4) Quality management and organization management are not closely connected. Due to the above-mentioned problems, statistical quality management cannot meet the needs of the development of modern industrial production and needs to be further developed[16].

(3) Total quality management stage

In the 1960s, with the rapid development of social productivity and rapid changes in science and technology, many new challenges are presented to quality management. (1) Many sciences and technologies, large projects and precision products in the steel industry required the use of the concept of system to deal with quality issues. (2) The emerging behavior school advocated focusing on the role of people in management. (3) The rise of the movement to protect the interests of consumers required companies to assume quality and economic responsibilities. (4) The rapid changes in various markets, including the steel market, and the business decision making required continuous development of higher-quality products. In this context, it was very difficult to guarantee and improve product quality only by relying on quality inspection and statistical methods. Moreover, it was obviously inappropriate to delegate the quality functions fully to dedicated quality management engineers and technicians.

Feigenbaum, the quality manager of General Electric in the United States, was the first to put forward the concept of total quality management. In 1961, his work *Total Quality Control* was published[17]. It emphasized that it is the responsibility of all employees in the company to implement quality functions. All employees in the enterprise should be made quality aware and responsible for quality. He argued that total quality control is intended to perform market research, design, production and service at the most economical level and giving full consideration of user requirements, so as to form an integrated and effective system for the quality development, maintenance and improvement activities of various departments in the enterprise. Since then, Feigenbaum's total quality control concept was gradually accepted by countries all over the world. Countries formed a quality management model with their own characteristics as appropriate to respective situations.

Since the implementation of total quality management in 1978, China has developed rapidly both in practice and theory. Total quality management has been gradually implemented from industrial enterprises to transportation, post and telecommunications, commercial enterprises and township enterprises. Some concepts and methods of quality management had been formulated as national standards. On the basis of drawing from the experience and lessons of total quality management, extensive enterprises had further comprehensively and deeply implemented this modern international quality management method.

The stages of quality management evolution are not isolated and mutually exclusive, but interrelated. The former stages are the foundations of the latter ones, which in turn are the inheritance and development of the previous one. From the perspective of the evolution history of quality management, the methods and means used to solve quality problems are constantly developing and improving, and are closely related to the development of science and technology and social productivity. It is foreseeable that with the rise of the new technological revolution and the resulting quality challenges, the methods and means to solve quality problems will be enriched and improved, and quality management will develop to an updated stage, i.e. modern quality management engineering stage.

1.2.2 Total Quality Management Methods

According to Yao's research, the implementation of TQM in enterprises should be man oriented, and corresponding improvement measures were recommended as follows: (1) enhance quality awareness, establish and implement the customer-centered quality policy; (2) reduce longitudinal variation of organization; (3) strengthen quality education and training, and establish a "people-oriented" training mechanism; (4) establish an effective quality incentive mechanism, adopt correct incentive strategies, and regularly evaluate employee satisfaction.

According to Chen Bingquan, the problems in quality management can be summarized into three areas[18]: (1) the implementation of the enterprise's quality system is out of touch with the actual quality management work; (2) the enterprise quality data is scattered, showing "information island phenomenon" and lacking of correlation, so the value of the quality data has not been discovered; (3) difficult to trace product quality and low efficiency in handling quality issues.

According to Sui Lihui, at present, there are many problems in the implementation of total quality management in Chinese enterprises, including insufficient attention by enterprise leaders and employees, insufficient theoretical learning, and imperfect enterprise management system[19]. Therefore, if an enterprise wants to effectively implement total quality management and improve the implementation effect, it must start from the following aspects: (1) the business leaders attach great importance to it; (2) establish a long-term effective TQM training mechanism; (3) establish a complete quality management system.

From the above studies, Chinese enterprises still have a long way to go in implementation of TQM, and they should continue to improve it in the process of implementation for a better effect.

In addition, Chinese scholars also have conducted thorough research on the relationship between the implementation of TQM and enterprise performance. Main representative works are as follows:

The results of analysis by Tang Xiaofen and Jin Guoqiang[20] indicate that: (i) there is a strong correlation between personnel management and employee participation; (ii) the employee participation is highly correlated with customer satisfaction; (iii) it has a very small coefficient of correlation between supplier quality and other indicators. The results of Wang Renpeng's research show that: (i) the implementation effect of quality management is related to many factors, and customer focus has the biggest direct effect on the results while leadership has the biggest indirect effect on the results; (ii) "Leadership" and "Information and Analysis" are the starting and supporting point of the whole model respectively, and their effects on the results are mainly embodied by their influence and promotion on other factors; (iii) the implementation of "customer-focused" decision-making and optimization of human resources can directly enhance the quality management effect; (iv) "process management" can directly and indirectly enhance the effect of quality management; (v) "human resources" have obvious direct effect on process and results[21].

Through study using path analysis, Chen Qinghua mainly concluded that: (i) all factors of TQM have a positive impact on enterprise performance; (ii) among the six factors of TQM, customers, markets and resources have the most significant direct effects on enterprise performance; (iii) among the indirect effects on enterprise performance, leadership plays the biggest role, followed by measurement, analysis and improvement; (iv) there is a certain positive correlation among the factors of TQM; (v) to improve enterprise performance through the TQM, enterprises should start from the following key aspects: customer and market, resources, leadership and measurement, analysis and improvement.

1.2.3 Basic Content of TQM Theory

After continuous development, TQM has gradually become a sound quality management system, paving the way for rapid and stable development of enterprise quality management level from ideological guidance and practical operation.

(1) Definition of TQM

(i) "Total" is relative to "statistics" in the statistical quality control. In other words, in order to produce products that meet customer requirements and provide services satisfactory to the customers, it is not enough to control the production process only by statistical methods, and it is necessary to give full play to the role of each member of the organization by combining various management methods and means, so as to solve quality problems more comprehensively[22].

(ii) “Total” is also relative to the manufacturing process. Product quality has a process of production, formation and realization, including market research, R&D, design, standards development, process planning, procurement, equipment allocation, manufacturing, process control, inspection, sales, after-sales service, etc. Their mutual restriction and combined effect determines the final quality level, so it is far from enough to control the manufacturing process only [12].

(iii) Quality should be a perfect combination of “the most economical level” and “fully meeting customer requirements”, otherwise it would be meaningless to talk about quality apart from benefits[22].

The proposed TQM concept mainly includes three dimensions: (i) advanced system management thought; (ii) emphasis on the establishment of an effective quality system; (iii) for benefits of users and society.

(2) Application basis and work content of TQM[23].

The application of TQM is based on system engineering and management (system engineering), sound technical methods (control engineering) and effective interpersonal relationship (behavior engineering), so TQM can be advanced only by combination of the above three.

As determined by the comprehensiveness of TQM, the TQM shall include quality management in the following four processes: design process, manufacturing process, supporting process and application process[24].

(i) Design process

Being the primary link of TQM, quality management in this design process involves all product design processes before official production of products, including market survey, product design, process preparation, trial production, evaluation, etc. It is to develop products that meet the user’s requirements with higher use value and ensure good economic benefits when products are put into production in consideration of the production technology level and the process. Its main work contents are as follows: establishment of product quality objectives, design review and joint verification, trial production and identification of new products[25].

Quality Function Deployment (QFD) is considered as the most beneficial tool for quality design in these advanced production modes. When applied to the product planning and design stage, it can, to the maximum extent, meet the customer needs with the fastest speed, the lowest cost and excellent quality, so it is an important tool for the TQM of the company[26].

(ii) Manufacturing process Manufacturing process is the process of directly processing products. Being a key link in the production process of an enterprise, it is aimed to guarantee the product quality and is realized by controlling six factors (5M1E), i.e., manpower, machine, material, method, measurement, and environment. It is the key of quality management in this process to establish a production system that can stably produce qualified products and high-quality products. Both “check” and “prevention” are required for this. Its main work contents are as follows: implementing process specifications, strengthening process management, organizing the technical inspection, choosing reasonable inspection method, establishing a special inspection team consisting of both specialties and masses.

(iii) Supporting and service process

Supporting process is the process of providing various materials and technical conditions to ensure the normal manufacturing process. The supporting production process of manufacturing enterprises includes power production, equipment maintenance, tool manufacturing, etc. Service process is the process of material procurement and supply, warehousing, transportation services, etc. Both of them are important contents of TQM. Many quality problems arising from the manufacturing process are often related to the work quality of these departments. The basic task of quality management in the supporting process is to provide high-quality services and good technical conditions to ensure and improve product quality. Its main work contents are as follows: doing a good job in the quality management of material procurement and supply to ensure the procurement quality, carrying out strict check and acceptance on the incoming materials, and providing all kinds of materials required for production on quality, quantity and time; organizing the equipment maintenance work and maintaining good equipment technology, etc.

(iv)Application process

For ultimate goal of quality management, it is to better meet customer needs. Being a process of testing the actual quality of products, this process is the continuation of internal quality management of enterprises and the starting point and ultimate purpose of TQM. The basic task of quality management in this process is to improve service quality (including pre-sales and after-sales service quality), ensure the actual application effect of products, and constantly urge enterprises to continue R&D and improvement of product quality. Its main work contents are as follows: actively carrying out technical service work and dealing with the quality problems of ex-factory products; investigating and studying the application effect of products and user requirements to fulfill a quality policy of true customer satisfaction.

1.2.4 Implementation Principles of TQM

In order to achieve the ultimate management objectives, TQM shall be implemented by adhering to the following quality management principles[27].

(1) Customer focus. Organizations depend on their customers and therefore should understand current and future customer needs, should meet customer requirements and strive to exceed customer expectations. In today's economic activities, all organizations rely on their customers. Organizations or enterprises gain the power and source to survive by meeting customer needs or exceeding their expectations. Under the principle of customer focus, TQM is to meet the customer needs by means of continuous quality improvement in PDCA cycle[28].

(2) Leadership. Leaders establish unity of purpose and direction of the organization. They should create and maintain the internal environment in which people can become fully involved in achieving the organization's objectives. The entire enterprise, ranging from the general managers to the employees, shall participate in quality management activities. Most importantly, the decision makers of the enterprise must pay enough attention to quality management.

(3) Engagement of people. People at all levels are the essence of an organization and their full engagement enables their abilities to be used for the organization's benefit. Engagement of people, as a core of TQM, requires employees ranging from executives to front-line operators to engage in the implementation of TQM with a positive attitude.

(4) Process approach. A desired result is achieved more efficiently when managing activities and related resources are managed as a process, which means, the related resources and activities involved in TQM shall be managed as a process. PDCA cycle is actually used to study a process, so we must focus on the whole process of product production and quality management.

(5) System approach to management. An organization may achieve its objectives in a more effective and efficient manner by identifying, understanding and managing interrelated processes as a system. When carrying out a quality improvement activity, it is necessary to firstly establish, identify and determine objectives, and understand and manage a system consisting of interrelated processes in a unified manner. All departments of the organization shall participate in the product production to satisfy the customer needs to the maximum extent, since it is not just a matter for production departments.

(6) Improvement. Improvement of the organization's overall performance should be a permanent objective of the organization. In fact, it is not difficult to do just one thing right, but it is not easy to do a simple thing right thousands of times. Therefore, improvement is the core idea of TQM, and the application of statistical technology and computer technology is just for a better improvement.

(7) Evidence-based decision making. Effective decisions are based on the logical and intuitive analysis of data and information. Therefore, TQM, as the most scientific quality management by far, shall be based on facts, and it would be meaningless if deviated from the factual basis.

(8) Relationship management. An organization and its suppliers are interdependent and a mutually beneficial relationship enhances the ability of both to create value. Mutually beneficial relationships between them can enhance their ability to create value, thus laying a foundation for further cooperation and greater common interests.

1.2.5 Quality Management of Nanjing Iron and Steel Company

In this section, quality management information systems established by Nanjing Iron and Steel Company and their corresponding functions will be introduced. According to the total quality management system of iron and steel enterprises introduced in the first four sections of this chapter, Nanjing Iron and Steel Company has established some information systems based on its own business characteristics and formed a quality management system to control, warn and alarm the quality problems in the production process, so as to improve the product quality and competitiveness in the industry.

Nanjing Iron and Steel Company has established ERP (Enterprise Resource Planning) system, MES (Manufacturing Execution System), SCADA (Supervisory Control and Data Acquisition), SPC (Statistical Process Control) and other information systems, and initially formed a quality management system with PDCA (plandocheckact) as the core for product quality cycle control, so as to help the company achieve high-level quality management.

ERP is a planning, using the overall resource optimization technology with financial analysis and decision making as the core, that is used for production plan development, inventory control and financial management, with focus on the optimization of production organization, production management and business decision. In terms of software function composition, ERP includes main production plan, production operation plan, material requirements planning, sales management, purchase management, cost management, inventory management, financial management and production data management. ERP also have additional industrial functions, such as formula management, conversion of measurement units, operation management of related products and by-products process and maintenance management.

MES, an intermediate-level manufacturing execution system combing production and management, mainly uses advanced control technology with product quality and process requirements as indicators and improved operation, control and technology management for production process with comprehensive production indicators as targets. MES is mainly used for production management and scheduling, and improves manufacturing competitiveness and provides a unified platform by controlling all factory resources including materials, equipment, personnel, process instructions and facilities. MES enables to realize information conversion between ERP and process control, i.e., it implements the plan specified by ERP, adjusts production and makes scheduling based on real-time production information, and provides accurate information about resource utilization and inventory to ERP system in real time. MES can also automatically convert production objectives and production specifications into process settings. Besides, MES compares and analyzes the production data acquired from the control system with the quality indicators, and provides closed-loop quality control. It shows that MES serves as the information link between ERP and control system, and it is the core of industrial implementation.

SCADA, a process control system (PCS), is mainly used for monitoring the production process and the operating state of production equipment (usually specific equipment) in real time, and controlling the production process. SCADA is usually used for data acquisition, actual production feedback, etc.

SPC, abbreviation of statistical process control, is a method of measuring and controlling quality by monitoring production process. Quality data is used to evaluate, monitor and control the process[29] after being collected in the form of product or process measurements or readings of various machines or instruments. With the advancing information level, some iron and steel enterprises introduce SPC for product quality management and control. Quality managers can monitor the state of product manufacturing process with SPC system to ensure the production process is in the specified state, so as to reduce the possibility of product quality change.

Nanjing Iron and Steel Company has established a preliminary quality management system with PDCA (plandocheckact) as its core by introducing some of the above information management systems. This system has made certain contributions to the product quality control and improvement for Nanjing Iron and Steel Company, and provided customers with satisfactory steel products, thus improving the company's competitiveness in the field of iron and steel. However, due to some of its defects, this quality management system can't be used well. The defects of the quality system will be described in the following three points:

(1) SPC system puts forward higher requirements for users. Before SPC analysis, the standard value of the project shall be specified first. The accuracy of standard values will greatly affect that of SPC analysis, which puts forward higher requirements for the capacity of quality standard managers of enterprises. In terms of accuracy, SPC cannot fully adapt to the quality management of enterprises. For example, when the control value of on-site process parameters exceeds the predetermined target value, the PCS will

give the user an alarm that quality problems may (in fact, usually not) occur. When the on-site control value of the equipment does not exceed the predetermined target value, SPC system cannot give warning for product quality control, but may cause quality defects of finished products[30]. After Nanjing Iron and Steel Company introduced the SPC system, the production line workers and managers could not make full use of SPC tools for analysis and intervention on real-time process control ability. The statistics shows that SPC gets involved in the quality management at a rate of 50%-75% throughout the process, which deviates from the original intention of whole-process control and thus causing unstable quality of some products.

(2) A large quantity of quality-related data has been accumulated in the above-mentioned quality management system, but the accumulated data are still far from effective utilization. Data analysis is still in the off-line analysis stage, and the data use is far from meeting the requirements of quality control in production process[31]. The quality management of iron and steel products still stays in the post-control stage, mainly relying on the final sampling of products after the completion of the process to conduct experiments or tests so as to judge whether there are quality problems. If the experimental data or test results meet the user's requirements, the products will be produced as planned, otherwise, they need be reprocessed in the subsequent process, and the quality of products cannot be predicted on-site in real time. Take the tensile test as an example: when the steel plate is cut by the manufacturer, the production personnel will cut the sample based on the specified time and shape according to the specifications and instructions. After collecting a certain number of samples, the sample delivery personnel group the samples and prepare a power of attorney. Sample delivery personnel deliver samples once every 4 hours. After the experimental samples arrive at the experimental testing center, the experimental testing center schedules and tests the experimental samples according to the experimental items and requirements and with reference to the power of attorney. After the samples are collected, mark the sample number on the samples, arrange the processing plan according to the processing conditions of the test items, and the processing workshop conducts processing according to the processing plan. Ordinary steel is milled to the specified thickness by a double-sided milling machine, and then the sample is formed by a forming machine. After

forming, mark the specific No. on the sample. The wear-resistant steel is cut by sawing machine and other professional equipment. After being cut, it is put on a grinder for grinding, and then processed by a double-sided milling machine. After being processed, the samples are sent to the mechanics laboratory, where they are stretched to fracture by a stretcher, and the experimental results are collected systematically. After the test is completed, the system will determine whether the product meets the specification requirements according to the test results.

(3) A lot of quality data are collected in each system, including the quality information in product manufacturing process, process parameter information and the quality of each product, which form the whole life cycle of the corresponding product. However, with different responsibilities and functions in different information systems, each information system is only responsible for storing data with corresponding functions. The data of different information systems are independent from each other to form data islands, and the data cannot be well utilized for quality analysis, management and control.

1.3 Introduction to AI Development

Artificial Intelligence (AI), as an important branch of computer science, is currently known as one of the world's leading technologies. The 1950s and early 1960s were the initial stages of ai development, and research during this period focused on the adoption of heuristic thinking and the application of domain knowledge. In "Computing machinery and intelligence"[32] by Alan Turing, the possibility of mechanization of human intelligence was discussed and the theoretical model of Turing machine was proposed, which laid the theoretical foundation for the emergence of modern computers. At the same time, the article also proposed the famous Turing criterion, which has now become the most important intelligent machine standard in the field of artificial intelligence research. In literature [33] the authors proved that a certain type of neural network that can be strictly defined can in principle be able to calculate a certain type of logical function. There are two major categories of current artificial intelligence research: symbolism and joint conclusions. Since 1963, people began to use natural language communication, which marked another leap in artificial intelligence. How to make computers understand natural language, automatically answer questions, analyze images or graphics, etc. has become an important goal of AI research. Then AI research entered the second stage. In the 1970s, after a lot of explorations on the scientific reasoning of human experts, a batch of expert-level program systems came out one after another. The knowledge expert system has been developed rapidly all over the world, and its application range extends to all fields of mankind, and has produced huge economic benefits. In the 1980s, AI entered the stage of knowledge-based development. More and more people realized the importance of knowledge in simulated intelligence, and made more in-depth exploration around knowledge representation, reasoning, machine learning, and new cognitive simulation combining problem domain knowledge[34].

The research and application of artificial intelligence mainly include three aspects: expert system, machine learning and artificial neural network.

Expert System (ES) is an important branch in the field of artificial intelligence. It transforms the general thinking method into the application of specialized knowledge to solve specialized problems, and realizes a major breakthrough in the development of artificial intelligence from theoretical research to practical application. Expert system can be regarded as a kind of computer intelligent program system with specialized knowledge. It can use the specialized knowledge and experience provided by experts in a specific field, and use the reasoning technology in artificial intelligence to solve and simulate various complex problems that can only be solved by experts. Generally speaking, the expert system is a kind of intelligent software using heuristic method to solve problems, which generally have no algorithmic solution. Herein, the difference with traditional computer programs is that it often requires drawing conclusions based on incomplete, inaccurate or uncertain information. In different fields and types, the architecture and function of expert systems are consequently different, but their composition principle is basically the same.

Machine Learning (ML) gives computers the capability to learn without being explicitly programmed. ML is one of the most exciting technologies that one would have ever come across. As its name implies, it enables the computer to think and act more similar to humans: having the ability to learn[35]. Machine learning is to program computers to optimize performance criteria using example data or past experience. It has to firstly define a machine learning model with a bunch of hyperparameters to be fine-tuned. The learning process means that the parameters of the learnt model is optimised by training against the historical data in an iterative way. The learnt model shall thus obtain the capability to make predictions for unseen data[36].

Machine learning applies the theory of statistics in building mathematical models, because the core task is making inferences based on a sample. The role of computer science is twofold: First, in training, we need efficient algorithms to solve the optimization problem, as well as to store and process the massive amount of data we generally have. Second, once a modelis learned, its representation and algorithmic solution for inference need to be efficient as well. In certain applications, the efficiency of the learning or inference algorithm, namely, its space and time complexity, are as important as its predictive accuracy[36].

Artificial neural network (ANN) is a network consisting of a large number of processing units, namely neurons, which is often referred to as neural network[37]. A neural network is an operational model composed of a large number of nodes (or neurons) and their interconnections. It is an abstraction and simulation of some basic characteristics of human brain or natural neural network. Its purpose is to simulate some mechanisms and mechanisms of the brain, so as to realize some functions. Generally speaking, artificial neural network is the result of simulating biological neural network. To be more detailed, artificial neural network is a technique to obtain the solution of a specific problem, according to the biological neural network mechanism mastered, according to the idea of control engineering and mathematical description method, establish the corresponding mathematical model and adopt the appropriate algorithm, and determine the parameters of the mathematical model[38]. Artificial neural network has a strong self-learning ability, it does not rely on the "expert" mind, while automatically summarizing the rules from the existing experimental data. As a result, artificial neural networks are good at dealing with complex and multi-dimensional nonlinear problems. They can not only solve qualitative problems, but also quantitative problems. At the same time, they also have large-scale parallel processing and distributed information storage capabilities. Organizational and strong learning, association, fault tolerance and better reliability[39].

1.4 The Research Goals of This Thesis

Section 1.2.5 introduces the quality management system that Nanjing Iron and Steel Company has established, which has made a certain contribution to the improvement of the product quality of Nanjing Iron and Steel Company and has improved the company's competitiveness in the industry. However, this quality management system also has some shortcomings, which make it not able to be applied well, for example: (1) It has high requirements on the data statistical basis of business personnel, which leads to its limited utilization rate. (2) A lot of quality data has been collected in each system. In the past, it was mainly focused on the post-analysis and processing of the data, and the quality problems of the products were mainly tested by sampling-experiment method, so that the quality of the products could not be tested in time or predicted in advance. The sampling-experiment method has a long detection cycle and high cost. (3) In the quality control system, different information systems have different responsibilities and functions. Each information system is only responsible for storing the data of the corresponding function. The data between different information systems are independent of each other to form data islands. The iron and steel production process belongs to the process industry, and the data in multiple information systems are combined with each other to conduct in-depth analysis, prediction and early warning of product quality. Therefore, it is necessary to introduce some new product quality control methods in the iron and steel industry.

In recent years, Industry 4.0 based on the deep integration of the Industrial Internet and the manufacturing industry has developed rapidly[40]. Some countries have already released their intelligent manufacturing plans. Therefore, intelligent manufacturing has become the core of all industries in the world, and it will also affect the future of a country. The implementation of intelligent manufacturing will also bring huge profits to enterprises. It can not only improve the automation and information level of enterprises, but also help reduce the number of personnel and improve efficiency. Most importantly, it helps to improve the quality and reduce the cost of products[41].

Quality control is an important part of intelligent manufacturing and also the most critical link in enterprise production[42], especially in the iron and steel industry. The iron and steel industry are a process industry. In the production process, each production line is continuously running, and each has a relatively large production scale. Therefore, when there is a quality problem, the same batch of products will have similar quality problems, and there will also be greater economic losses due to continuous production. Quality control in the iron and steel industry is thus an important aspect of intelligent manufacturing. First of all, good product quality management determines the reputation of an enterprise. Therefore, quality control methods have always been the focus of enterprises research. With the promotion of Industry 4.0 and intelligent manufacturing, the iron and steel enterprises have also introduced some new quality control methods[43] to improve their competitiveness in the industry, which are mainly based on big data analysis technology and AI technology.

At the same time, the data accumulated in the industry is expanding rapidly and growing in terms of scale. At present, big data technology has been widely used in e-commerce, O2O, logistics and other fields. The use of this new technology can help enterprises develop new businesses and improve their innovation capabilities[44]. The application of big data analysis methods in quality control will enable enterprises to generate accurate quality analysis and good predictions based on the data distributed throughout the factory.

The ERP, MES, SCADA, SPC and other systems that Nanjing Iron and Steel Company has established as introduced in Section 1.2.5 have accumulated a large amount of data related to the product life cycle, including order data, quality information during product manufacturing, price and quality information of raw materials, process parameter information, and energy consumption information during production. These data form the full life cycle of the corresponding iron and steel products. With these data, a quality data warehouse can be created to extract quality data from various systems and to break data islands, and then the data are verified, cleaned, summarized and modeled with the data mining technology, and an online quality prediction model can thus be established to help product quality detection[45].

Therefore, big data technology and machine learning algorithms realize online adjustment of production process by connecting all these data together to avoid unqualified products[46]. This is not only conducive to the positive significance and effect of enterprises product quality control, but also conducive to reducing product costs.

So far, in the era of Industry 4.0, it has become the focus of attention of enterprises to use the data accumulated in enterprises by traditional information systems to guide the production[47]. In this regard, it is considered to a very promising method to collect data from traditional information systems and analyze such data to control product quality and cost.

Based on the shortcomings and pain points of traditional quality control systems and the advantages of new methods based on big data analysis in the context of intelligent manufacturing, and on the basis of a large amount of data related to the product life cycle, this thesis, by taking the iron and steel manufacturing industry as the research object, discusses the cost control and quality improvement of three links in iron and steel production process by machine learning and deep learning according to the sequence of production processes. The main content of this thesis covers: 1) intelligent rating and acceptance of the quality of scrap steel as raw material for steelmaking, improving the quality of each grade of scrap steel, and reducing the acquisition cost of scrap steel; 2) predicting the longitudinal cracks on the surface of continuous casting plate blank, and giving warning for the casting slab with quality problems to avoid the casting slab with longitudinal cracks entering the later production link; 3) predicting the mechanical properties of steel plates, establishing an online product prediction system, and improving the traditional sampling-experiment method for performance test to improve the speed and reduce the cost of steel plate performance test.

Under the background of Industry 4.0 and Made in China 2025, this research focuses on the application of AI algorithm in the iron and steel quality field to solve the pain points in some traditional quality management and control systems, so as to provide new ideas and methods for quality prediction, management and control and cost control of iron and steel products. The successful operation of the system will bring significant benefits to the iron and steel industry in reducing direct investment. Moreover, it will also greatly improve the efficiency of the industry.

1.5 Structure of This Thesis

Chapter 1 introduces the research background of this thesis.

Chapter 2 introduces AI technology and its application in iron and steel quality control and other fields.

Chapter 3 introduces the scrap steel intelligent rating system.

Chapter 4 introduces the prediction system of longitudinal crack on continuous casting slab surface.

Chapter 5 introduces the prediction methods for mechanical properties of steel plates based on integrated learning.

Chapter 6 reviews the research content of this thesis and looks forward to the future research content

1.6 Summary

This chapter firstly introduces the main processes of steel smelting and the main functions of each process in details, followed by the comprehensive quality management system of steel enterprises, the quality management system already established by Nanjing Iron and Steel Company and its shortcomings, as well as a brief introduction to artificial intelligence. Lastly, the objectives of the research conducted in this thesis and the overall structure of the thesis are illustrated.

Chapter 2

Literature Review

The concept of artificial intelligence (AI) was put forward by computer expert John Maccarthy at a conference held by Dartmouth College in 1956. So far, the development of AI has experienced two rises and falls. This stage is the third rise of AI. In the first peak of AI, computers were widely used in the fields of mathematics and natural language to solve algebra, geometry, and English problems. AI ushered in its first trough as the problems such imperfect performance, insufficient data and high complexity of computers could not be solved. At the second peak, an expert system combining "knowledge base and inference engine" was mainly proposed, which could solve somecommercialization problems at that time. By 1987, when the performance of desktop computers produced by Apple and IBM surpassed that of the inference engine used by the expert system, AI ushered in the second trough. Since the mid-1990s, as the AI technology, especially the neural network technology, was developing gradually, and people began to have an objective and rational cognition of AI, AI technology began to enter a period of steady development. In 2006, Hinton made a breakthrough in the field of deep learning of neural networks, enabling mankind once again to see the hope of machines catching up with mankind, which was also a landmark technological advancement.

In the development of AI, AI methods have been applied to a certain extent. Of course, AI methods also have been widely applied in the field of iron and steel[48]. Following the development of AI, this chapter will introduce some applications of AI methods in the field of iron and steel. Section 2.1 introduces the expert system and its application in the iron and steel industry; Section 2.2 introduces the machine learning algorithms and their applications in the iron and steel industry; Section 2.3 introduces the deep learning and its application in the iron and steel industry; and Section 2.4 summarizes the chapter.

2.1 Expert System and its Application in the Steel Industry

An expert system refers to a large quantity of expert-level knowledge and experience in one or more fields contained in the computer, based on which the decision-making process is simulated using computer to perform inference and judgment. The expert system includes six parts: human-computer interaction interface, knowledge base, inference engine, interpreter, comprehensive database, and knowledge acquisition, and among them, the knowledge base and inference engine are two important links[49]. The knowledge base is used to store the knowledge provided by the expert, and to simulate the empirical knowledge of the model required by the expert depends on the knowledge base. The inference engine is based on the knowledge in the knowledge base and the problem conditions that need to be solved or known information to simulate the way experts think about and solve problems and match the corresponding rules. With the development of the information age, the application fields of expert systems are becoming wider and wider, and the diagnostic accuracy in some fields even exceeds that of human experts[49].

The role of expert system in the field of steel is also very important. Kawasaki Chiba Plant in Japan has developed a heating furnace combustion control expert system with which to select the objective function and constraint conditions of the optimization model and determine the set temperature[50]. There are solids, liquids, and gases in the blast furnace, and the chemical reactions are complex, so it is difficult to obtain an accurate mathematical model. Therefore, the blast furnace operation expert system is adopted in the development process. According to some characteristics of steel production companies, literature [51] applied the Multi-agent expert system to model the electric furnace steelmaking and control the steelmaking process. In the article, the knowledge base of agents is generated by an adaptive neuro-fuzzy inference system. Literature [52] constructed a set of expert system for LD process steelmaking, this system can intelligently calculate the charge and interactively guide the LD process operators to complete a complete cycle. Aiming at the Decarburization-Annealing Furnace of Tandem Annealing Line of Rourkela Steel Plant, Somnath Mitra et al.developed a heating control system based on a closed-loop expert system, which is improving product quality and reducing energy[53]. Significantly improved plant performance in terms of consumption. Aiming at the problem that it is difficult to detect the particle size of pulverized coal online, the literature[54] proposed a soft measurement method of pulverized coal particle size based on case-based reasoning technology, which acquires knowledge from a large amount of historical data accumulated in the process of pulverized coal production. A significant effect has been achieved. It can be seen that the application of the expert system has improved the automation level of steel production.

2.2 Machine Learning and its Application in Iron and Steel Industry

2.2.1 Introduction to Machine Learning Algorithms

Machine learning is a multi-field interdisciplinary subject, involving multiple subjects such as probability theory, statistics, approximation theory, convex analysis, and algorithm complexity theory[50]. It is specialized in the study of how computers simulate or realize human learning behaviors in order to acquire new knowledge or skills and reorganize the existing knowledge structure to continuously improve its performance.

Machine learning mainly includes four learning methods: supervised learning, unsupervised learning, semi-supervised learning and reinforcement learning. Supervised learning is to build a model and learn under the condition of training samples, compare the prediction results with the actual results of the training data, and continuously adjust the prediction model until the prediction result of the model reaches an expected accuracy rate. Both classification and regression belong to supervised learning. The difference lies in the type of output variable. Quantitative output is called regression, or the prediction of continuous variables; the qualitative output is called classification, or the prediction of discrete variable. First of all, this chapter will introduce the machine learning algorithms used in this thesis, such as linear model, support vector machine

(SVM), kernel ridge regression and k-nearest neighbor.

2.2.1.1 Linear Model

The linear model is simple and easy to construct, but it contains some important basic ideas in machine learning algorithms. Introducing hierarchical structure or high-dimensional mapping on the basis of linear model can get a more powerful nonlinear model. In addition, the vector form of the linear model is $f(x) = \theta^T X + b$, parameter θ indicates the importance of each attribute in the regression process, and the linear model has good interpretability[55].

When using machine learning methods to make predictions, the trained model often has a good effect on the train set, but the effect on the test set is poor. This phenomenon is called overfitting. In order to prevent overfitting, scholars have optimized linear regression and produced lasso[56], ridge[57] and ElasticNet[58] regression. In short, the main difference between these three types of regression and linear regression is that the complexity penalty factor is added in the objective function $J(\theta)$, that is, the regularization term. These three regressions respectively use L1 norm, L2 norm and a combination of L1 norm and L2 norm.

Among them, the L1 norm is used for regularization in LASSO, and the objective function calculation equation is as follows:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(h_\theta(x_i) - y_i)^2 + \lambda \sum_{j=1}^{n} |\theta_j|. \quad \text{(Equation 2.1)}$$

In Equation 2.1, the first term is the objective function of the linear model, and the second is the L1 regular term. By determining λ, a non-negative coefficient, can make the model reach a balance between variance and bias. θ_j is the weight in linear regression.

The L2 norm is used for regularization in Ridge, and the objective function calculation equation is as follows:

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{n}(h_\theta(x_i) - y_i)^2 + \lambda \sum_{j=1}^{n} {\theta_j}^2, \quad \text{(Equation 2.2)}$$

where the L2 norm is used for the regularization of the second term, and the other parameters have the same meaning as in Equation 2.1.

Elastic Net: L1 norm and L2 norm are used in combination for regularization, and the objective function calculation equation is as follows:

$$J(\theta) = \sum_{i=1}^{n}(h_\theta(x_i) - y_i)^2 + \lambda_1 \sum_{i=1}^{n} {\theta_i}^2 + \lambda_2 \sum_{i=1}^{n} |\theta_j|. \qquad \text{(Equation 2.3)}$$

The meanings of the parameters in Equation 2.3 are the same as those in Equation 2.2 and Equation 2.1.

2.2.1.2 Support Vector Machine

Support vector machines (SVM)[59] were first proposed by Vapnik. Like multi-layer perceptron networks and radial basis function networks, support vector machines can be used for pattern classification and nonlinear regression. The main idea of support vector machine is to establish a classification hyperplane as a decision surface to maximize the isolation edge between positive and negative examples; the theoretical basis of support vector machine is statistical theory, more precisely, support vector machine is approximate realization of structural risk minimization. This principle is based on the fact that the error rate (generalization error rate) of the learning machine on the test data set is bounded by the sum of the training error rate and a term that depends on the VC dimension (Vapnik-Chervonenkis dimension). In the separable case, the support vector machine has a value of zero for the previous term and minimizes the second term. Therefore, support vector machines can provide good generalization performance in the pattern classification problem, though it does not use the internal problems in the problem domain.

The basic model of the support vector machine is a linear classifier with the largest margin defined in the feature space. The largest margin makes it different from the perceptron. Support vector machine also includes kernel trick. Kernel can be said to be the essence of support vector machine. Its purpose is to make the original linearly inseparable samples become linearly separable in the new kernel space by mapping the original input space to the high-dimensional feature space, which makes it a substantial non-linear classifier, in which the Radial Basis Function (RBF) (kernel function) is more widely used[60]. The learning strategy of support vector machines is to maximize the margin, which can be formalized as a problem of solving convex quadratic programming[61]. It is also equivalent to the problem of minimizing the regularized hinge loss function.

Support vector machine learning methods include constructing models from simple to complex: linear support vector machine in linearly separable case, linear support vector machine, and non-linear support vector machine[62]. The simple model is the basis of the complex model, also a special case of the complex model. When the training data is linearly separable, a linear classifier, that is, linear support vector machine in linearly separable case, also known as hard margin support vector machine, is learnt through hard margin maximization; when the training data is approximately linearly separable, a linear classifier, that is, linear support vector machine, also known as soft margin support vector machine, is also learnt through soft margin maximization; when the training data is linearly inseparable, non-linear support vector machine is learnt by using the kernel function and soft margin maximization.

When the input space is Euclidean space or isolated set and the feature space is Hilbert space, the kernel function represents that the inner product between the feature vectors obtained by mapping the input space to the feature space is to be input. Non-linear support vector machine can be learned by using kernel functions, which is equivalent to implicitly learning linear support vector machines in a high-dimensional feature space.

The support vector machine algorithm has the following advantages:

(1) Versatility: able to construct functions in a wide variety of function sets;

(2) Robustness: no need for fine-tuning;

(3) Effectiveness: it is always one of the best methods in solving practical problems;

(4) Simple calculation: the realization of the method only needs to use simple optimization techniques;

(5) Perfection in theory: based on the framework of VC generalization theory.

Support vector machines can be classified into two-class support vector machines and multi-class support vector machines[63], with the definitions as follows:

1. Two-class support vector machine

The specific form of the common two-class support vector machine model is as follows:

1). Let the known train set be: $T = \{(x_1, y_1), \cdots, (x_l, y_l)\} \in (X \times Y)^l$, where $x_i \in X \in R^n$, $y_i \in Y = \{1, -1\}, i = 1, 2, \cdots, l$; x_i is the feature vectors.

2). Select the appropriate kernel function $K(x, x^{'})$ and the appropriate parameter C, construct and solve the optimization problem:

$$\begin{cases} \min\limits_{\alpha} \quad \dfrac{1}{2}\sum\limits_{i=1}^{j}\sum\limits_{j=1}^{l} y_i y_j \alpha_i \alpha_j K(x, x^{'}) - \sum\limits_{j=1}^{l} \alpha_j \\ s.t. \ \sum\limits_{i=1}^{l} y_i \alpha_i = 0, 0 \leq \alpha_i \leq C, i = 1, 2, \cdots, l \end{cases} \qquad \text{(Equation 2.4)}$$

Get the optimal solution: $\alpha^* = (\alpha_1^*, \cdots, \alpha_l^*)^T$.

3). Select a positive component of α^*, and calculate the threshold based on this:

$$b^* = y_j - \sum_{i=1}^{l} y_i \alpha_i^* K(x_i - x_j. \qquad \text{(Equation 2.5)}$$

4). Construct a decision function:

$$f(x) = sign\left(\sum_{i=1}^{l} y_i \alpha_i^* K(x, x_i) + b^*\right). \qquad \text{(Equation 2.6)}$$

2. Multi-class support vector machine (SVM)

The support vector machine algorithm was originally designed for two-class classification problems. To deal with multi-class classification problems, it is necessary to construct suitable multi-class classifiers. At present, there are two main types of methods for constructing multi-class SVM classifiers: one is the direct method, used to modify the objective function directly, with the parametric solutions for multiple classifying planes combined into an optimization problem, and realize multiple classifications at a single time by solving this optimization problem. This type of method seems simple, but its computational complexity is relatively high, and it is relatively difficult to implement, so it is only suitable for small-scale problems. The other type is the indirect methods, used to which implement the construction of multi-class classifiers mainly by combining multiple two-class classifiers, with one-to-many and one-to-one methods as common.

(1) One-to-many method: during training, samples of a certain class are classified into one class in sequence, and the remaining samples into another class, so that samples of k (number) classes construct k support vector machines. During classification, classify the unknown samples into the class with the largest value of classification function.

(2) One-to-one method: in this method, a support vector machine is designed between any two classes of samples, so $k(k-1)/2$ support vector machines need to be designed for these k classes of samples. When an unknown sample is classified, the class with the most votes is the class of the unknown sample.

(3) Hierarchical support vector machine: in the hierarchical classification method, each class is first divided into two sub-classes, which then are further divided into two secondary sub-classes, and so on until a single class is obtained.

2.2.1.3 Kernel Ridge Regression

Kernel ridge regression[64] refers to ridge regression with kernel techniques. The principle is to map the features of the data to a higher-dimensional space through a function, so that the number of data samples or features can be reversed.

Firstly, a nonlinear mapping function Φ is defined to map the low-dimensional space data into the high-dimensional space, and then the linear regression method is applied in the high-dimensional space to solve the nonlinear regression problem. Replace X with $\Phi(X)$ to get the objective function as Equation 2.7.

$$J(\theta) = \|y - \Phi(X)\theta\|^2 + \lambda \|\theta\|^2 . \qquad \text{(Equation 2.7)}$$

The solution of the parameter θ is expressed as Equation 2.8.

$$\theta = (\Phi(X)^T \Phi(X) + \lambda I)^{-1} \Phi(X)^T y, \qquad \text{(Equation 2.8)}$$

where I is the identity matrix.

Usually the nonlinear mapping function is unknown, but we can obtain the inner product by defining the kernel function.The kernel function is defined as follows:

$$K(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle = \Phi(x_i)\Phi(x_j)^T. \quad \text{(Equation 2.9)}$$

We now define $\alpha = \lambda^{-1}(y - \Phi(X)\theta)$, $\mathcal{K} = \Phi(X)\Phi(X)^T$, and we get the following result:

$$\alpha = (\mathcal{K} + \lambda I)^{-1} y. \quad \text{(Equation 2.10)}$$

Therefore, the prediction formula of kernel ridge regression is Equation 2.11.

$$\hat{y}_k = K(x_k, X)(\mathcal{K} + \lambda I)^{-1} y. \quad \text{(Equation 2.11)}$$

In comparison with kernel ridge regression, the shortcomings of linear regression are relatively limited, but from the perspective of computational complexity, when the amount of data N is much larger than the number of dimensions d, the linear regression has higher computational efficiency. The kernel ridge regression is flexible due to the use of kernel techniques and is suitable for complex fitting. We have found that the computational complexity of this method is related to the amount of data. For a large amount of data, this method is not suitable. Therefore, the difference between the linear method and the kernel method lies in the balance and trade-off[64] between the computational efficiency and flexibility of complex problems.

2.2.1.4 K-Nearest Neighbor (KNN)

k-nearest neighbor (KNN)[65] is a very basic classification and regression method. The main difference between KNN classification and regression lies in the different decision-making methods during prediction. When KNN is used to perform classification prediction, the majority voting method is usually selected, and the nearest k samples in the train set and the predicted sample features have the largest number of classes. When KNN is used to perform regression, the average method is generally selected, that is, the average value of the sample output of the nearest k samples is used as the regression prediction value[66].

In the KNN algorithm, three important factors are mainly considered: the selection of the value of k, the measurement method of distance, and the classification decision rule.

The selection of the value of k will have a significant impact on the results of the k-nearest neighbor method. If a smaller value of k is selected, it is equivalent to predicting with the training instance in the smaller neighboring regions, the approximation error in "learning" will be reduced, and only the training instance that is closer (similar) to the input instance will have an effect on the prediction result. But the disadvantage is that the estimation error in "learning" will increase, and the prediction result will be very sensitive to the neighboring instance points. If the neighboring instance points happen to be noise, the prediction will be wrong. In other words, the decrease of the value of k means that the overall model becomes complicated, and it is prone to overfitting.

If a larger value of k is selected, it is equivalent to predicting with the training instances in the larger neighboring regions. The advantage is that the estimation error in learning can be reduced, but the disadvantage is that the approximate error in learning will increase. At this time, the training instance far away from (dissimilar to) the input instance will also contribute to the prediction, making the prediction wrong. The increase of the value of k means that the overall model becomes simple.

In the case that $k = N$ (N is the number of training instance samples), then no matter what the input instance is, it will simply be predicted that it belongs to the class with the most training instances. At this time, the model is too simple, and it is not advisable to completely ignore the large amount of useful information in the training instances. In application, a relatively small value of k is generally taken, and the cross-validation method is usually used to select the optimal value of k.

The distance between two instance points in the feature space reflects the similarity between the two instance points. The feature space of the k-nearest neighbor model is generally an n-dimensional real number vector space R^n. The distance used is usually the Euclidean distance, but it can also be any other distance, such as a more general L_p distance or Minkowski distance.

Let the feature space X be an n-dimensional real number vector space R^n, $x_i, x_j \in X$, $x_i = (x_i^1, x_i^2, \cdots, x_i^n)^T$, $x_j = (x_j^1, x_j^2, \cdots, x_j^n)^T$, and the L_p distance of x_i, x_j is defined as

$$L_p(x_i, x_j) = \left(\sum_{l=1}^{n} \left| x_i^l - x_j^l \right|^p \right)^{\frac{1}{p}}, \tag{Equation 2.12}$$

where $p \geq 1$. x_i^l and x_j^l are the l-th component of vector x_i and vector x_j respectively.

when $p = 2$, the distance is the Euclidean distance, namely

$$L_2(x_i, x_j) = \left(\sum_{l=1}^{n} \left| x_i^l - x_j^l \right|^2 \right)^{\frac{1}{2}}, \tag{Equation 2.13}$$

when $p = 1$, it is called Manhattan distance, namely

$$L_1(x_i, x_j) = \sum_{l=1}^{n} \left| x_i^l - x_j^l \right|, \tag{Equation 2.14}$$

when $p = \infty$, it is called Chebyshev distance, it is the maximum value of each coordinate distance, namely

$$L_\infty(x_i, x_j) = \max_l \left| x_i^l - x_j^l \right|, \tag{Equation 2.15}$$

In the k-nearest neighbor algorithm used for a classification problem, the decision rule is usually to perform the principle of majority voting after k samples are taken: the class with the most occurrences is that of the sample to be tested. The principle of weighted voting may also be adopted as the decision rule: the closer to the sample to be tested a sample is, the greater its weight. After the weights of all classes are summed, the class with the largest weight sum is selected. In the k-nearest neighbor algorithm used for a regression problem, the decision rule is usually to obtain the average or weighted average of the results after k samples are taken.

2.2.1.5 Integrated Algorithm

Integrated learning is a learning algorithm that combines a variety of different machine learning models to improve the prediction performance[67, 68]. A set of machine learning models can be called basic learners or weak learners[69]. In general, multiple learners in integrated learning are homogeneous “weak learners”. Generally, integrated learning can help achieve stronger generalization capabilities than a single model[69, 70]. Bagging, Boosting, and stacking are the three most common methods used in integrated learning to solve different problems[67]. These algorithms balance the deviation, reduce the variance, and are not easy to cause overfitting[55]. In addition to the above three methods, there are Voting, Averaging, etc.[67, 69], with the details described as follows.

1) Voting

Voting can also be seen as a model fusion, a method of voting systems. For example, in the classification, the one with more votes is determined to be the final classification.

2) Averaging

Averaging the results of model fusion by averaging or weighting the results of the model fusion has produced good results on many machine learning problems and different evaluation standards. For regression problems, a simple and straightforward idea is averaging. A slightly improved method is to obtain the weighted average. The weights can be determined by sorting.

3) Bagging

Bagging (Bootstrap aggregating) algorithm is the earliest integrated learning algorithm. It is an integration technology that trains the classifiers by reselecting S (number) new data sets through random sampling with replacement on the original data set, that is to say, these new data sets allow duplication. Use the trained classifier set to classify the new samples, and then count the classification results from all classifiers through the majority voting or averaging of the output. The class with the highest result is taken as the final label. In order to improve the variance of the model, bagging is used to randomly extract data from the train set when each model to be combined is trained.

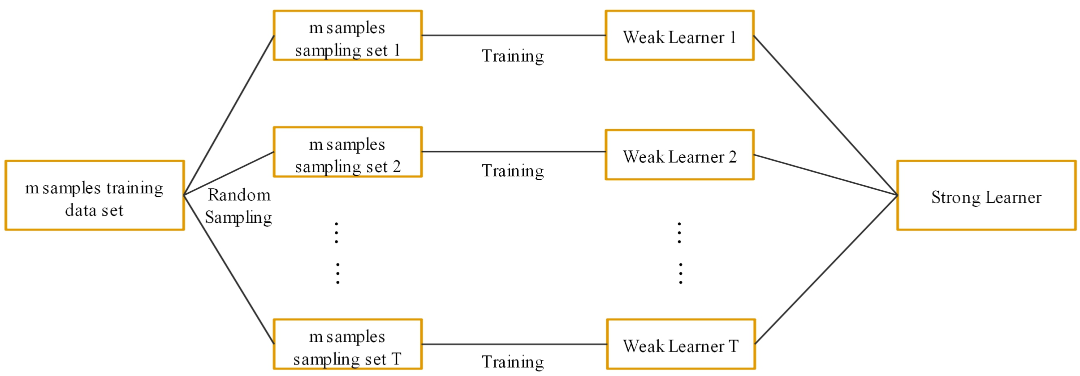


Figure 2.1: Bagging algorithm principle.

The principle of bagging algorithm is as shown in Figure 2.1. In the bagging algorithm, samples of the same number of the train set samples, m, will generally be randomly taken through random resampling with replacement. In this way, although there are the same number of samples obtained for the sample set and the train set, the contents of these samples are different. If the train set with m samples is randomly sampled T (number) times, the T sample sets are different due to randomness. The sampling method used by the bagging algorithm is random sampling with replacement. The so-called random sampling with replacement (bootstrap) is to collect a fixed number of samples from the train set, but each sample will be replaced after being collected. In other words, the samples collected before may be collected again after being replaced.

The typical algorithm of bagging is random forest, which is a typical example based on bagging algorithm. The basic classifier used is a decision tree, which will be described in detail later.

4) Boosting

The boosting algorithm allows parallel processing, and it is an iterative method. For example, there are n (number) points in the train set. We can assign a weight to each point in the train set W_i, $(0 \leq i<n)$ to show the importance of this point. At the beginning, each weight is the same. Through training, the weights of the points are modified. If the classification is correct, the weight will decrease, and if the classification is wrong, the weight will increase. All the procedures are executed to obtain M (number) models, corresponding to $y_1(x), \cdots, y_M(x)$ in the following figure, and combine them into the final model by weighting $y_M(x)$.

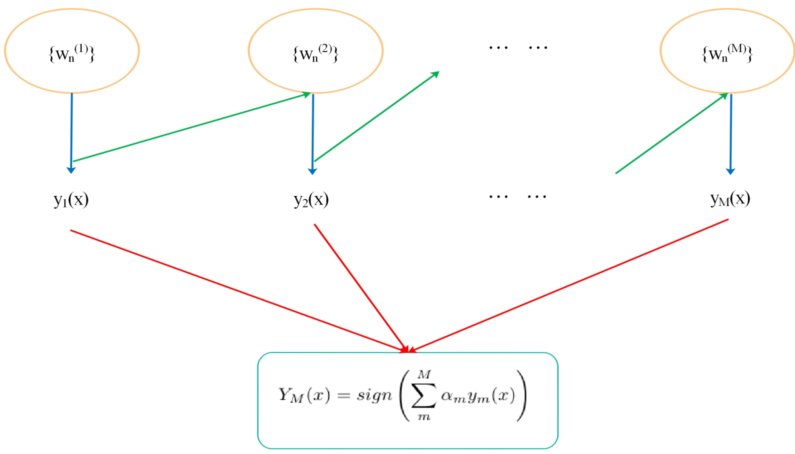


Figure 2.2: Boosting algorithm.

5) Stacking

Stacking is used to enhance prediction accuracy and robustness by combining a series of low-level individual learners as the input into the meta-learner[71, 72].

Stacking principle: all training base models are used to predict the entire train set. The predicted value of the j-based model of the sample i will be used as the feature value j of a new train set of the sample i. Finally, the value is trained based on the new train set. Similarly, a new test set also needs to be generated through all the basic model predictions in the prediction process, and finally the prediction is made on the test set.

2.2.1.6 Random Forest

The random forest algorithm was proposed by Leo Breiman and Adele Cutler[73], which combined Breimans's "Boot-strap aggregation" idea and Ho's "random subspace" method. Its essence is a classifier which contains many decision trees which are formed by random methods, so they are also called random decision trees. There is no correlation between trees in random forest. In fact, when the test data enter the random forest, they are classified by each decision tree, and the final result is the one with the most classification results in all decision trees[74]. This voting mechanism can avoid the overfitting of decision trees. Therefore, the random forest is a classifier containing multiple decision trees, and its output class is determined by the mode of the output class of the individual tree.

Similar to Bagging algorithm, Random forest algorithm is based on resampling specified in Bootstrap method to generate multiple train sets. The difference is that random forest algorithm adopts the method of randomly selecting split attribute sets when constructing the decision tree. The detailed flow of random forest algorithm is as follows:

(1) resample by Bootstrap method to generate T train sets randomly $S_1, S_2, \cdots, S_T$.

(2) Use each train set to generate corresponding decision tree $C_1, C_2, \cdots, C_T$. Before selecting attributes on each non-leaf node (internal node), m attributes are randomly selected from M attributes as the split attribute set of the current node, and the node is split in the best split mode among these m attributes (generally speaking, the value of m remains unchanged during the whole forest growth process).

(3) Every tree grows completely without pruning.

(4) The test set sample X is tested by each decision tree to get the corresponding class $C_1(x), \cdots, C_T(x)$

(5) Use voting method, and take the class with the most output in T decision trees as the class to which test set sample X belongs.

2.2.1.7 Introduction to Gradient Boosting Decision Tree Algorithm

GBDT (Gradient Boosting Decision Tree)[75] is also called MART (Multiple Additive Regression Tree), and GBRT (Gradient Boosting Regression Tree) has good effects on many data. At the beginning, it was considered as an algorithm with strong generalization ability together with support vector machine [76].

Assume that the input sample is $T = \{(x_1, y_1), (x_2, y_2) \cdots, (x_m, y_m)\}$, $x_i \in X \subseteq R^n$, $y_i \in Y \subseteq R$, the maximum number of iterations is T, the loss function is $L(y, f(x))$, and $f(x)$ is the output strong learner.

GBDT algorithm mainly includes the following three steps:

(1) initialize the weak learner

$$f_0(x) = arg\min_c \sum_{i=1}^{m} L(y_i, c), \qquad \text{(Equation 2.16)}$$

where c is a constant.

(2) for $t = 1, 2, \cdots, T$

(a) for $i = 1, 2, \cdots, m$

Calculate the negative gradient:

$$r_{ti} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{t-1}(x)}, \qquad \text{(Equation 2.17)}$$

(b) Use $(x_i, r_{ti})(i = 1, 2, \cdots, m)$ to fit CART regression tree to obtain the t-th regression tree. The corresponding leaf node area is R_{tj}, $j = 1, 2, \cdots, J$, where J is the number of leaf nodes in the regression tree t.

(c) Calculate the best fitting value of the leaf region $j = 1, 2, \cdots, J$:

$$c_{tj} = arg\min_c \sum_{x_i \in R_{tj}} L(y_i, f_{t-1}(x_i) + c), \qquad \text{(Equation 2.18)}$$

(d) Update the strong learner

$$f_t(x) = f_{t-1}(x) + \sum_{j=1}^{J} c_{tj} I(x \in R_{tj}). \quad \text{(Equation 2.19)}$$

Here, the indicator function $I(\cdot)$ has the value 1 if its argument is true, and zero otherwise.

(3) Obtain the expression of the strong learner $f(x)$:

$$f(x) = f_0(x) + \sum_{t=1}^{T} \sum_{j=1}^{J} c_{tj} I(x \in R_{tj}). \quad \text{(Equation 2.20)}$$

The first step of the algorithm is to estimate the constant value that minimizes the loss function, which is a tree with only one root node. Step 2(a) calculates the value of the negative gradient of the loss function in the current model, and uses it as an estimate of the residual. Step 2(b) estimates the leaf node area of the regression tree to fit the approximate value of the residual. Step 2(c) uses linear search to estimate the value of the leaf node area to minimize the loss function. Step 2(d) updates the regression tree. In step 3, the final model $f(x)$ of the output is obtained.

2.2.1.8 Introduction to XGBoost Algorithm

XGBoost[77] is a gradient boosting algorithm, which has attracted people's attention recently. Compared with GBDT, XGBoost has significantly improved the speed and accuracy. The full name of XGBoost is extreme gradient boosting, which is a gradient boosting machine implemented in C++. The significant feature of XGBoost is that it can automatically use the multithreading parallelism of CPU, which improves the algorithm and accuracy[77].

Compared with traditional GBDT, XGBoost also supports the linear classifier. At this time, XGBoost is equivalent to logistic regression (classification) or linear regression (regression) with L1 and L2 regularization. Traditional GBDT only uses the first derivative information for optimization, while XGBoost performs the second-order Taylor expansion on the cost function, and adds the regularization to get the optimal solution of the global objective function, and gives the proposal of balancing the objective function decline and the model complexity[75].

In order to prevent overfitting, shrinkage and column subsampling are added when the shrinkage is similar to the learning rate. After each step of tree promotion, the shrinkage is increased by a parameter n (weight), thus reducing the influence of each tree and providing space for optimizing the model for the following trees. Column sampling (feature) is said to be learned from random forest side. Compared with traditional line sampling (including line sampling function), it has better effect of preventing over-fitting, which is beneficial to the parallel processing algorithm mentioned later.

2.2.1.9 Introduction to LightGBM Algorithm

LightGBM[78] is a new boosting framework introduced by MSRA. As the name implies, LightGBM contains two key points: Light refers to lightweight and GBM refers to gradient. Its advantages: high training speed, low memory occupancy, high accuracy, support for parallel learning and be capable for process large-scale data.

There are two main reasons why LightGBM can achieve such good performance improvement: first, the decision tree algorithm based on histogram is used to replace the traditional pre-ranking algorithm; second, the leaf growth strategy based on depth restriction is used to replace the traditional hierarchical decision tree growth strategy, which improves the accuracy and avoids the danger of overfitting. The leaf growth may produce a deeper decision tree, which leads to overfitting. Therefore, LightGBM adds a maximum depth limit on the leaf to prevent overfitting, while ensuring high efficiency.

Machine learning theory has made long-term development in recent decades, and the related theoretical framework has been constantly improved by scholars. Nowadays, machine learning algorithm have been applied in many fields, especially industrial manufacturing. Machine learning algorithm has been integrated into the production line by a large number of factories because of its fast response, high accuracy and low construction cost.

2.2.2 Application of Machine Learning in Iron and Steel Field

Machine learning in AI mainly studies how to make machines acquire new knowledge or skills through self-learning by writing corresponding computer programs to solve problems to be solved. Machine learning is one of the core research directions in the field of AI, which can provide many technical supports for the intelligent development of iron and steel field. With the continuous development of machine learning algorithm, its application in iron and steel field is increasing gradually.

There is a certain defective rate in the production of iron and steel enterprises, which will directly affect the production efficiency and cause serious waste of resources. Therefore, the detection and classification of steel defects mainly rely on experienced workers. The disadvantages of this method are that it requires a large number of workers, takes a long time, and has a high probability of misjudgment. The application of machine learning can largely avoid the above problems. Cluster analysis, decision tree, random forest and support vector machine can all be used to solve the classification problems. Cluster analysis is a process of grouping physical or abstract collections into multiple classes composed of similar objects. Decision tree is a decision analysis method to evaluate the project risk and judge its feasibility on the basis of knowing the occurrence probability of various situations, which can better solve the multi-classification problem.

In addition to the problem of product defect detection, there are also the problems of steel shelf configuration and ex-warehousing/warehousing. The stacking problem of large steel coils mentioned in the literature is actually a multi-objective comprehensive optimization problem. By adopting partition classification storage strategy and cluster analysis method to design shelves, and decision tree method to allocate goods positions, the overall objective of high storage space utilization and high ex-warehousing/warehousing efficiency is achieved, which meets the requirements for safe and high-quality industrial production, and the production has the characteristics of accuracy, rapidity and reliability. Therefore, the decision tree for classification can also solve the optimization problem of steel storage.

It is also very important to identify and diagnose defects in the production process. In the literature, on-line diagnosis system of continuous casting plate blank quality is developed, and production abnormal event model and neural network diagnosis model are established, which can identify production abnormal events, make preliminarily diagnose for defect types, and be combined with process data to determine the key factors causing abnormal events.

In addition, various machine learning methods have played a great role in solving the problems of plate shape prediction, defect detection, production process control optimization, procurement and inventory management, dispatching optimization, production optimization and so on.

With the rapid development of industrial artificial intelligence, the manufacturing system of modern manufacturing industry is becoming more complex and dynamic. It is of great significance to make full use of all possible means to meet the needs of users for high-quality products more effectively. At present, the rapidly developing machine learning field is an interdisciplinary scientific field, which ensures the accuracy and availability of evaluation and results[50]. Machine learning is thought that it can solve many new and old challenges in manufacturing industry and give answers, which aroused extensive discussion among researchers and practitioners. However, the definition of machine learning is broad, and the successful theoretical knowledge and algorithm implementation is still in a rapid development process. New algorithms and models are constantly developing and changing dynamically, which brings great challenges and obstacles to application of machine learning.

Machine learning allows machines to learn by themselves according to design rules. It allows computer programs to automatically improve the performance of models in certain tasks through learning[79]. Manufacturing industry is an important application field of machine learning. Although not widely used in manufacturing industry, the machine learning technology plays an important role.

Many applications of machine learning are knowledge-based systems for production process automation, such as expert systems for decision support and intelligent dispatching systems for parallel production, which form such a system by collecting expert knowledge. However, the process of collecting knowledge is still a problem. Machine learning technology can automatically acquire expert knowledge when building an expert system. In this process, the development speed is improved, the development cost is lowered, and the time needed to learn from experts and knowledge engineers is reduced. Automation may obtain knowledge that is easily ignored in the process of knowledge acquisition.

The field of machine learning is closely related to pattern recognition and statistical inference. A lot of researches in the field of machine learning focus on classification, and the training model can correctly identify the new samples of the same population from a series of known samples. These classification models are widely used in manufacturing, telecommunications, marketing and scientific analysis[80]. Many problems in manufacturing belong to the classification category, but before establishing the classification model, industry experts are required to specify a category label for the objects or scenes as the target value of model training according to the specific values of a set of parameters.

Machine learning models and algorithms include inductive learning algorithms (such as decision tree induction and rule induction), case-based learning, neural networks, genetic algorithms and Bayesian learning algorithms. Among various machine learning methods for classification, example induction is the most commonly used method in practical application[81]. Inductive learning technology is faster than other technologies, with another advantage of being simple and the generated model being easy to understand. Finally, compared with other classification techniques, inductive learning classifier can achieve similar or even better classification accuracy.

In modern times, the manufacturing industry is experiencing unprecedented high-speed data growth, including data from production line sensors, environmental data, machine parameters and so on. Moreover, with uneven quality, these data have various formats and semantics. Therefore, in order to cope with the industrial development under the new situation, experts and scholars have put forward different concepts, such as Industry 4.0 (Germany), Intelligent Manufacturing (USA), Made in China 2025 Plan (China) and Intelligent Factory (Korea). Data is collected in large amounts, which is usually called big data. The availability of quality-related data provides the possibility for continuous improvement of process and product quality[74]. However, it is recognized that too much information brings challenges and negative effects, because it may lead to delays in inferring correct actions, and even lead to wrong conclusions. Generally speaking, in order to benefit from the existing data, such as quality improvement plan, manufacturing cost estimation or process optimization,customer demand enhancement, etc., the manufacturing industry has realized and acknowledged that there is a strong need to introduce machine learning in high dimension and complexity, as well as the dynamics for data processing.

2.2.2.1 Application of Machine Learning in Dynamic Dispatching

With the continuous popularization of flexible manufacturing systems (FMS), the machine learning plays an increasingly important role in the manufacturing industry. System dispatching rules are usually used to dynamically schedule production. However, one disadvantage of using dispatching rules is that their performance depends on the state of the system[82]. Unfortunately, there is currently no uniform rule applicable to all possible countries where the system may be applicable, but without better rules. If the best rule can be used for each specific situation, this defect will be further eliminated. Therefore, a dispatching method based on machine learning is proposed. The "scheduling knowledge" is obtained by using this advanced technology and analyzing the early performance of the system[83]. So, the correct dispatching rule can be determined at each specific time. Three different machine learning algorithms can be used to obtain "dispatching knowledge": inductive learning, neural network based on back propagation and case-based reasoning (CBR).

Machine learning-based planning technology is widely used in manufacturing systems for all tasks from product design to production planning and control. One of the tasks is to automatically generate the control sequence of the whole production system. Therefore, the final plan can be directly used as a continuous control program to guide the operation of the production system[84]. As a sequence control program of manufacturing system, Hybis is a hierarchical nonlinear planner, aiming to obtain a detailed partial orderly plan. At present, sequence control program is manually generated by modeling tools. Using machine learning technology will help to improve the efficiency of Hybis problem[85]. It realizes the deductive learning method of automatically generating control knowledge, and solves the drawing problem by generating bounded interpretation problems. This learning method is based on Hamlet, which is a system for learning and controlling knowledge in the form of control rules.

Facing the increasing pressure of just-in-time (JIT) production, many manufacturing enterprises are developing dispatching systems. In the process of research and development, the enterprises should consider both the traditional penalty of being late and the penalty of being early. Therefore, the implementation of JIT dispatching is very complicated. However, many JIT dispatching systems are currently based on traditional dispatching methods, such as mathematical planning, branching and constraint, heuristic algorithm, etc. Therefore, they can't deal with the uncertainty that prevails in the current manufacturing system. For example, in case of urgent work, machine failures usually occur in the workshop. Static scheduling method based on traditional optimization model often fails because it cannot adapt to real-time workshop environment. To solve this problem, Literature[86] has proposed a kind of real-time dynamic adaptive dispatching control system based on reinforcement learning. These systems are composed of adaptive agents that can learn from actual or simulated experience and adapt to dynamic environment. However, these systems only aims to meet the demand rate, or to minimize conventional measures, such as completion time that actually refers to the total time required to process a given number of parts. In this respect, these measures are not applicable to the order production environment with unknown reasons. Every part of the order production environment is related to the due date, and it is also very important to produce on time. Therefore, the requirement for development is not only to adapt to the dynamic changes of the workshop environment, but also to adapt to the adaptive dispatching methods of JIT-oriented production environment.

Centralized learning and distributed workshop control algorithm can help solve the above problems. In this algorithm, components and machines control their behavior by minimizing inventory (by minimizing lead time) and delay, so as to minimize the mean squared difference (MSD) cost of accurate delivery date and the installation cost, whiling ensuring JIT production. It is an important goal to minimize the installation cost by minimizing the installation time, because the machine time of parts processing can be used to improve the productivity of the production system.

The efficient cooperation among the sub-processes of iron and steel production is of great significance, which directly affects the production cycle and energy consumption of products. Min Kong et al.[87] introduced an improved variable neighborhood search algorithm to study the coordinated rolling operation and batch distribution scheduling problem with the sum of the total departure time and the distribution cost performance as the index. At the same time, the study considered continuous casting and The interval processing time in the rolling process can effectively solve the problems encountered in actual coordination by verifying the algorithm. Aiming at the planning and scheduling problems of steelmaking and continuous casting operations, the literature [88] applied the simulated annealing method to solve the problem. This method is due to the greedy algorithm in almost all cases and performs well in the examples. Kun Peng et al.[89] proposed an improved artificial bee colony algorithm to solve the NP-difficult Steelmaking-Refining-Continuous Casting (SCC) scheduling problem. The results show that and can be used as a good method to solve the SCC scheduling problem. MHF Zarandi et al.[90] combined particle swarm optimization algorithm and fuzzy linear programming to solve the SCC scheduling problem. By solving different problems and comparing with the algorithms in the literature, the results show that the algorithm can better solve the SCC problem. Literature [91] established a multi-objective optimization mathematical model with the goal of advancing or delaying the minimum start-up time, the shortest heating waiting time of each process, and the shortest converter idle time to solve the scheduling problem in "furnace-cast matching" production. This model Based on genetic algorithm, greedy strategy and linear programming, a simulation experiment was carried out with actual production as an example. The result showed that it matched the actual process well, and it has been applied to a large steel plant in China.

The dispatching management of machine learning in flexible manufacturing has brought new methods to users, and solved the problems that users cannot solve before.

2.2.2.2 Application of Machine Learning in Quality Control

The growing market demand for product quality and process efficiency is prompting enterprises to consider new and innovative ways to optimize their products. In the field of high-tech industrial products, small changes in production status during the production process can lead to increased costs, increased workload and even scrap. At the same time, the quality of high-tech engineering products is increasingly improved. In the face of these changing trends, manufacturing enterprises must address the rapidly increasing complexity in their manufacturing and business processes to achieve desired product quality.

Describing the status of individual products throughout the production process, including all relevant usage information, such as the adjustment of process parameters, is a way to meet quality requirements and remain competitive[92]. Ideally, the information collected can be analyzed directly. Once a serious trend or event is identified, the corresponding action, such as an alert, can be triggered. Due to the complexity and high dimensionality of modern manufacturing systems, traditional approaches based on causality models have encountered potential bottlenecks. These traditional methods are not very efficient or effective. We need new approaches to deal with this complexity and high dimensionality. These methods shall also be able to produce useful results at a reasonable cost. On this basis, Wuest T, Irgens C, Thoben KD et al. proposed a clustering analysis and supervised machine learning method to generate a prediction system based on product status data in production plan.

In recent years, the manufacturing industry, especially the aerospace and automotive industries, has become increasingly demanding in terms of quality control. No defective material is allowed in the manufacturing process, as a small defect in a component may cause disaster in later use. In the production of steel plates, the rolling process is often the last process that affects the microstructure of the steel plates. The cost of generating defects in the rolling process is high because it takes more than 5,000 kWh of energy to produce one ton of steel. Early detection of defects can reduce product damage and manufacturing costs.

As a pattern classification problem, the study of this problem belongs to the field of defect detection modeling. There are two important issues that need to be addressed in the process of defect detection in manufacturing. First, it is necessary to explore, extract and optimize effective defect features. Defect features describe the surface defects and determine the complexity of the classification system. The second important issue is the design of the classifier, which determines the performance of the whole vision system in terms of classifying defects. Jia H, Murphey YL, Shi J, et al. developed a support vector machine-based machine learning system[93].

Improving the process quality of industrial products (QI) requires the collection and analysis of data to solve quality-related manufacturing problems. Qualified intermediary programs, such as Design for Six Sigma (DFSS) and improvement, continually encourage the collection of data to address quality issues. With the development of automation and computer systems, more and more data is available for manufacturing processes. While traditional data analysis tools have been used successfully to improve the quality of products and processes, today better tools, such as machine learning, are used to mine the large data sets collected by computer systems.

Mechanical properties are an important indicator to measure the quality of steel products. Literature[94] introduces support vector machines combined with metallurgical mechanism models to predict the hardness of ultra-high-strength stainless steels. Support vector machines use radial basis kernel functions as their kernel functions. The values of the two important parameters of the function are selected by genetic algorithm, and the mechanical properties obtained through prediction are basically consistent with the experimental results. Li et al.[95] also used the stacking integrated learning method to predict the mechanical properties of the steel plate, and the prediction results are basically consistent with the experimental results. The work [96] aimed to devise a deep learning model, to predict mechanical properties of industrial steel plate including yield strength (YS), ultimate tensile strength (UTS), elongation (EL), and impact energy (Akv). Based on Principal Component Analysis (PCA) and Gradient Lifting Decision Tree (GBDT) methods, literature [97] takes the tensile strength as the research object and constructs a prediction model for the mechanical properties of hot rolled strip steel.

Non-metallic inclusions will inevitably be produced in the process of casting steel, which will cause adverse effects such as mechanical strength reduction. Literature[98] applies machine learning algorithms to determine the number of inclusions and performance of tire reinforcement castings for classification, and compares model predictions with actual values. In comparison, the random forest model is the best performing model.

Quality problems may involve many input and output variables that are not easily modeled and optimized. As pointed out by Yang and Trewn (2004), data mining (DM) and knowledge discovery in databases (KDD) have been successfully used to address the quality and control of many variables at different stages of the product (process) life cycle[99]. However, this problem-solving approach needs to be tested in order to guide practitioners.

In the modern manufacturing environment, a large amount of data is collected in areas related to product and process design, assembly, material plan, quality control, schedule plan, maintenance, etc. Data mining has become an important tool for obtaining knowledge from industrial databases. Choudhary A K, Harding J A and Tiwari M K reviewed the literatures dealing with the application of knowledge discovery and data mining in many manufacturing areas, with particular emphasis on the types of functions applied to the data. The main data mining functions used include feature description, correlation analysis, classification, prediction, clustering, and evolutionary analysis[100]. Research has shown that data mining has been growing rapidly over the past three years in both manufacturing processes and enterprise applications.

2.2.2.3 Application of Machine Learning in Manufacturing Condition Monitoring

There is an increasing demand from manufacturing enterprises for the development of models for predicting mechanical failures, remaining useful life of manufacturing systems or parts and components. Unlike traditional manufacturing enterprises, intelligent manufacturing enterprises use information and communication technology (ICT), intelligent automation systems and sensor networks to monitor the status of machines, diagnose the root cause of failures, and predict the remaining useful life (RUL) of machine systems or components.

Traditional model-based or physics-based predictions usually require a deep physical understanding of the system in order to build a closed mathematical model. However, a priori knowledge of system behavior is not always known, especially for complex manufacturing systems and processes. To complement model-based prediction, data-driven approaches are increasingly being applied to machine prediction and maintenance management, transforming traditional manufacturing systems into intelligent manufacturing systems characterized by artificial intelligence. Although previous studies have demonstrated the effectiveness of data-driven approaches, most prediction methods are based on classical machine learning techniques such as artificial neural networks (ANNs) and support vector regression (SVR). With the rapid development of artificial intelligence technology, various machine learning algorithms have been developed rapidly and are widely used in many engineering fields.

Prediction methods fall into two categories: model-based prediction and data-driven prediction. Model-based prediction is an approach based on mathematical models of system behavior, where the mathematical models are derived from the physical laws and formulas of probability distributions. Model-based prediction methods include Wiener and Gamma processes, Hidden Markov Models (HMMs), Kalman filters, and particle filters[101]. One of the limitations of model-based prediction is that it requires insight into the underlying physical processes that causesystem failures. Another limitation is that it is based on the assumption that the underlying process follows a certain probability distribution, such as a gamma or normal distribution. Although probability density functions can quantify uncertainty, the distribution assumptions may not match the facts.

To complement the model-based prediction, the data-driven prediction approach is introduced. This method uses learning algorithms and extensive data training to build prediction models. Classical data-driven prediction models are based on autoregressive (AR) models, multiple adaptive regression, fuzzy set theory, ANNs and SVR. The unique advantage of a data-driven approach is that a deep understanding of the physical behavior of the system is no longer a prerequisite. In addition, the data-driven approach does not assume any potential probability distribution. These assumed probability distributions may not be applicable to practical applications[102].

Traditional statistical process control (SPC) control chart techniques are not applicable in many process industries where data are often autocorrelated. It occurs mainly in highly automated and integrated parts in manufacturing. Experts and scholars have made some attempts to extend the traditional SPC technique to deal with autocorrelation parameters. However, these extensions bring some serious limitations. Widodo A and Yang BS demonstrated that support vector machines (SVMs) in autocorrelation processes can effectively minimize Class I errors (the probability that the method incorrectly states that the flow is not controlled or generates false alarms) and Class II errors (the probability that the method fails to detect actual displacements or trends in the system) [103]. Even when using the simplest polynomial kernel functions, support vector machines can do a good job of detecting displacements in the paper writing process and viscosity datasets with performance comparable to or better than that of traditional machine learning. It can be seen that support vector machines are good at reducing class I and class II errors even when monitoring unrelated processes. When testing in the literature, they again performed at or better than the classical Shewhart control charts and other machine learning methods.

Due to the harsh production environment and operating factors, production equipment is vulnerable to damage, so continuous maintenance and inspection are required. The development of continuous and reliable monitoring will ensure the safe use of these structures and help extend their service life. Literature [104] developed a continuous diagnosis and equipment condition monitoring system based on machine learning algorithms for a stainless steel pipe manufacturing company, which reduced equipment maintenance costs, improved product quality, optimized equipment management conditions, and verified it in production. To improve its feasibility. Matthias Karner, etc. applied support vector machines, linear models, decision trees and other machine learning methods to predict the status of cutting equipment in iron and steel enterprises online[105].

Semiconductor manufacturing is the most technology-oriented and cost-intensive industrial sector that has a significant impact on people's daily lives. In fact, semiconductor-based devices are everywhere: personal computers, cell phones and cars are the most direct examples[106]. As a result, semiconductor manufacturing enterprises spend a lot of effort and resources to improve quality. It's no surprise that the way is open for smaller, faster and better-quality devices.

2.3 Deep learning and its application in iron and steel industry

Machine learning algorithms often require a large number of high-quality characteristic variables that are constructed artificially in advance and then the construction of machine learning models. In other words, machine learning models built by algorithm engineers with different experience have mixed results in practical applications. The reason is that special variables need to be constructed artificially. However, new techniques like deep learning tend to use end-to-end learning, and directly use models to automatically learn features in the data instead of relying on artificial construction, which greatly reduces the problem of insufficient human experience and achieves higher accuracy.

2.3.1 Development history and introduction of classification algorithm

2.3.1.1 Development history of classification algorithm

With the development of AI, the applications of computer vision, for example, face recognition and vehicle license plate recognition, has been successfully implemented. The successful implementation of computer vision technology has been made possible by the development of deep convolutional neural networks.

Before convolutional neural networks were proposed, the commonly used image classification algorithms include machine learning algorithms such as SVM[59] and KNN[65], as introduced earlier. These algorithms often require the image features to be designed manually using such methods as SIFT[107], HOG[108] and so on. Contrary to traditional manually-designed features, convolutional neural networks can learn features from large-scale data, without manual intervention in the training process, and can transfer the learned features to various recognition tasks. With the increase in data volume and accuracy requirements, the attention of research scholars has shifted to designing better network architectures for learning.

The AlexNet neural network has become famous in the ILSVRC vision competition in 2012, and its effects have substantially outperformed traditional methods. On the million-volume Imagenet dataset, the recognition rate has improved from the 70%+ (by using traditional algorithms) to 80%+, making deep learning official step onto the stage. After that, new records are set by deep learning on ILSVRC every year, and the development of convolutional neural networks became unstoppable. In addition to higher accuracy, they are commonly characterized by the fact that the layers of the network are getting deeper and deeper.

VGG neural network[109] won the ILSVRC vision competition in 2013. It is further divided into VGG16 and VGG19, with the number of layers increased to 16 and 19 respectively on the basis of AlexNet. In addition to being excellent in recognition, it also has a good recognition effect for image target detection. The GoogLeNet neural network won the ILSVRC vision competition in 2014. In addition to increasing the number of layers to 22, the main innovation of GoogLeNet lies in its Inception structure, a nets-within-nets structure, i.e., the original node is also a network. The width and depth of the entire network structure can be expanded with the use of Inception, which can bring a great performance improvement. The ResNet neural network[110], which won the ILSVRC vision competition in 2015, can directly increase the depth up to 152 layers, and its main innovation lies in the residual network, of which the proposal is to essentially solve the problem that the network cannot be trained in case of a relatively increased number of layers. This network draws on the idea of Highway Network, which is equivalent to opening a channel on the side so that the input can go straight to the output, and the optimization goal is changed from the original fitted output $H(x)$ to the difference between the fitted output and input, $F(x) = H(x) - x$, where $H(x)$ is the original expected mapping output of a layer and x is the input. The Inception-ResNet-v2 neural network, which won the ILSVRC vision competition in 2016, is a relatively new classical model that integrates depth and width, and is a combination of Inception v3 and ResNet.

Deep convolutional neural networks have made a series of breakthroughs[111, 112], in the field of images. It integrates low-, medium- and high-level features in an end-to-end multi-layer fashion, and the level of features can be enriched by the number of stacked convolutional/pooling layers (depth) or by the special structure (width) of the neural network. Recent evidence shows that the depth and width of the network are critical. The those with leading results on the challenging ImageNet dataset all use very deep or very wide network models[110, 111], and many other difficult visual recognition tasks also benefit greatly from very deep neural networks[111, 113].

2.3.1.2 Introduction to artificial neural network

Before introducing convolutional neural network, an introduction to artificial neural networks shall be made first. Artificial neural network (ANN) is a computational model inspired by the biological neural network process of the human brain for processing information. Artificial neural networks have caused a great surge in research in machine learning industry thanks to its groundbreaking results in voice recognition, computer vision and word processing.

An artificial neural network is a feed-forward neural network in which nodes between adjacent layers are connected unidirectionally from front to back, namely from input nodes, to hidden layer nodes (if any), then to output nodes. Feedforward neural networks contain two main types: single-layer perceptrons and multilayer perceptrons. Single-layer perceptrons do not contain hidden layers, and multilayer perceptrons contain one or more hidden layers. The structure of a multilayer perceptron is shown in 2.3.

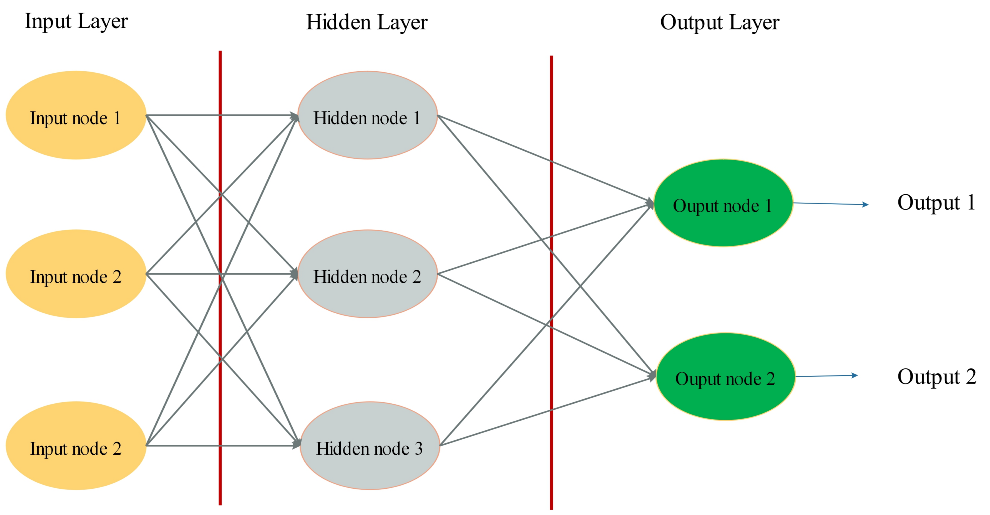


Figure 2.3: Structure of feedforward neural network.

The input layer is responsible for passing the information to the hidden layer, which performs the calculation and then passes the information to the output layer, which is responsible for calculating and outputting the result. If there is an error between the output result and the actual result, the weight value and offset of the hidden layer are adjusted by back propagation to finally reach a suitable output.

2.3.1.3 BP neural network

BP neural network[114] is a multilayer feed-forward neural network, which is mainly characterized by forward propagation of input signals and back-propagation of errors. In the forward propagation of the input signal, the input signal passes from the input layer to the hidden layer for processing and then passes to the output layer, and the neuron status of each layer only affects the neuron status of the next layer. If there is an error between the predicted value of the output layer and the true value, it will shift to the back propagation of errors, and the weights and thresholds of the neural network will be adjusted according to the prediction error, so that the predicted output of the BP neural network continuously approximates the true result. The topological structure of BP neural network is as shown in Figure 2.4.

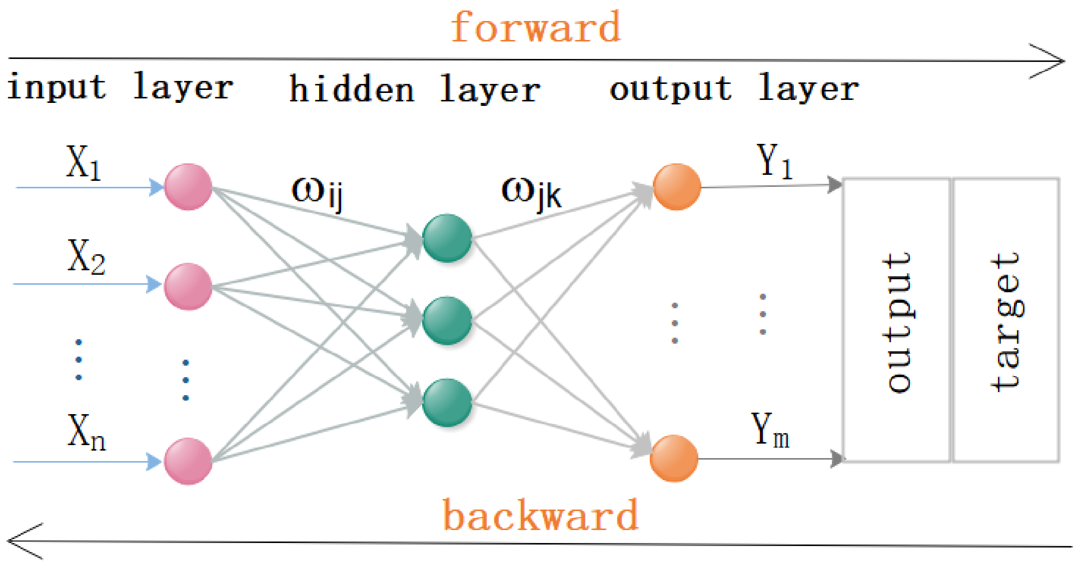


Figure 2.4: BP neural network.

In Figure 2.4, $X_1, X_2, \cdots, X_n$ is the input value of BP neural network, and $Y_1, Y_2, \cdots, Y_m$ is the predicted output value of BP neural network, and ω_{ij} and ω_{jk} are the weighted values of neural network. It can be seen from Figure 2.4 that BP neural network can be regarded as a nonlinear function, and the network input value and predicted value are independent variable and dependent variable of the function respectively. When the number of input nodes is n and the number of output nodes is m, BP neural network indicates the functional mapping relationship from n independent variables to m dependent variables.

To make the BP neural network predictive, the network must first be trained to have memory and predictive capabilities. The training process of the BP neural network includes the following steps:

Step 1: Network initialization. Determine the number of input layer nodes n, output layer nodes l and hidden layer nodes m according to the input-output sequence, and then initialize the connection weights ω_{ij} and ω_{jk} among input layer, hidden layer and output layer neurons, and then initialize hidden layer threshold a, and output layer threshold b, and give out the learning rate and activation function.

Step 2: Calculation of output of hidden layer. Calculate the output H of the hidden layer with the input variable X, the connection weights ω_{ij} between the input layer and the hidden layer and the hidden layer threshold a, where l is the number of hidden layer nodes and f is the activation function of hidden layer.

$$H(j) = f\left(\sum_{i=1}^{n} \omega_{ij} x_i - a_j\right), j = 1, 2, \cdots, l. \qquad \text{(Equation 2.21)}$$

Step 3: Calculation of output of output layer. Calculate the predicted output value O of BP neural network with the output H of hidden layer, connection weight ω_{jk} and threshold b.

$$O(k) = \sum_{j=1}^{l} H_j \omega_{jk} - b_k, k = 1, 2, \cdots, m. \qquad \text{(Equation 2.22)}$$

Step 4: Error calculation. Calculate the network error e with the network predicted output value O and the true value Y.

$$e(k) = Y_k - O_k, k = 1, 2, \cdots, m. \qquad \text{(Equation 2.23)}$$

Step 5: Weight update. Update the network connection weights ω_{ij} and ω_{jk} according to the network forecast error e.

$$\omega_{ij} = \omega_{ij} + \zeta H_j (1 - H_j) x_i \sum_{k=1}^{m} \omega_{jk} e_k, j = 1, 2, \cdots, l; j = 1, 2, \cdots, m. \qquad \text{(Equation 2.24)}$$

$$\omega_{jk} = \omega_{jk} + \zeta H_j e_k, j = 1, 2, \cdots, l; k = 1, 2, \cdots, m. \qquad \text{(Equation 2.25)}$$

Where, ζ is the learning rate of neural network.

Step 6: Threshold update. Update the network node thresholds a and b according to the network forecast error e.

$$a_j = a_j + \zeta H_j (1 - H_j) \sum_{k=1}^{m} \omega_{jk} e_k, j = 1, 2, \cdots, l. \qquad \text{(Equation 2.26)}$$

$$b(k) = b_k + e_k, k = 1, 2, \cdots, m. \qquad \text{(Equation 2.27)}$$

Step 7: Determine whether the algorithm iteration is over, if not, back to step 2.

2.3.1.4 Introduction to Convolutional Neural Network and Its Components

Convolutional neural network is one of the most classical models in deep learning. Convolutional neural network can be found in all classical models of deep learning today. By using a few weights cleverly, it attains an effect that a fully connected neural network can't achieve.

In the fully connected neural network structure, the spatial structure of the image is not considered, which leads to its inability to correctly predict the category of the object in the picture when the position of the object in the training set picture is different from that in the test set picture. But the convolutional neural network makes use of this spatial structure and it uses local areas to perceive pictures, thus reducing the number of parameters, and making the training speed very fast. And it can also be used to train deep networks. The convolutional neural network is based on three basic ideas: local perception domain, weight sharing and pooling.

The convolutional neural network mainly includes input layer, convolutional layer, pooling layer, fully connected layer and output layer. Among them, the convolutional layer is the hidden layer for convolution operation, and the pooling layer is the hidden layer for pooling operation. The main structure of convolutional neural network is shown in Figure 2.5:

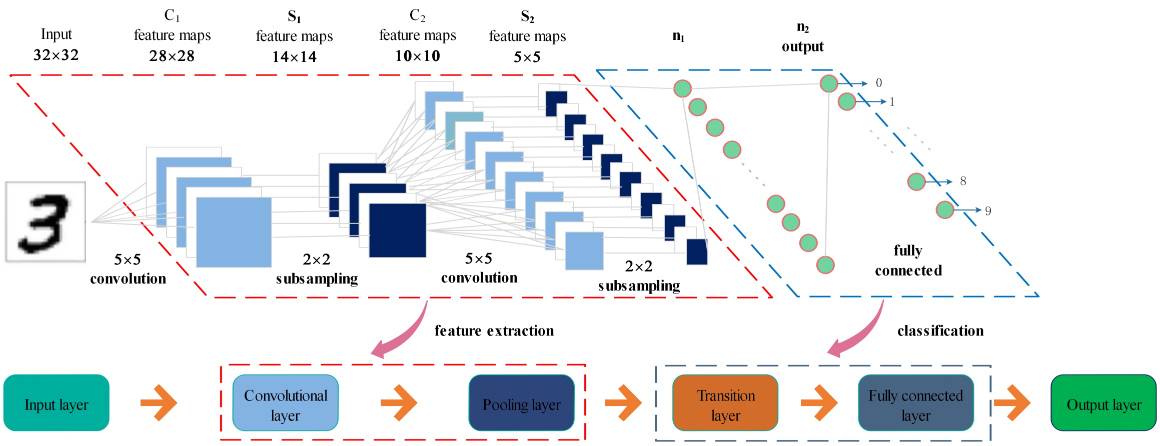


Figure 2.5: Structure of convolutional neural network.

From the above figure, we can see the data flow in the convolutional neural network (from input to output), specifically, during this process, a picture will be used as a network input, it enters the convolutional layer through local connections, and then its features will be extracted in the convolutional layer, and then the dimensionality of the convolutional layer is reduced through the pooling layer. The dimensionality reduction process will cause the loss of the feature information of the picture, and then through the convolutional layer, the pooling layer, etc., the picture finally enters the fully connected layer and the softmax layer (output layer), there, the target objects in the picture will be classified.

2.3.1.5 Introduction to VGG Neural Network

In 2014, the Visual Geometry Group of the University of Oxford and the researchers from the Google DeepMind Company developed a new deep convolutional neural network: VGGNet [109], which won the first place and the second place on ILSVRC localization and classification respectively. It is also a ground-breaking CNN architecture, while the most striking part of it was its depth. At that time, it was definitely a very, very deep network architecture. This paper mainly focuses on the influence of the depth of the convolutional neural network on the recognition accuracy of large-scale image sets. The main contribution of it is that it proposes the use of a very small convolution kernel (3×3) to construct a convolutional neural network structure of various depths. And these network structures were evaluated, and finally proved that a better recognition accuracy may attain with a network depth of $16 - 19$ layers. They are VGG-16 and VGG-19 that are commonly used to extract image features.

It can be seen from Table 2.1 that three 3×3 convolution kernels are used in the VGG neural network instead of 7×7 convolution kernels, and two 3×3 convolution kernels are used instead of 5×5 convolution kernels. The main reason for doing this is to improve the depth of the network and improve the effect of the neural network to a certain extent while ensuring the same receptive field.

2.3.2 Development history and introduction of object detection algorithm

2.3.2.1 Brief introduction to object detection algorithm

Object detection is a computer vision task that distinguishes the object in an image or video from other uninteresting regions, judges whether there is an object, determines

Table 2.1: Schematic diagram of VGG neural network structure [109].

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC1000					
softmax					

the position of the object, and identifies the type of the object. Object Detection is one of the basic tasks in the field of computer vision, which has been studied in academia for nearly twenty years. Most of the early research on object recognition are based on template matching technique[115], focusing on specific objects (such as human faces) with a roughly fixed spatial layout. Before 1990, the mainstream of object recognition was based on geometric representation[116, 117]. Later, its emphasis was shifted from geometric and previous model matching to using statistical classifiers based on appearance features, such as neural network[118], SVM[119] and Adaboost[120, 121]. These successful object detectors laid a foundation for most subsequent researches in this field[122].

At the end of 20th century and the beginning of 21st century, great progress has been made in object detection research. Appearance features change from global representation[121, 123] to local representation with invariance of translation, scaling, rotation and occlusion. For hand-made local invariant features, great popularity has been gained from the start of Scale Invariant Feature Transform (SIFT) features[124], so the progress of various visual recognition tasks is basically based on the use of local descriptors[125], such as Haar-like features[121], SIFT[124], Shape Contexts[126], histograms of oriented gradients (HOG)[108], local binary patterns (LBP)[127] and covariance[128]. These local features are usually integrated with simple cascades or feature pool encoders, such as the Visual Word Bag method[129] introduced by SivicandZisserman[130] and Csurka, spatial pyramid matching (SPM)[131] based on BoW model and Fisher vectors[122].

In 2012, an important turning point occurred in the field of computer vision when the deep convolutional neural networks set new records in the image classification tasks. The successful application [112] of deep convolutional neural networks in image classification was transferred to object detection, which resulted in the R-CNN detector proposed by Girshick et al.[132]. Since then, the field of object detection has made great progress, and many object detection methods based on deep learning have been developed: from OverFeat[133] proposed in 2013, to Fast R-CNN[134], Faster R-CNN[135], SSD[136], YOLO[137] series, and then to Pelee recently proposed in 2018. In less than five years, the object detection techniques based on deep learning have been developed in terms of network structure: from two stage to one stage; from bottom-up only to Top-

Down; from single scale network to feature pyramid network; and from PC-oriented to mobile phone-oriented. Many good algorithm techniques have emerged. These algorithms have excellent detection effect and performance on public object detection data sets.

There are many uncertain factors in the process of object detection, such as the uncertain number of objects in the image, the different appearances, shapes and postures of objects, and the interference of illumination, occlusion and other factors in case of object imaging, resulting in a certain degree of difficulty in the detection algorithm. Since entering the era of deep learning, the development of object detection mainly focuses on two directions: one-stage algorithm and two-stage algorithm. The main difference between the two is that the two-stage algorithm needs to create a proposal, and then perform fine-grained object detection (for example, R-CNN series), while one-stage algorithm can directly extract features from the network to predict the classification and location of objects, without the need to search for candidate regions separately, typically including SSD/YOLO.

2.3.2.2 Traditional object detection algorithms

Most early object detection algorithms are constructed based on manual features. People had to design complex features for representation because of the lack of effective image representation at that time. Three classical traditional target detection methods will be introduced below:

Viola Jones Detectors[121, 131], launched in 2001, is mainly used for face detection. Viola Jones Detectors perform the detection with the method of sliding windows, that is, going through all positions and scaling the image to see if each window contains a face. The core of Viola Jones Detectors is the combination of Haar feature, Adaboost classifier and the idea of Cascade. Haar feature is the difference between white pixels and black pixels, which is a texture feature[121]. The idea of cascade makes it possible that sliding windows without objects can be discarded quickly, thus greatly improving the detection efficiency. But it has a big disadvantage that it only uses very weak features. As for this method, Adaboost classifier which is a weak classifier is used, and its accuracy depends on implementing one-vote veto voting by multiple weak classifiers to improve the hit rate. The number of classifiers is also determined by experience or with the trial-and-error method.

The histograms of oriented gradients (HOG) feature descriptor was first proposed by N. Dalal and B. Triggs in 2005[108]. In order to balance feature invariance and nonlinearity, the HOG descriptors are designed to be calculated on evenly spaced unit dense grids, and the overlapping local contrast normalization is used to improve the accuracy. Although HOG can be used to detect object categories, it is mainly used for pedestrian detection. To detect objects of different sizes, the HOG detector will scale the input image several times while keeping the size of the detection window unchanged. For many years, the HOG detector has been an important foundation for many object detectors[138, 139, 140] and computer vision applications.

DPM is the acme of traditional object detection methods. This algorithm has been playing a role in object detection for a long time before the deep learning method is matured. It was first proposed by P. Felzenszwalb[140] in 2008 as an extension of HOG detector, and then subjected to improvements made by R. Girshick[139, 141]. DPM follows the detection principle of divide and conquer where training can be simply regarded as the correct method of learning to decompose an object, while reasoning can be regarded as the collection of detecting different parts of the object. A DPM detector consists of a root-filter and many part-filters. A weakly supervised learning method is proposed in DPM, which can automatically learn all configurations of part-filters as potential variables instead of manually specifying all configurations of part-filters. Later, R. Girshick further improved and expanded [142] the model to deal with objects in the real world under more meaningful changes.

Although the current object detectors have far exceeded DPM in detection accuracy, many methods, such as hard negative mining and bounding box regression, are still deeply affected by the value of DPM.

2.3.2.3 Object detection algorithm based on deep learning

The neural network is a basic network, which can be used for image classification. In order to provide the location information of multiple objects, Girshick et al. [132] proposed a region-based convolutional neural network (R-CNN) which takes proposals on images and objects from selective search as inputs [143], uses CNN to extract features, and uses a new simple regression and support vector machine (SVM) to locate and classify targeted objects. However, as R-CNN is not end-to-end training, it is slow to calculate and difficult to use, and three different training processes are needed for CNN, regression and SVM. In addition, the features extracted by CNN will be fed to SVM and regression in the forward propagation process of R-CNN, but the calculation will not be shared, which will lead to slow process and high cost. To deal with the shortcomings of R-CNN, He et al. [113] introduced a new CNN architecture called Spatial Pyramid Pooling Networks (SPP-Net). Compared with R-CNN, the speed of object detection is improved, but SPP-Net has a difficult training process similar to R-CNN, and its pre-trained CNN is not involved in the training process, which limits its accuracy. In 2015, Girshick proposed Fast R-CNN [134] to deal with the shortcomings of R-CNN and SPP-Net. Fast R-CNN trained end-to-end in one stage, which showed higher computing speed and accuracy than R-CNN and SPP-Net. Even though Fast R-CNN has better performance, it still relies on pre-computed object proposals, which is the same as R-CNN and SPP-Net. In addition, it is time-consuming to generate object proposals by external methods such as selective search, and it is not optimized in the training process, resulting in a slow overall computation speed and limited accuracy of Fast R-CNN. Ren et al. [110] proposed a region proposal network (RPN) to generate suggested objects by combining the extraction of object proposals with Fast R-CNN, and detect objects by combining RPN with Fast R-CNN. This method is called Faster R-CNN. Faster R-CNN reduces the computational cost by sharing features between RPN and Fast R-CNN, and improves the accuracy by training the network of object detectors and suggestion generators end to end.

In the field of object detection, besides the two-stage R-CNN series, some one-stage algorithms are also developing. Joseph Redmon et al. put forward a new object detection algorithm YOLO [137], which regards the object detection problem as a regression problem with bounding box separated from category probability prediction, and only uses a single network structure, thus reducing the number of networks selected by the suggestion box in the two-stage algorithm. Liu Wei et al. put forward SSD object detection algorithm [136], which completely eliminates the stage of suggestion generation and subsequent pixel or feature resampling, and encapsulates all calculations in a single network; of which the core is to carry out convolution calculation by using smaller convolution kernel so as to predict the category score and box offset of a set of fixed default bounding boxes. In order to achieve higher detection accuracy, predictions of different scales are generated from feature maps of different scales. It is faster and more accurate than YOLO algorithm. Later, Joseph et al. improved YOLO algorithm and proposed YOLO9000 algorithm [144] in 2017, which is able to better balance accuracy and speed, and real-time detection process can be achieved, while the detection accuracy is lower than SSD and Faster R-CNN with Resnet as the skeleton network. In 2018, Joseph et al. improved YOLO9000 algorithm and proposed YOLOv3 algorithm [145], which has a deeper network and higher accuracy, while its speed is three times that of SSD. YOLOv3 algorithm has high-speed and accurate performance, which makes it popular in industry.

According to the history introduction for the field of object detection, after entering the era of deep learning, the development of object detection mainly focuses on two directions, i.e. one-stage algorithm and two-stage algorithm, typically represented by SSD, YOLO series and R-CNN series respectively. Two-stage algorithm has high accuracy, but its running speed is slow, so it is difficult to realize real-time detection. One-stage algorithm runs fast, but its accuracy is slightly inferior to that of two-stage algorithm. Since the object studied in this thesis has high requirements for real-time performance, one-stage algorithm is applied.

YOLOv3 algorithm has a good balance between speed and accuracy, and it is widely used in industry, so it has become the first choice of object detection algorithm in this paper. In order to better understand YOLOv3 algorithm and its improvement process, this chapter will briefly introduce YOLO series algorithms respectively.

2.3.2.4 Introduction of YOLOv3 Neural Network

YOLOv3[145] was invented in 2018 and has become a very classical one-stage algorithm for object detection, including darknet-53 network structure, anchor box, FPN and other excellent structures. In the whole YOLOv3 structure, there is no pooling layer and fully connected layer. In the process of forward propagation, the size transformation of tensor is realized by changing the step size of convolution kernel, for example $stride = (2, 2)$, which is equivalent to reducing the side length of image by half, that is, reducing the area to one quarter of the original one. In order to achieve better classification results, the author designed and trained darknet-53 neural network. The experiments on ImageNet by the author show that darknet-53 has strong performance. Compared with ResNet-152 and ResNet-101, darknet-53 has not only better classification accuracy, but also faster computing speed, with fewer network layers. The diagram of YOLOv3 neural network structure is shown in Figure 2.6:

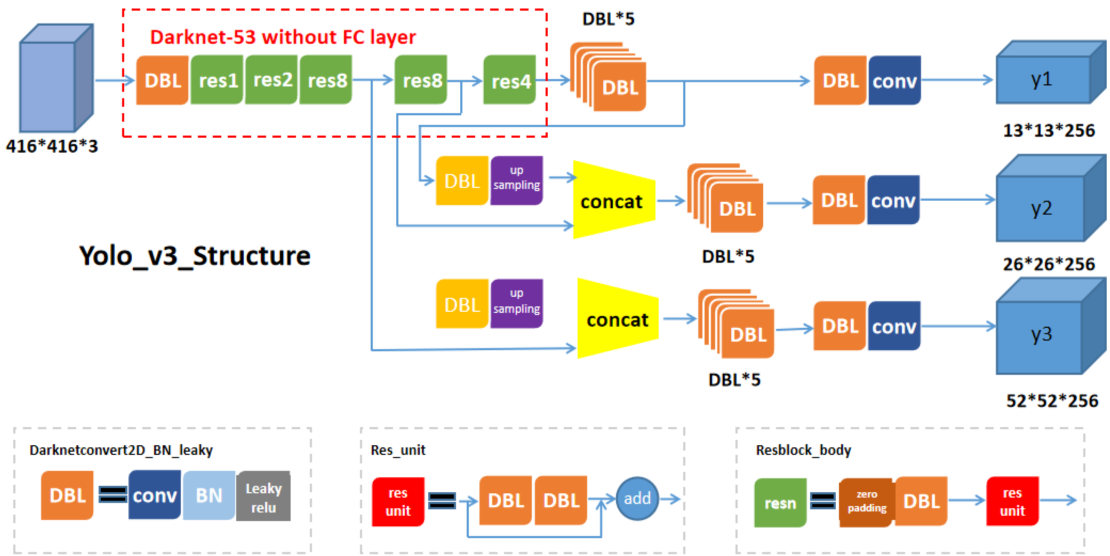


Figure 2.6: Schematic diagram of YOLOV3 neural network structure[145].

The three boxes at the lower part of the figure represent the three basic components of YOLOv3:

(1) DBL: the smallest component of YOLOv3 neural network structure, which is composed of convolution, Batch Normalization, and Leaky relu activation functions.

(2) Res unit: which, learning from the residual structure in Resnet network, can make the network building deeper.

(3) Resn: n stands for number, including Res1, Res2..., Res8, etc., indicating how many Res units there are in the Res block. Composed of a DBL and n residual components, it is a big component in YOLOv3 neural network structure. YOLOv3 began to learn from ResNet residual structure, which can make the network structure deeper. DBL in front of each Res unit module plays a role in downsampling. In addition

to the basic components mentioned above, there are tensor concatenate and add basic operations.

In YOLOv3 neural network structure, backbone adopts the first 52 layers of darknet-53 neural network, and removes the fully connected layer. On the one hand, YOLOv3 basically adopts full convolution, on the other hand, it introduces residual structure, and in order to reduce the gradient negative effect brought by pooling, the author directly abandons the pooling layer and uses stride in convolution to realize down sampling. In this network structure, a convolution with a step size of 2 is used for down sampling. Advantages of using residual structure: (1) A key point of depth model is whether it can converge normally. Residual structure can ensure that the network structure can still converge at a very deep layer and the model can be trained continuously. (2) The deeper the network is, the better the features are expressed, and the effects of classification and detection will be improved. (3) In 1×1 convolution of residuals, the idea of networkin network is used to greatly reduce the channel for each convolution; on the one hand, it reduces the number of parameters (the larger the number of parameters, the larger the model to be saved), on the other hand, it reduces the amount of computation to a certain extent.

In order to enhance the accuracy of the algorithm for small target detection, YOLOv3 uses a method similar to FPN's upsampling and fusion to fuse three scales, i.e., $y1$,$y2$,$y3$ in the figure, so that detection can be done on multiple scales of the feature map, respectively, at 32-fold downsampling, 16-fold downsampling, and 8-fold downsampling. In the three detections, each detection corresponds to a different field of perception, 32-fold downsampling has the largest field of perception, which is suitable for detecting large targets, 16-fold downsampling is suitable for detecting objects of average size, and 8-fold has the smallest field of perception, which is suitable for detecting small targets.

Each cell in the feature map will predict three bounding boxes, and each bounding box will predict three things: (1) the position of each box (four values, which are the abscissa and ordinate of the center, the height and width of the box); (2) whether it is the confidence of the object; (3) category confidence. Logistic regression is applied to predict the bounding box in YOLOv3 neural network.

In the category prediction, the original single-label classification is improved to multi-label classification, so the network structure is replaced by a softmax layer for single-label multi-classification and a logistic regression layer for multi-label multi-classification. The logistic regression layer mainly uses the sigmoid function, which can constrain the input in the range of 0 to 1. Therefore, when the output of a certain category after feature extraction is constrained by the sigmoid function, if it is greater than 0.5, it means it belongs to that category.

2.3.3 Application of Deep Learning in Field of Iron and Steel

Deep learning refers to the processing and analysis of various forms of information representing things or phenomena to describe, identify, classify and explain them. Deep learning is an important part of artificial intelligence, which is mainly applied to image classification, object recognition, image segmentation, and speech recognition, etc. Early deep learning is limited by the lack of data, the lack of hardware computing power resources, and the inability to continuously update and obtain high-quality model weights after the layers are too deep. With the development of big data in recent years, more and more data are stored. With the development of computer hardware, computing power resources are no longer the bottleneck of deep learning models. The invention of self-adaptive adjustment of the weight of the post-feedback gradient descent neural network makes it possible to build hundreds of layers of deep learning models on a s-ingle PC. Therefore, the deep learning has been mature for application. Naturally, deep learning, a new technology, is urgently needed in industrial iron and steel production to solve many problems left over from history.

The most widely used artificial neural network in the steel manufacturing industry is the post-feedback (BP) neural network. In the literature[146], the BP neural network is used to establish an inverse model reflecting the relationship between the performance indicators of hot-rolled products, the chemical composition of steel and the rolling process parameters, that is, the performance indicators of hot-rolled products are used as input, and the chemical composition of steel is related to rolling. The best value of the process parameter is taken as the output, and the relationship between input and output is established through BP neural network. In the 3Cr23Ni8Mn3N heat-resistant steel flow stress prediction under thermal deformation, BP neural network was used to study[147]. The improved Johnson-Cook (J-C) phenomenological model, the improved Zerilli-Armstrong (Z-A) as a physically-based model and the BP neural network are combined to predict the high temperature fluidity of the 5CT-L80 steel grade[148]. BP neural network is used to study the corrosion behavior of carbon steel Q345R in different concentrations of diamine mixtures, and it is better than the support vector machine model with radial basis function as the kernel function [149]. The high-temperature deformation behavior of hypereutectoid rigid in the range of strain rate $(0.001 - 1)s^{-1}$ and deformation temperature $(950 - 1100)$°C is predicted by using a three-layer BP neural network, which shows excellent results[150]. Li et al. used BP neural network to predict the temperature of molten steel in ladle furnace refining, and the performance was good[151].

In Xunqian Xu's study[152], to reduce the stress of the top pavement layer of the longitudinal stiffeners, an optimization method that combines orthogonal experimental design, BP neural network, and genetic algorithm (GA) is presented. This experiment has shown that the presented approach improved fatigue reliability and established the efficacy of the design strategy and optimization method. The paper[153] uses a combination of artificial neural network (ANN) and genetic algorithm (GA) to predict the bond strength in concrete-encased steel structures. BP neural network is also used for surface defect detection task. Yue, Xiaofeng et al.[154] proposed that the LBP feature extraction algorithm and WG-BA-BP neural network were combined to detect 6 defects on the surface of rigid surfaces, including inclusion, patches, crazing, pitted, rolled-in, and scratches.

The role of deep learning based on computer vision in the field of steel is even more important. During the production process of steel products, defects may occur on the surface, making them inferior steel or even scrap steel that cannot be exported, which seriously affects the efficiency of the company. However, the efficiency of artificial visual inspection is very low. Therefore, machines are used instead of labor to detect defective images. And recognition is extremely important. When the defect feature is a picture, image processing can be used to identify and detect the defect[106].

Xu Ke et al. proposed a new adaptive multi-scale geometric analysis method to detect and identify defects on the surface of continuous cast slabs, hot-rolled steel sheets and cold-rolled steel strips[155]. Uizhong Fu et al. proposed an end-to-end surface defect detection neural network based on pre-trained SqueezeNet to detect the defect detection data set compiled by Northeastern University[156]. Literature [157] uses a neural network with GoogleNet as the back-bone to detect and identify strip steel surface defects, and the model fully meets the actual production requirements. At the same time, Jiawei Zhang et al. proposed a fuzzy-measured surface defect detection method to detect defects on the surface of strip steel[158]. Jie Liu et al. proposed a Dual Prototype Auto-Encoder (DPAE) semi-supervised detection method to detect defects on the steel surface and achieved good performance[159]. Literature[160] proposed a new type of defect detection system based on deep learning, which has the potential of real-time detection. Hongwen Dong et al., based on the image pyramid feature fusion and global context attention mechanism, have studied the challenges of similarity between classes of surface defects and differences within classes, and the results show that this method is very effective in practical applications [161]. The semi-supervised defect detection method that combines the generative adversarial network and the residual network is used to solve the challenge of steel plate surface defects [162].

The literature [163] proposed a CP-YOLOv3-dense (classification priority YOLOv3 DenseNet) neural network to quickly and effectively detect six surface defects of steel strip. In reference [164], the newly proposed RepVGG algorithm is used to detect 7 typical defect types in the Xsteel Surface Defect Dataset (X-SDD) of hot rolled strip steel, and the results show that the accuracy of the algorithm is much higher than other similar algorithms. Praveen Damacharla et al.[165] proposed an approach that uses U-NET (TLU-net) framework based on transfer learning to detect steel surface defects. The results show that the transfer learning method is 5%(absolute) better than the random initialization method in defect classification.

It can be seen that the application of deep learning in the field of iron and steel not only reduces the unnecessary consumption of human resources, but also improves the detection accuracy, thus improving the enterprise benefits.

2.4 Summary

Expert system, machine learning and deep learning play a vital role in the development of artificial intelligence. Of course, these technologies have also been widely used in the iron and steel industry. This chapter introduces the application of expert system, machine learning algorithm and deep learning algorithm in the iron and steel industry, which proves that the emerging artificial intelligence technology can provide technical support for the production quality control and other multi-field problems in the iron and steel industry. At the same time, this chapter also introduces the machine learning algorithm and deep learning algorithm used in detail.

Chapter 3

Research and Application of Scrap Steel Intelligent Rating System

According to the production process of iron and steel products mentioned in Chapter 1, iron and steel production by smelting can be classified into long-process production and short-process production. Scrap steel plays a very important role in both long-process and short-process production. Its quality has a certain influence on the quality of steel products, as well as a great influence on the purchasing cost of scrap steel.

This chapter covers the following main contents: Section 3.1 describes the key challenges of scrap steel intelligent rating, Section 3.2 introduces the overall scrap intelligent rating system mentioned in this chapter, Section 3.3 mainly introduces the experiment on scrap steel intelligent rating system based on computer vision in scrap steel intelligent rating system, Section 3.4 mainly includes the summarization and future outlook.

3.1 Problem Statement

Scrap steel refers to steel wastes (such as scrap edge and scrap end) that do not constitute products in the production process of iron and steel plants and steel materials in scrapped equipment and components, with a component of steel. The scrap steel is sourced from self-produced scrap steel and social scrap steel. Self-produced scrap steel refers to scrap such as scrap edge and scrap end produced by iron and steel production company. Social scrap steel includes waste scrap steel and processed scrap steel. Waste scrap steel refers to the recycled automobiles, steel products, equipment, rails, ships, etc., while processed scrap steel refers to the scrap produced by various industrial sectors when processing steel. At present, the total amount of scrap steel produced in the world is about $300-400$ million tons per year, accounting for about $45-50\%$ of the total steel production, of which $85-90\%$ is used as raw material for steel-making, and $10-15\%$ is used for casting, iron-making and recycled steel.

Scrap steel is a renewable resource of ferrous metals, which is the product of society and the crystallization of energy. It is an important raw material for steel-making and foundry production for iron and steel enterprises. The metal content of scrap steel is $4-5\%$ higher and carbon content is $3-4\%$ lower than that of pig iron, with less sulfur, phosphorus and other impurities. Therefore, making steel with scrap steel instead of pig iron has the advantages of less solvent consumption, short smelting time, less energy consumption, low cost, high metal yield, less steel slag discharged and less environmental pollution, so scrap steel is the only option to replace iron ore for iron and steel smelting. Scrap steel is mainly used as steel-making additive for long-process production in converter or main raw material for short-process production in electric furnace. Energy consumption and cost can be reduced if scrap steel is used for steel-making, so iron and steel production companies will buy a large number of scrap steel for steel-making every year.

Scrap steel is an important resource for the sustainable development of iron and steel industry, especially an important and indispensable raw material for steel-making in electric furnace, and it is also the best coolant for steel-making in converter. In order not to affect the normal process of steel-making and ensure the quality of finished steel parts, high-quality scrap steel raw material must be selected. The recycled scrap steel is classified into several grades according to cleanliness, dimensions and thickness. At present, most scrap steel is judged by experienced inspectors by visual inspection so as to determine the value of scrap steel. It has many disadvantages, however. First, the scrap steel customers come from all over the world. When judging the grade of scrap steel, judging shoddy goods as quality goods, artificial adulteration, "judging scrap steel with private relationship", etc. often occur, which make the steel plant suffer losses, mainly reflected in the following aspects: 1) the judgment criterion may be different for the scrap steel from different customers due to “private relationship”; 2) internal and external illegal personnel colluded with each other and maliciously adulterated scrap steel, which damaged the interests of steel production plants; 3) the visual inspection result has subjectivity and uncertainty. For the scrap steel of the same quality, the acceptance result may be different if it is inspected in different time periods of a day, in different days, or inspected by different inspectors. Second, the scrap steel contains a lot of dust, which will diffuse in the air during the acceptance of scrap steel, and is harmful to the respiratory system of inspectors. In addition, it will produce high decibel noise during the unloading of scrap steel to the yard, which will damage the hearing system of inspectors. Therefore, an intelligent acceptance system for scrap steel is very necessary, which not only helps to reduce the phenomenon of “judging scrap steel with private relationship”, but also benefits the health of inspectors.

Based on the development of image-related algorithm in artificial intelligence, some steel production companies in northern cities of China have rated scrap steel based on computer vision system, while hardly any of those in southern cities of China done so. That is because the scrap steel in northern China is basically waste industrial scrap steel. Therefore, all such scrap steels in a same truck is of same type and in same thickness, making it easy to identify the thickness and size of scrap steel. However, the scrap steel in the south covers domestic scrap steel and waste industrial scrap steel, with complex material types in general, and each grade of scrap steel involving different varieties and thicknesses, so it is difficult to identify the thickness of each scrap steel by image.

Nanjing Iron and Steel Company, located in the Yangtze River Delta region in southern China, is also faced with the problem that the material type of scrap steel is complex on the whole, and each grade of scrap steel involves different varieties and thicknesses, so it is difficult to identify the thickness of each scrap steel by image. However, there are two channels for Nanjing Iron and Steel Company to purchase scrap steel, namely, (1) purchasing scrap steel from upstream large scrap steel recycling and processing companies, which will be transported to Nanjing Iron and Steel Company by ship, and (2) purchasing scrap steel from adjacent small scrap steel recycling plants or individuals, which will be transported to Nanjing Iron and Steel Company by vehicle. The scrap steel purchased from upstream large scrap steel recycling and processing companies is of better quality relatively, and its impurities are mainly dust and a small amount of other impurities. The scrap steel intelligent rating system based on computer vision system and manual experience method proposed in this chapter is mainly used to rate the scrap steel purchased from upstream large scrap steel recycling and processing companies and transported by ship. The system mainly includes three parts: detection of chucks and carriages, identification of scrap steel grade and rating of the material type for the whole truckload of scrap steel.

In this chapter, at first, YOLOv3 object detection algorithm is used to detect chucks and carriages, to delete some invalid images useless for scrap steel rating, and at the same time, to obtain the input needed for the next recognition algorithm - carriage zone. Then, VGG16 neural network is used to identify the sub-grades of scrap steel, delete the images empty in carriage zones, rating the non-empty carriage zones and calculate the proportion of each grade. Finally, the whole-vehicle rating results are modeled according to the analysis results of all images collected for the whole truckload of scrap steel and manual labels. At present, by comparing the system ratingresults with the manual rating results, we can see that the average matching degree of main material type reaches 80%, that of impurity deduction reaches 86%, and that of attached material type is about 10%. It can be seen that the average matching degrees of the main material type and that of impurity deduction obtained based on our scrap steel intelligent rating system get satisfactory results.

By far, annual scrap steel purchase volume of Nanjing Iron and Steel Company is, based on investigation, about 1.6 million tons, and the current purchase unit price fluctuates between CNY $2,500-2,800$, with a price difference of about CNY $20-100$ for different scrap steel grades judged. On-site rating is made as follows: first, on-site inspectors rate it based on their experience, and then the shift leader makes the final judgment according to the opinions of the inspectors. Therefore, because of the different grades judged, the annual amount difference is roughly within CNY 96 million (calculated based on an average price of CNY $2,650$ and the difference of CNY 60 by grade). However, the scrap steel purchased from upstream large scrap steel recycling and processing companies and transported by ship accounts for the vast majority of scrap steel purchased by Nanjing Iron and Steel Company. Therefore, using the scrap steel intelligent rating system, which can make standardized and rational judgment of scrap steel grade, can reduce some artificial subjective factors and emotional factors, thus saving significant costs for Nanjing Iron and Steel Company.

3.2 Introduction of Scrap Steel Intelligent Rating System

This section mainly introduces the components of scrap steel intelligent rating system and the functions of its main modules of each part.

The scrap steel intelligent rating system proposed in this chapter mainly includes some ERP systems and metering systems existing in iron and steel production companies, an artificial intelligence platform for rating scrap steel through computer vision, and some hardware devices, such as cameras and self-service machines. Components of scrap steel intelligent rating system are shown in Figure 3.1:

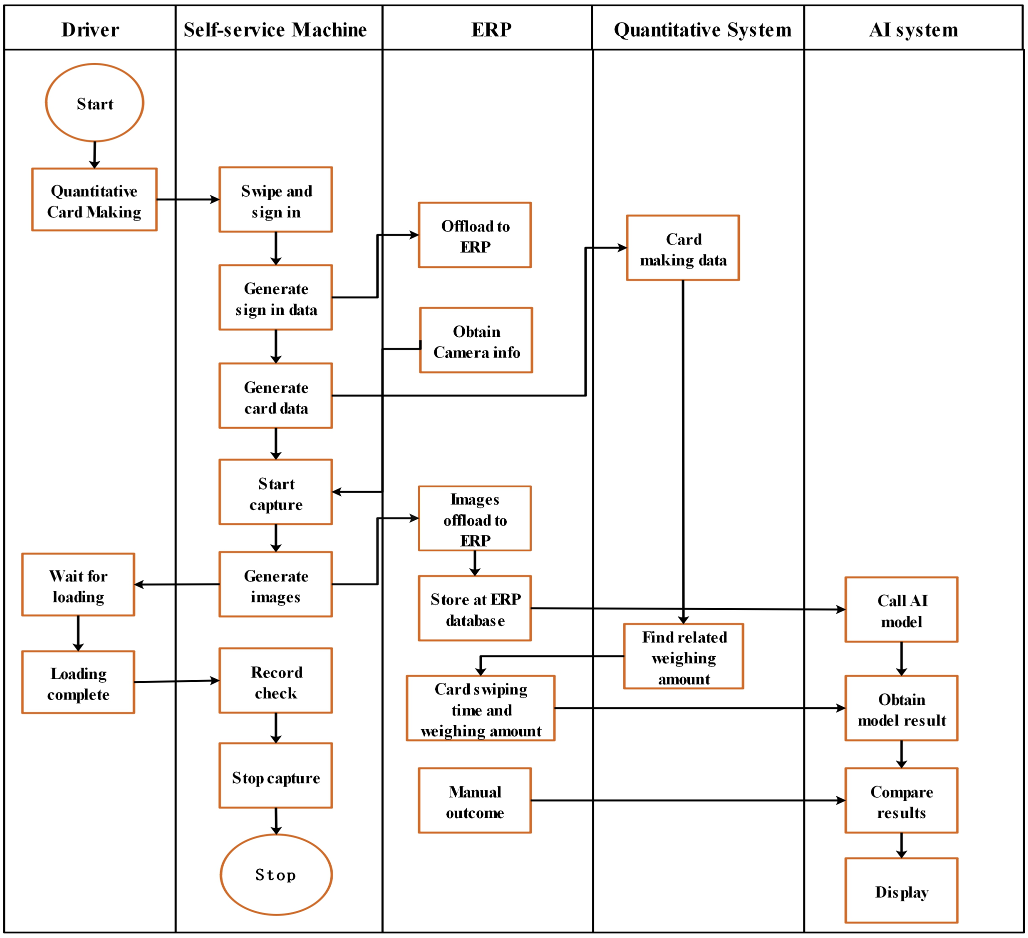


Figure 3.1: Flow Chart of Scrap Steel Intelligent Rating System.

(1) Driver operation module:

In this system, the driver is responsible for swiping card for check-in on the self-service machine after stopping the truck at the loading point, and swiping card again for ending after the truck is full. The function of check-in by swiping card is to trigger the camera to start taking images, while the function of ending by swiping card is to trigger the camera to stop taking images. If the driver of the previous truck forgets to indicate ending by swiping card, the image taking of the previous truck will automatically end when the driver of the next truck makes heck-in by swiping card.

The module for driver making check-in by swiping card refers to that the driver places the IC card on the card reader of the self-service machine, selects the check-in function, loading portal crane and loading point on the interface of the self-service machine, and inputs the license plate number. Then, the system reads the physical card-number of the IC card, obtains the latest card-making record in the metering database, and displays the planned order number and ship number in the card-making record on the interface of the self-service machine for the driver to verify and confirm. After confirming it, the driver can click OK.

The module for driver confirming ending by swiping card refers to that the driver places the IC card on the card reader of the self-service machine and selects the confirmation function on the interface of the self-service machine. Then the system obtains the physical card number of the metering card, finds the corresponding check-in record, and changes the check-in status to confirmation, after which, the system stops taking images of the truck and post back the data.

Metering card making refers to that: after the unloading truck arrives at the dock, the driver should obtain a card made in the dispatching room of the logistics center in the office building, that is, to bind the license plate number, planned order number, ship number and other information with the metering IC card. It does not require drivers to make cards for each transport, but only when the information of ship number and planned order number changes.

(2) Self-service machine system module:

The modules of checking in and confirming ending by swiping card is for the interaction between the driver and the self-service machine, and the specific operation has been introduced in the driver operation module.Generating check-in data refers to that: after the driver swipes the card to check in on the self-service machine, the system stores the check-in information verified by the driver, including license plate number, planned order number, check-in timestamp, loading point, ship number, sign for starting snapping and other information into the database.

Generating card-making data refers to that: after the driver swipes the card to check in on the self-service machine, the system will send the check-in information verified by the driver, including license plate number, planned order number, check-in timestamp, ship number and other information, and post the check-in information to the metering system synchronously, and the database of the metering system will add such check-in records.

Self-service capture module of the system, including two sub-modules: starting capturing and generating local images, refers to that a check-in record forms after thedriver checks in. After finding the corresponding crane and loading point information and the sign to start capturing through the check-in record, the system identifies the information of the camera to capture images through the interface and calls the camera for capturing. The system has set up regular snapshots, which are taken every 40 seconds. The captured images are named and saved to the local server by combining the check-in timestamp and the number of snapshots.

(3) ERP system module:

Camera information maintenance module in ERP system: because this scrap steel intelligent rating involves several portal cranes for loading scrap steel and each portal crane has two sides, there are many cameras, making it very important to call the corresponding camera to take images after the driver checks in. Therefore, the information of each camera should be maintained in the ERP system so that the camera can be called by the self-service machine system. The main information of the camera includes IP, portal crane number, loading point and other information.

Storage of images captured: the image capture program of the all-in-one machine will be triggered after the driver checks in by swiping card. At that time, an image will be captured per 40 seconds, and the captured image information will be stored in the ERP database through interface. At the same time, the captured image will be stored under the corresponding server path, and the two will be correlated to facilitate the subsequent calling for AI rating.

Obtaining of weighed net weight: AI intelligent rating gives the proportion of main material type, attached material type and impurity deduction, but it is necessary to give the weight of main material type, attached material type and impurity deduction of scrap steel on each truck, so it is necessary to accurately calculate the weight of main material type, attached material type and impurities of scrap steel on each truck to obtain the net weight by weighing. In the ERP module, the weighing data is posted from the metering system to ERP in the form of DI so that the weighing data is correlated with the captured images to obtain the weighed net weight of the current truck. Then, the weight of the main material type, attached material type and impurities of each truck given by the AI rating model are calculated by setting batches and batch size.

(4) Metering system module:

Weighing quantity matching refers to that after a truck is weighed, each weighing data is spliced with the check-in time of the truck (the check-in time is determined by screening out the latest check-in record by the plan number and the car number) to form a new weighing record, which will be posted back to ERP by executing the operation at the metering end regularly.

(5) AI rating module:

In this system, the AI rating module is mainly responsible for giving the final rating results through AI algorithm based on the input pictures of scrap steel. The flow of AI rating module is as shown in Figure 3.2. When pictures are input into AI rating module, the following steps will be taken:

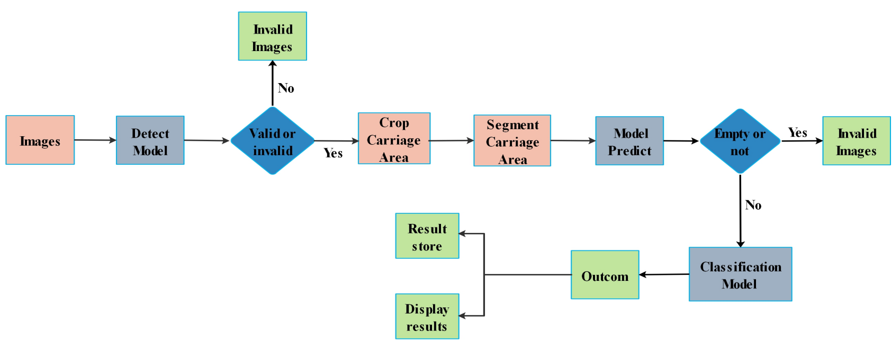


Figure 3.2: Flow of AI Rating Module.

Step 1: Firstly, the picture will be subjected to the carriage and electromagnetic chuck detection algorithm, if no carriage is detected, or if carriage and electromagnetic chuck are detected and the carriage is shielded by the electromagnetic chuck, the picture is deemed as invalid and the model stops operating. If only the carriage is detected or the carriage and electromagnetic chuck are detected and the carriage is not shielded by the electromagnetic chuck, the picture is regarded as valid and the carriage zone in the picture is taken for the next operation;

Step 2: Divide the carriage zone taken from the picture into 12 disjoint small blocks;

Step 3: Classify each small block obtained in step 2 by using a classification algorithm, and if all the small blocks are empty carriages, the picture is regarded as invalid, and further operation will be stopped. If any one small block is a non-empty carriage class, calculate the proportion of that class to the non-empty carriage class. The classification algorithm applied is VGG16 neural network, it has been introduced in details in section 2.3.1.5.

Step 4: Combine the proportion of each material type calculated in Step 3 and the rating data accumulated in the process of acceptance by manual inspector to give the main material type, the proportion of the main material type, the attached material type, the proportion of the attached material type, and the proportion of impurities for each carriage of scrap steel.

Step 5: Results display. The display content in the page of AI intelligent rating platform mainly includes the rating report of whole truckload of scrap steel, the scrap steel rating matching degree report and the real-time display image for intelligent rating of scrap steel. The real-time display image for intelligent rating of scrap steel mainly covers the information such as captured current picture, ship name, license plate number, time of check-in through card swiping, as well as the main material type, the proportion of the main material type, the attached material type, the proportion of the attached material type, and the proportion of impurities given by module intelligent rating. The rating report of whole truckload of scrap steel mainly covers the information such as ship name, license plate number, time of check-in through card swiping, net weight weighted, as well as the main material type, main material type weight, attached material type, attached material type weight, and impurity weight obtained by manual rating and system rating. The scrap steel rating matching degree report mainly shows the matching degree of manual rating and system rating in main material type weight, attached material type weight and impurity weight.

(6) Main hardware construction: In the intelligent scrap steel rating system mentioned in this chapter, some hardware is required to support the image acquisition and storage, image trigger signal and the input of current loading information. The equipment used in the system and the functions are introduced in detail below:

A. Construction of self-service machine:

The self-service machine mainly realizes the trigger function of automatic capture by the on-site camera and the function of input the basic information of the vehicles carrying scrap steel. It is considered to install self-service machines near the portal cranes, and the portal cranes that need to be provided with self-service machines are No.2, No.3, No.6, No.7, No.8 and No.10. The self-service machine mainly consists of a terminal and a rainproof outer cover.

B. Construction of camera:

In this chapter, the scrap intelligent rating scenario is the process from unloading the scrap steel from the ship to loading it to the vehicle on the wharf, with the focus on loading process. Therefore, the monitoring system mainly monitors the loading process of scrap steel and captures at regular intervals to provide guarantee for the demand of the scrap steel intelligent rating system in the background. According to the system planning, the portal cranes that need to be provided with cameras are No.2, No.3, No.6, No.7, No.8 and No.10. Each side of the portal crane is equipped with one starlight intelligent dome camera; as a result, 12 cameras in total are needed to be installed on 6 portal cranes for image acquisition.

C. Construction of server:

In the scenario for which the system is operated, the scrap steel intelligent rating process requires each camera to capture scrap steel pictures with an interval of 40 seconds. The camera shall be of 8 megapixel UHD camera, and the size of each picture will reach 15MB. According to the investigation on the actual number of operating vehicles at the site, there are about 140 vehicles carrying scrap steel at the wharf every day, and the average loading time of each vehicle is about 20 minutes, which means that the storage space required for captured pictures is about 0.1TB a day. According to the actual demand, the pictures need to be saved for at least one year, thus about 40TB storage space is needed for storage of pictures captured in one year. As a result, a file server with huge storage space is needed to store the captured pictures of scrap steel.

3.3 Data Experiment

This chapter aims to rate the social scrap steel recovered by Nanjing Iron and Steel Company. According to the scrap steel rating business of Nanjing Iron and Steel Company, five indicators are required for scrap steel to be rated in each vehicle: the main material type, the proportion of the main material type, the attached material type, the proportion of the attached material type, and the proportion of impurities. According to the three experiments in this section, i.e., the experiment in carriage and chuck detection, the experiment in scrap steel rating and the experiment in impurity deduction, the five indicators of the whole truckload of scrap steel can be quantified.

3.3.1 Introduction of Basic Concepts

This section first introduces the scenario of scrap steel intelligent rating of Nanjing Iron and Steel Company and the concepts of valid images and invalid images.

(1) Introduction of the scenario of scrap steel intelligent rating of Nanjing Iron and Steel Company.

At present, the scrap steel warehouse of Nanjing Iron and Steel Company is divided into indoor warehouse and outdoor warehouse. The indoor scrap steel warehouse is relatively small, and the scrap steel transport vehicle can only unload through dumping, which will lead to the accumulation of scrap steel. At this time, only a small amount of scrap steel can be captured, which is not sufficient to represent the material type of the scrap steel on a whole truck. The outdoor scrap steel warehouse occupies a very large area, so it is not a good way to define a certain point for unloading and acquiring pictures. If a certain point is specified as the scrap steel unloading point, it may be necessary to carry out secondary transportation to dump the scrap steel to the point where it really needs to be unloaded, which will increase the transportation cost and reduce the overall scrap steel rating efficiency. If multiple scrap steel rating points are established to avoid secondary transportation, a large number of rating points need to be built, which will reduce the storage capacity of the scrap steel warehouse and may affect the steel production process. Taking into account various factors, Nanjing Iron and Steel Company's scrap steel intelligent rating point is arranged at the wharf, and the process of transporting scrap steel unloaded from ships to the scrap steel warehouse in batches via transport trucks is set as the process of scrap steel intelligent rating.

The cameras for acquiring pictures are installed on both sides of the crane for unloading from the ship and loading to the vehicle. Since the crane can rotate, the process of scrap steel loading and unloading can be carried out on both sides. The start and end of image acquisition is controlled by the driver triggering the self-service all-in-one metering machine. The overall scene diagram of the intelligent scrap rating point is shown in Figure 3.3. Since the real scene is too large to be captured comprehensively in a single picture, we apply 3D diagram to restore the real scene. In Figure 3.3, the blue wave shape represents the ship transport harbor; the brick-red machine represents the crane, which is used to load the scrap steel in the ship in batches to the truck on the shore; the green and white machines represent the self-service all-in-one machine, which is used to trigger the camera to take pictures, the camera and the patch light are installed on the left and right of the crane, the specific physical scene is shown in Figure 3.4; the camera and the patch light are installed in the blue operation room of the crane. The camera and patch light are installed in the blue operation room of the crane on the large disk below, marked by red boxes in the figure.

When the driver triggers the camera to take pictures by swiping the card, the camera will take pictures once every 40 seconds (the interval between the crane's electromagnetic chuck appearing above the carriage). Some of the pictures taken contain electromagnetic chuck shielding, open space, etc. These pictures have no contribution to the intelligent rating of scrap steel, so no judgment will be made on them. Therefore, it is necessary to establish a model to identify whether the picture only contains open space and whether there are a large overlapping zone between the electromagnetic chuck and the carriage.

(2) Introduction of valid and invalid pictures

Valid pictures refer to those that contribute to the intelligent rating of scrap steel, that is, pictures with scrap steel not shielded. Pictures only contain the carriage with scrap

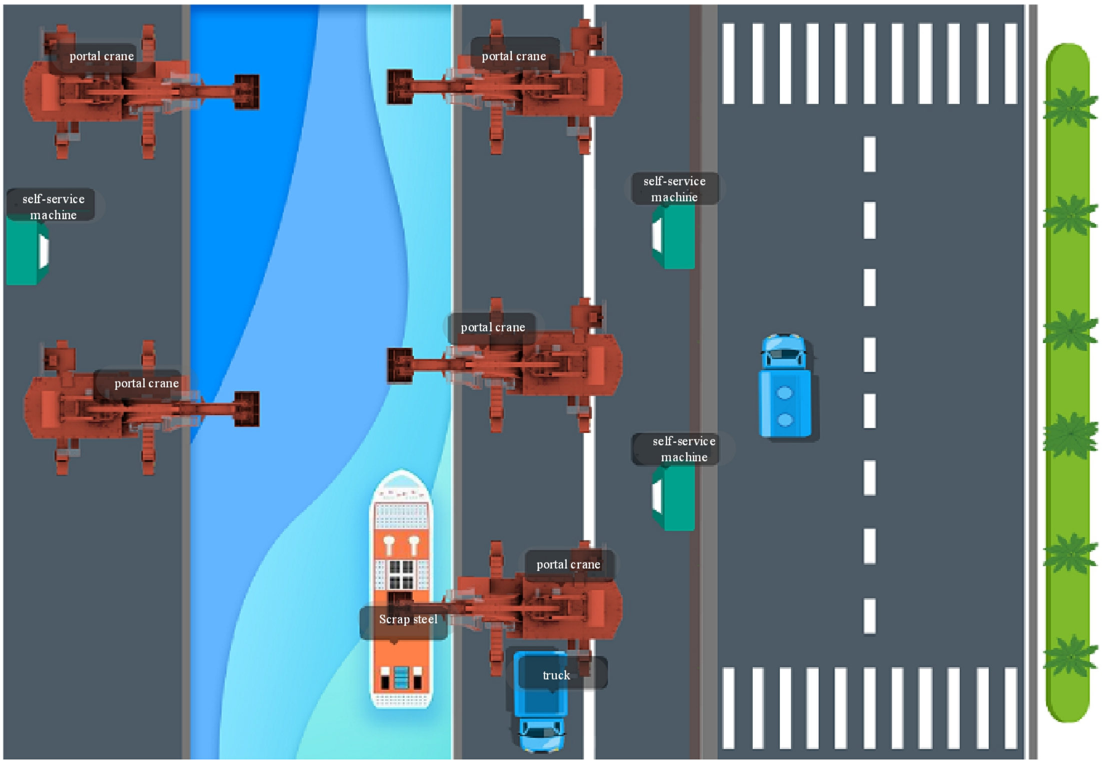


Figure 3.3: 3D view of the intelligent scrap rating scene.

steel, without electromagnetic chuck, or with the electromagnetic chuck not shielding the carriage with scrap steel, are deemed as valid, as shown in the Figure 3.5 below:

It can be seen from the figure that the picture in Figure 3.5(a) contains electromagnetic chuck and the carriage with scrap steel, but the electromagnetic chuck does not shield the carriage, and the scrap steel in the carriage is clearly visible, so it is valid. The picture in Figure 3.5(b) only contains the carriage with scrap steel, and there is no other interference and shielding, so it is valid.

Invalid pictures refer to those that have no contribute to the intelligent rating of scrap steel, that is, pictures with shielding or pictures without scrap steel. Pictures only contain electromagnetic chuck, empty carriage, open space and those having overlapping between electromagnetic chuck and carriage are deemed as invalid, as shown in Figure 3.6:

It can be seen from Figure 3.6 that the picture in Figure 3.6(a) contains electromagnetic chuck and the carriage with scrap steel, and the electromagnetic chuck is located above the carriage with scrap steel, resulting in a large area of shielding on the carriage with scrap steel. The scrap steel of Nanjing Iron and Steel Company is loaded from the

Figure 3.4: Schematic diagram of the crane and camera installation position in the actual scene.

(a) No overlapping between the carriage and the chuck

(b) Only capturing the carriage

Figure 3.5: Schematic Diagrams for Valid Pictures.

(a) With overlapping between the carriage and the chuck

(b) Only capturing the chuck

(c) Open space

(d) Empty cars

Figure 3.6: Schematic Diagrams for Invalid Pictures.

ship with electromagnetic chuck into the truck carriage. Figure 3.6(a) shows a situation where the electromagnetic chuck have not yet moved away or just moved to the right position to put the scrap steel into the carriage, resulting in the invisibility of the newly loaded scrap steel. Because the scrap steel is clearly shown in the previously captured picture, the visible scrap steel captured this time does not contribute to the intelligent rating, then we consider such pictures to be invalid. The picture in Figure 3.6(b) only contains the chuck, the picture in Figure 3.6(c) only contains an open space, and the picture in Figure 3.6(d) only contains an empty carriage. These pictures do not capture the contents to be rated, so they are regarded as invalid.

3.3.2 Carriage and Chuck Detection Experiment

In this section, whether the picture contains electromagnetic chuck, whether there is an overlapping zone between electromagnetic chuck and carriage, and whether the picture contains carriages is determined by detecting the position information of carriages and electromagnetic chuck. If the detection result contains any of Figure 3.6(a), Figure 3.6(b) or Figure 3.6(c) in Figure 3.6, it is an invalid picture. If the detection result contains any one of Figure 3.5, it is a valid picture and the carriage zone is taken according to the carriage location information. If the detection result is as shown in Figure 3.6(d), then we also consider the picture as valid but will remove it during scrap steel material type identification.

There are $2,000$ pictures containing carriages and chucks in the data set of this experiment. Among them, $1,488$ pictures are used as train set, 204 pictures are used as validation set and 308 pictures are used as test set. The dataset contains pictures with electromagnetic chucks and carriages in daytime and nighttime scenarios, pictures with locations of electromagnetic chucks and carriages including cases with and without overlapping zones, pictures with carriages only, and pictures with chuck only.

The algorithm used in this experiment is the YOLOv3 neural network. The structure of this neural network and its specific process of object detection have been described in Chapter 2. The YOLOv3 neural network achieves a significant improvement in detection speed under the condition of obtaining considerable accuracy. Generally, it is $1,000$ times faster than R-CNN and 100 times faster than Fast R-CNN The target to be detected by the industry requires high real-time performance. The YOLOv3 neural network has been well received by the industry for its fast and accurate characteristics. Therefore, YOLOv3 neural network is also used for training and reasoning in the carriage and electromagnetic chuck detection experiments in this section.

During the process of using YOLOv3 neural network to train the data set, the training environment is a GPU server containing two TeslaP100 graphics cards. During the training process, the batch size is set to 64, subdivisions is set to 16, momentum is set to 0.949, decay is set to 0.0005, and learning rate is set to 0.001. During the training process, the decline curve of loss and the change curve of mAP on the validation set of the model change with the number of iterations as shown in Figure 3.7. The blue and red curves in the figure represent the change curves of loss and MAP with the increase of iteration times respectively. It can be seen from the figure that the curve represents that loss first decreases rapidly with the increase of iteration times, then slowly decreases, and finally tends to be stable. The average loss in the stable stage is about 0.3638. The curve represents that mAP increases first and then tends to be stable with the increase of iteration times, and the final mAP is about 99%. This illustrates that the YOLOv3 neural network fits the validation set of the carriage and chuck dataset very well, and the model is also very meaningful for using the test set to conduct test.

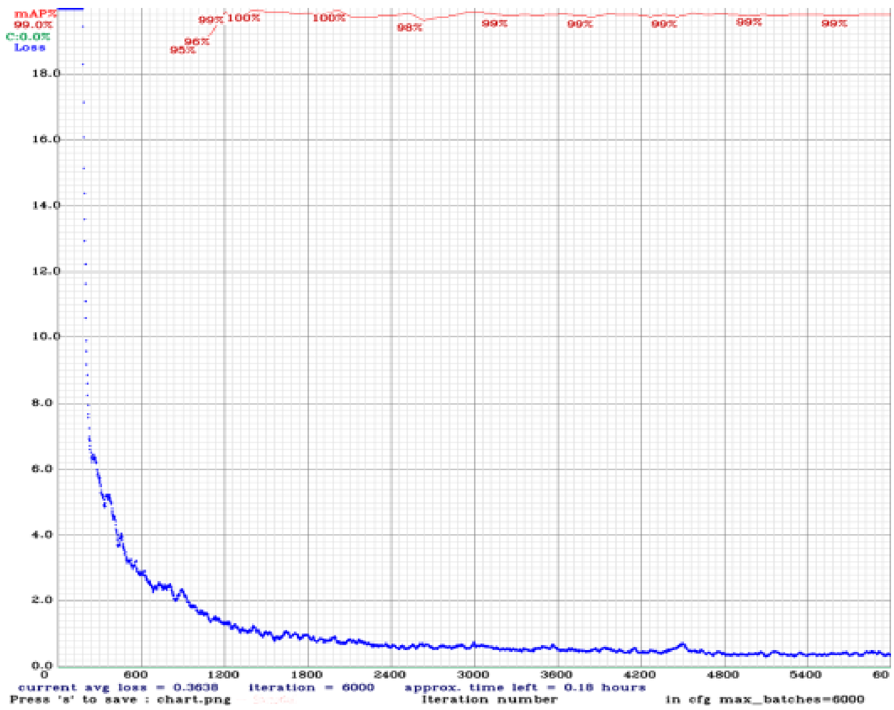


Figure 3.7: The Optimization Process of the Hidden Layer in BP Neural Network.

Use the above trained weight file combined with the forward network of YOLOv3 in opencv and NMS (non-maximum suppression) algorithm to reason on the test set, and the mAP on the test set is 99.27%. In the test set, the number of electromagnetic chucks is 308, and there are 9 unrecognizable ones, and these 9 pictures only contain a small part of the chuck, and there is no overlapping zone with the carriage, which will not affect the subsequent operation of the system. The number of carriages in the test set is 305, and there are 8 carriages that are not recognized by the model, which means that one of the pictures taken for a truck when it is loading with scrap steel is not identified, which will hardly affect the subsequent operation of the system. The above test results show that the model can detect carriages and electromagnetic chucks well, which provides the first guarantee for the scrap steel intelligent rating model. The detection results of carriages and electromagnetic chucks are shown in Figure 3.8, and the positioning effect of the model on carriages and electromagnetic chucks is as shown in Figure 3.9.

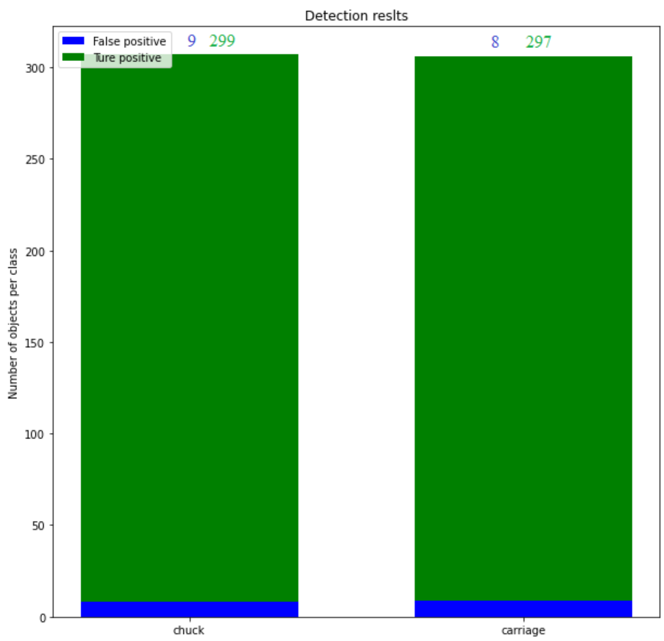


Figure 3.8: Detection Results of YOLOv3 Model on the Test Set.

3.3.3 Scrap Steel Material Type Rating Experiment

The classification process of scrap steel material type is to take the carriage zone in the picture detected as valid in the previous section, and then divide the taken carriage

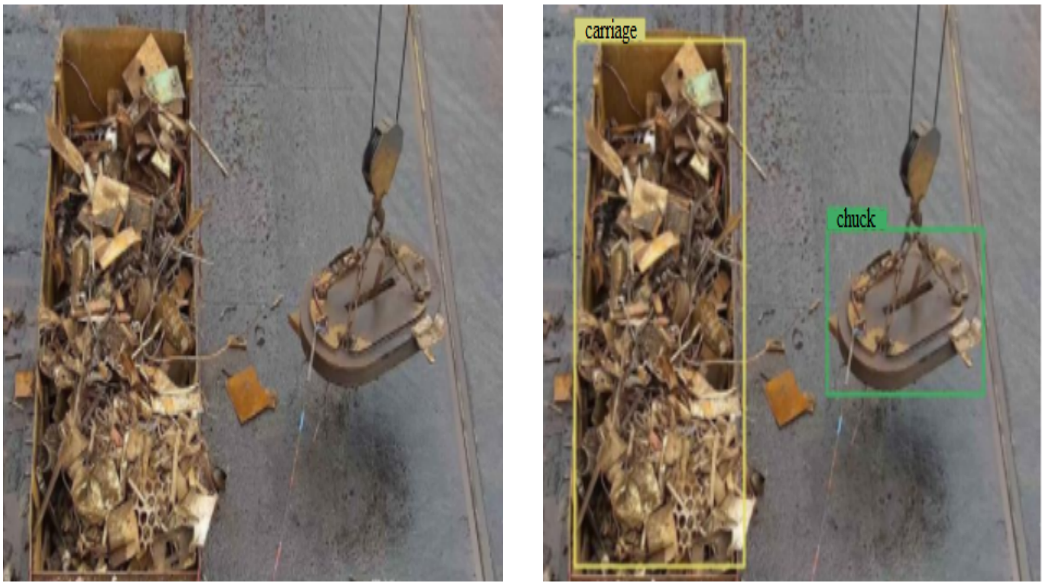


Figure 3.9: Positioning Effect of YOLOv3 Model on Carriages and Electromagnetic Chucks.

zone into 12 small pictures, and mark the material type of each small picture and sort it into a data set. The labels were rated by human experts manually. If all the 12 small pictures contained in the carriage zone are empty, the carriage zone is regarded as an invalid region, and if there contain non-empty pictures, the carriage zone is regarded as a valid region.

According to the actual situation of scrap steel recovery, fine cutting, heavy scrap and shear material can all be regarded as composed of three sub-materials, and some carriages may be empty at the beginning of unloading or loading. Therefore, the classification of materials introduced in this section can be divided into four categories. The visualization of the three material types (i.e., fine shear, heavy scrap and shear material) is shown in Figure 3.10: 3.10(a) for the fine shear material type, 3.10(b) for the heavy scrap material type and 3.10(c) for the normal shear material type.

The algorithm used in this experiment is VGG16 neural network, and the structure of this neural network has been introduced in Chapter 2. By using the series of small convolution kernels to replace the large convolution kernels in other neural networks, there will be fewer parameters, and more nonlinear transformations than a single convolution layer, thus ensuring high accuracy. The layer of this neural network is less than that of ResNet series neural networks, so it operates very fast. Combining these two advantages, VGG16 neural network is used for training and reasoning in scrap steel

(a) fine shear (b) heavy scrap (c) shear material

Figure 3.10: The visualization of the three material types.

material type experiment in this section.

In this experiment, the data set contains $9,641$ pictures, including $6,621$ as training set, $1,893$ as verification set and 947 as test sets, including picture data in daytime and night scenarios. In this experiment, VGG16 neural network is used to train the data set. The training environment is a GPU server containing two TeslaP100 graphics cards, and the operating system is Ubuntu16.04. The model uses SGD as the optimizer, the learning rate is set to 0.001, the decay is set to 1e-6, and the momentum is set to 0.9. A total of $1,000$ epochs are trained in this experiment. The training process is divided into two stages. In the first stage, all convolution layers are frozen, the weights on imagenet are loaded as pre-training weights, only the final classification layer is trained, and 300 epochs are iterated; in the second stage, fine tune is further performed, and the last three layers of vgg16 convolution layer are set to a trainable state, and 700 epochs are continuously trained.

In the training process, the decline curve of loss and the change curve of accuracy of the model on the training set and the validation set vary with the iteration times are as shown in Figure 3.11. In the figure, the curve change before the green line is the curve change of the first stage training process, and the curve change after the green line is the curve change of the second stage training process. It can be seen from the figure that the accuracy in the second stage training process is higher than that in the first stage, and the loss in the second stage training process is lower than that in the first stage. Training and optimizing the last three layers of vgg16 convolution layer may improve the overall accuracy of the model.

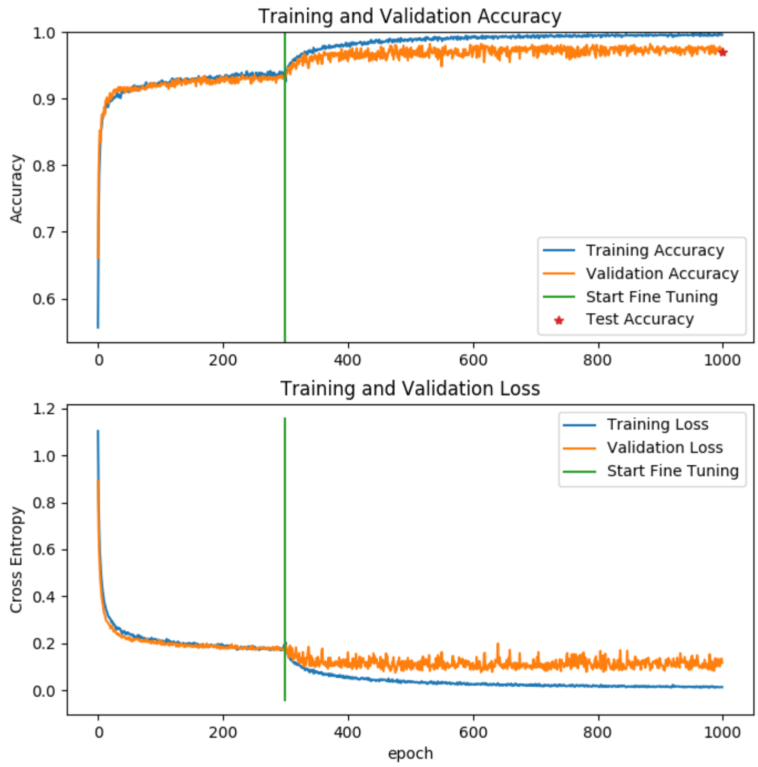


Figure 3.11: Change Curves of Loss and Accuracy during VGG16 Training.

Use the above trained weight file combined with the forward process of VGG16 neural network to reason on the test set, and the accuracy rate on the test set is 96.21%. According to the above test results, the model can well predict the scrap steel sub-material type and empty car picture, which provides the second guarantee for the scrap steel intelligent rating model.

After the VGG16 model outputs the proportion of each sub-material type, if the carriage zone is not empty and there are no empty carriage images in 12 small pictures, the proportion of each material type is calculated; if the carriage zone is not empty and there are several empty car images in 12 small pictures, the proportion of each material type is calculated after the empty carriage pictures is removed.

3.3.4 Prediction and Evaluation of the Results of the Whole Vehicle of Scrap Steel Material Type

This section mainly introduces the prediction and evaluation methods for the results of scrap steel material types. The final output result of scrap steel material type includes five parts: main material type, proportion of main material type, attached material type, proportion of attached material type and impurity deduction rate. The main material type refers to the scrap steel material type that accounts for a large proportion in the truck, while the attached material type refers to the scrap steel material type that accounts for a small proportion in trucks. Impurities refer to non-steel objects, such as wood, rubber, plastics, and muck, as well as steel wire ropes, and other objects that have an impact on the steel-making process.

3.3.4.1 Prediction of the results of the whole vehicle of scrap steel material type

As the scrap steel detection scenario of Nanjing Iron and Steel Company is from unloading from the ship and loading to the truck, the focus in this process is on the truck, that is, the detection scenario is the loading process. Impurities are easily blocked or infiltrated during loading. It is very inaccurate to deduct the impurities of the whole vehicle only by the impurity quantity shown in the detection pictures. Therefore, we convert the acceptance experience accumulated by the production site inspectors into data and deduct the impurities from the scrap steel in each truck.

To collect all the pictures collected from each car during loading, we shall select the valid pictures containing the carriage zone by using the carriage and electromagnetic chuck detection model, and then predict and reason the valid pictures of each vehicle to calculate the proportion of the three sub-material types using the scrap steel material type rating model. Combined with the results given by manual on-site acceptance, the data set contains 700 pieces of data, covering three types: fine cutting, heavy scrap and shear materials.

If a new data is available, calculate the Euclidean distance between this data and all the data in the data set, and select the manual acceptance result corresponding to the data closest to this data to calculate the predicted output of this data. If only one data is closest to this data and only one corresponding manual acceptance result is available, the manual acceptance result will be used as the prediction output of new data; if only one data is closest to this data and multiple corresponding manual acceptance results are available, the impurity deduction ratio is the average of impurity deduction ratios in multiple acceptance results, and the proportion of attached material types is the maximum value of multiple records, and the proportion of main material types equal to 1 minus the impurity deduction ratio and then minus the proportion of attached material types. If multiple data is closest to this data, the data with attached material type in the result is selected as the result output.

3.3.4.2 Assessment methods of model accuracy:

The accuracy of AI intelligent rating model of scrap steel is assessed based on the results of manual acceptance, the main assessment indexes include the matching degree of main material type, the matching degree of attached material type and the matching degree of impurities. The calculation method is as follows:

$$p_1 = \frac{W_{(pm==sm)}}{W_{pm}} \qquad \text{(Equation 3.1)}$$

In this equation, p_1 indicates the matching degree of main material type between system rating results and manual rating results; $W_{(pm==sm)}$ indicates the total weight of main material types when the main material types in system rating results match those in manual rating results; W_{pm} indicates the total weight of main material types in manual rating.

$$p_2 = \frac{W_{(sa==pa)|(sm==pm)}}{W_{pa}} \qquad \text{(Equation 3.2)}$$

In this equation, p_2 indicates the matching degree of attached material type between system rating results and manual rating results; $W_{(sa==pa)|(sm==pm)}$ indicates the total weight of attached material types when the attached material types in system rating results match those in manual rating results on the premise that the main material types of system rating results match with those of the manual rating results; W_{pa} indicates the total weight of attached material types in manual rating.

$$p_3 = \frac{W_{sz|(sm==pm)}}{W_{pz}} \qquad \text{(Equation 3.3)}$$

In this equation, p_3 indicates the matching degree of impurity deduction between the system rating result and manual rating result; $W_{sz|(sm==pm)}$ indicates the total weight of impurities deducted by the system on the premise that the system rating result of the main material type matches that of manual rating result; W_{pz} indicates the total weight of impurities deducted in manual rating.

The scrap steel intelligent rating system proposed in this chapter has been put into on-line trial operation in Nanjing Iron and Steel Company. When the above assessment methods are adopted to assess the system, the average matching degree of main material type reaches 80%, that of impurity deduction reaches 86%, and that of attached material type is about 10%. It can be seen that the average matching degrees of the main material type and that of impurity deduction obtained based on our scrap steel intelligent rating system get satisfactory results. The low average matching degree of attached material type is due to that different manual inspectors will give different results for the same material type. Some people think that it is enough to add attached material type, while others think that the main material type shall be degraded. However, according to historical data, the system predicts that a part of main material types shall be degraded, while attached material type shall be added for other parts of material types.

3.4 Summary

Under the background of industrial digitalization and intelligent transformation, with the rapid development of artificial intelligence technology, the scrap steel intelligent rating system is proposed in this chapter, this system combines the existing ERP system and metering system, image acquisition equipment and artificial intelligence platform of iron and steel enterprises to rate the scrap steel. The most important core of the scrap steel intelligent rating system is the computer vision algorithm in the artificial intelligence platform, which combines object detection and image classification algorithms to judge the material type of scrap steel. After the on-line trial operation of scrap steel intelligent rating system proposed in this chapter, the average matching degree of main material type given by the system and manual inspectors reaches 80%, that of impurity deduction given by the system and manual inspectors reaches 86%, so the accuracy of system proposed by us is very good.

In the acceptance results, although the accuracy of average matching degrees of main material type and impurity deduction given by the on-site inspector and the method proposed in this chapter is high, there are still some shortcomings in this system.

In terms of system effectiveness and stability, it is mainly reflected in the following aspects:

(1) At present, the driver can swipe the card again immediately after swiping the card, which makes the system only capture a few photos and cannot represent all the material types on the vehicle. Later, the function of control over the swiping time will be added to avoid swiping the card two times within a short time;

(2) When the previous driver forgets to swipe the card, the pictures will be taken continuously. When the next driver fails to swipe the card on time, the pictures of the next vehicle will be recorded in the data of the previous vehicle, which will affect the rating accuracy of scrap steel. The current method is that a swiping record can take photos for 30 minutes, and then the current photo time will be optimized according to the vehicle type and the loading time of each kind of vehicle.

(3) The crane for unloading scrap steel from ships and loading scrap steel on trucks is large, and the working arm of the crane can rotate 360 degrees, so the crane can operate at the front or rear end of the main body. The driver needs to choose which side to load when swiping the card, but sometimes the driver may choose the wrong direction. At present, there is no control over this situation, and the assessment and alarming functions will be added later. When the driver chooses the wrong direction, the system will assess the captured pictures and finally give an alarm.

(4) At present, in the process of loading, the driver may sometimes forget to swipe the card, and fails to confirm whether there is acceptance when unloading, this situation will lead to that the acceptance of scrap steel in this vehicle may be missed. Later, the metering system will confirm whether the card is swiped within an appropriate time before unloading.

(5) The current system has no message push function, and the driver doesn't know whether he/she has successfully swiped the card after swiping the card. This will lead to repeated card swiping by the driver, and the push function of swiping information and wrong point information will be added later to ensure the effective operation of the system.

(6) At present, the focal length of the camera is fixed to take photos, but the size of the truck loaded with scrap steel is not fixed, when the truck is a minivan, the material type of scrap steel in the truck will be slightly blurred. Later, we will consider autofocusing the camera according to the vehicle type, so that the photos of scrap steel type in the truck is very clear.

(7) At present, only a few material types can be identified by the system. In the future, other common material types will be trained and included in the identification range.

(8) Since the results of manual acceptance are somewhat different from those of system acceptance, it is considered that the manual acceptance inspector will mark the proportion of various material types in each picture in the future to improve the accuracy of the system model.

(9) The scrap steel intelligent rating method based on computer vision algorithm proposed in this chapter is used to deduct the impurity, the mass of impurity may be deducted accurately; but because there are few non-dust impurities in shipping scrap steel purchased from upstream large scrap steel recycling and processing companies, small data volume cannot train a good model for type identification. Therefore, through the accumulation of data, the types of impurities will be identified in the subsequent research.

The existing algorithms use equal divisions on the area of the effective carriages, and determine the scrap grade in combination with sub-material types. The Euclidean distance between the data is used to calculate the proportion of impurity. This algorithm can not predict the type and proportion of scrap steel and the type and proportion of impurity more accurately. In order to improve the accuracy of the current algorithm, in the following research, we mainly consider the combination of machine learning algorithm and image segmentation algorithm to determine the grade of scrap steel and calculate the weight of impurities. The image segmentation algorithm is used to identify and calculate the various types of steel scrap and the area of each type of steel scrap in the effective picture, the type of impurity and its corresponding area. The machine learning algorithm is used to fit the results produced by the image segmentation algorithm and the corresponding manually-given scrap grades and the weight of impurities.

Chapter 4

A Multiple-Factor-based Detection System of Longitudinal Surface cracks

According to the production process of iron and steel products introduced in the first chapter, iron and steel smelting can be divided into four processes: iron smelting, steel smelting, continuous casting and rolling. In the previous chapter, the intelligent rating method of quality of scrap steel, which is an important raw material in steel smelting process, is mainly introduced. According to the sequence of production process, this chapter will introduce the intelligent prediction method of longitudinal cracks on slab surface in continuous casting process. The successful application of this method can reduce the quantity of manually inspected slabs, thus reducing costs and increasing efficiency.

In this chapter, the main contents are as follows: Section 4.1 mainly puts forward the background of the surface longitudinal crack prediction of continuous casting; Section 4.2 introduces the details for the five-layer industrial Internet of Things platform proposed in this chapter; Section 4.3 mainly introduces the influence of single factor on longitudinal cracks on the surface of continuous casting slab; Section 4.4 mainly introduces the optimization process of pre-optimised post-feedback (BP) neural network; Section 4.5 mainly introduces the experiment and contrast of data collected by five-layer industrial Internet of Things platform proposed in this chapter of post-feedback (BP) neural network and pre-optimised BP neural network; Section 4.6 mainly includes the content summary of this chapter and the prospect of future work.

Since the rating results need to be output and displayed in real time during the intelligent rating of steel scrap, the AI intelligent rating model based on deep learning needs to be deployed to a high-performance, computationally powerful GPU server to make real-time predictions on the input HD pictures and feedback the rating results for reference and decision-making by buyers and suppliers.

D. Construction of light supplement equipment:

In the study of this chapter, the process of scrap steel rating involves the night scenarios. Because the wharf is an outdoor open place, it is difficult to supplement light, and the existing high-pressure sodium lamp will have a certain impact on the white light supplementing effect. After discussion with several optical equipment manufacturers, the high-power LED is finally selected for light supplement. Through simulation test, it was proved that high-power LED can realize local light supplement of 200lux.

4.1 Problem Description

Comparing with traditional method, many methods may be used to save the cost in the smelting process of iron and steel industry. For example, almost all steel plates can be produced by continuous casting technology, and hot charging and direct rolling technology are increasingly used in the production process[166, 167]. In order to save costs, the hot steel plate can be directly put into the heating furnace or rolled in the actual production process. For the above-mentioned two methods, the steel plate shall be kept at a high temperature without defects on the surface[146, 166]. However, due to the rapid solidification speed of continuous casting technology, longitudinal cracks on the surface of continuous casting slab have great influence on it. This kind of crack will further spread into the rolled products, so the product quality will be greatly reduced[148].

Theoretically, the longitudinal cracks on the surface of continuous casting slab have a high correlation with molten steel temperature, casting powder composition, liquid level fluctuation of crystallizer, copper plate of crystallizer and other factors. The longitudinal cracks on the surface of continuous casting slab generally occur at the middle of wide side of the slab, parallel to the casting direction, and sometimes occur at the corners of the slab. This kind of defect increases the cleanup amount on the slab surface, reduces the comprehensive yield of continuous casting, and even causes the whole slab scrap or breakout, bringing significant negative impact on the stable and smooth operation of the continuous casting production. Moreover, the longitudinal cracks on the surface of continuous casting slab may be inherited into the rolled products, which will seriously affect the quality of rolled products. The longitudinal cracks on the surface of the continuous casting slab are shown in Figure 4.1. Because the longitudinal cracks are comparatively long, they are shown horizontally for a better layout.

Figure 4.1: The longitudinal crack on the surface of the continuous casting slab.

At present, the detection methods of longitudinal cracks on the surface of continuous casting slab mainly include manual detection, physical detection and computer-aided control system detection. The manual detection method refers to that the on-site technicians inspect and judge the surface of the continuous casting slab through direct observation with naked eyes. The manual detection method requires experienced technicians, and the efficiency of manual detection is low. Visual inspection is prone to fatigue and error, and different technicians have different criteria for judging whether there are cracks on the surface. Physical detection methods include optical detection, thermal induction heating, electromagnetic ultrasonic testing and so on[168]. The detection method based on physical means needs complex detection technology, expensive equipment, has high requirements for equipment precision, and is easily interfered by production site environment and human factors, often leading to low detection accuracy.

The computer-aided control system detection method is a popular detection method in the industry at present, for example, the slab quality diagnosis expert system (QES) developed by CCTEC[95, 169] and other quality management systems. These methods are mainly based on expert system to track and monitor the single factors affecting the quality of continuous casting slab, and obtain the final prediction result of slab according to the defect probability of abnormal event set. Because longitudinal cracks on the surface of continuous casting slab are caused by many factors, it is not accurate to predict whether longitudinal cracks occur on the slab surface by monitoring the abnormality of single factor. Therefore, a new method with high accuracy, strong robustness, low cost and self-learning capability is urgently needed for the detection of longitudinal cracks on the surface of continuous casting slab.

Artificial intelligence (AI) technology has attracted more and more attention recently. Its successful application in different fields proves its ability to perform better and more effective tasks (such as data analysis and prediction). This inspires investigators and practitioners to regard AI technology as a valuable asset to solve the above problems in the methods currently used in the detection of longitudinal surface cracks of continuous casting slab.

Internet of things (IoT) has been popular among many areas, which connects a huge amount of physical devices over the Internet. It enables the development of advanced applications for many scenarios, such as home, medical care, automation, industry even oceania[170, 171, 172]. IoT also plays a crucial role in the service life cycle to better incorporate data and provide valuable insights and analysis to stakeholders[173].

Aiming at providing efficient and intelligent computing services, Industry 4.0, which was originated in Germany, has been the focus of industry to achieve a higher level of operation, automation and productivity[174, 175]. It will solicit the techniques from the Internet of Things (IoT), the information and communications technology and enterprise architecture[176]. Combining the industrial information to provide a knowledge intensive service is also termed as modern service industry, in which a middleware was presented to support the development[177]. However, how to present the industrial IoT system, particularly to deliver an intelligent steel manufacturing system, has been missing due to the complicated scenario of the underlying industry and the accessibility of the data infrastructures.

Therefore, this chapter aims to answer these two questions by presenting a five-layer service-oriented intelligent steel manufacturing system based on industrial IoT as the framework, and delivering an effective computational model for the prediction of the longitudinal surface crack based on multiple factors in the analysis of production process as the practice. Specifically, the following contributions can be summarised as relative to the recently published literature:

- We have summarised the challenges faced by the intelligent steel manufacturing industry based on industrial IoT, particularly for the predictive task;
- We have proposed a five-layer service-oriented intelligent steel manufacturing system based on industrial IoT, which include sensor layer, networklayer, database layer, model layer and application layer. This system is deeply based on our on-filed practice and has been deployed for improving productivity;
- We have conducted a systematic analysis of the longitudinal surface crack task, in which we have analysed the multiple factors causing the longitudinal surface crack. Particularly, the proposed five-layer system is employed for the surface crack task, in which a novel neural networkmodel is proposed and the performance is validated on the longitudinal surface crack prediction task. The model has been tested in the collected real data from different sources and the performance has demonstrated its superior capability and efficiency.

4.2 An Industrial Services Architecture for Intelligent Steel Manufacturing

In this section, we present the intelligent steel manufacturing system based on industrial IoT. Overall, the intelligent steel manufacturing system is a complex service-oriented system, which is composed of multiple layers. In accordance with the practice in steel manufacturing industry, we have designed a five-layer system architecture. This is so far the first comprehensive steel manufacturings industrial system architecture, which has been developed to tackle the latent challenges faced in enormous research and industry efforts, and it has well addressed the relationship of different techniques including IoT, information and communications technology and enterprise architecture.

By far, very few steel manufacturing industry systems with a wide and deep scope have been developed to take both the IoT data and the computing infrastructures deployment into account. Given several applications of computing methodologies and models for steel manufacturing in different scenarios, our proposed architecture articulates the requirements of future steel industrys information system by distilling the capability of IoT and artificial intelligence with the strong background of domain knowledge.

In our proposal, the five-layer system architecture consists of two major domains, which are IoT data acquisition and computing model construction. Figure 4.2 illustrates the details of the architecture.

Figure 4.2: The Architecture of Intelligent Steel Manufacturing Industry System.

4.2.1 Sensor Layer

Sensor layer in the smart steel manufacturing industry system plays a central role for IoT data acquisition. With the IoT sensors feeding information, the steel manufacturing system can be tracked, and the model setup data can also be investigated and properly set. Given the specialisation of steel industry, the information will focus on an entire life cycle of corresponding steel. In the sensor layer, various sensors, including those for temperature, rolling force, velocity, the roller gap, chemical composition and so on, will provide a great amount of valuable data for the steel manufacturing system to provide track records and facilitate the process control and management[95].

Although these sensors data and information are considered to provide the unique sources for identification of each continuous casting slab, it is generally considered to be difficult to maintain a database with regard to the extreme production environment[168]. This challenge has resulted in a high cost and demanded highly specialised technology, which will in turn require further research efforts and investments in the smart steel manufacturing industry.

4.2.2 Network Layer

To support the data acquisition from sensor layer, the networklayer of the intelligent steel manufacturing system is designed to establish the communication between sensors and facilitate the data transmission. Particularly, the networklayer will be designed with an industry level standard in terms of robustness and efficiency. Different communication methods may be considered, such as the industrial Ethernet and Internet [178], object linking and embedding for Process Control (OPC) protocol, MODBUS, IoT and so on[168, 169].

The debate of deploying communication protocols meeting industrial standard has been continued for nearly 30 years. While traditional Ethernet and field bus systems cannot fully support the data communication between nodes in a realtime manner, real-time Industrial Ethernet has been proposed[95]. A cycle time ranging from hours to 0.001s has been captured for an industrial control system for company, plant, area, machine and field work correspondingly. OPC is one of the dedicated protocol for real-time requirements based on distributed component object model (DCOM) interface. Supported by the different communication methods as for the networklayer in the intelligent steel manufacturing information system, the data acquisition stage will be completed within the database layer.

4.2.3 Database Layer

The database layer is designed to host the data collected from the production field. There have been a continuing development of complicated process management systems, such as the Enterprise Resource Planning (ERP) and Manufacturing Execution System (MES) systems[179, 180]. Generally, all the data generated during the steel production process will be collected among the projected process management system-

s. The different process management systems will work collaboratively to ensure the effectiveness of the communication and to keep the operators being aware of the overall status of the system.

Meanwhile, the implementation of a process management system, such as MES, poses various technical and organisational issues to be tackled for the intelligent steel manufacturing system[181, 182]. In our work, the systems of MES, Lab Execution System (LES) and SCADA are deployed as the database layer for the collection of data.

4.2.4 Model Layer

In the upper layers of the proposed system, the model layer is firstly designed to organise the data from the database layer in an informative manner. With the model layer, the large amount of heterogeneous data from different systems will be processed depending on their interior and external relationships. For example, the MES system collectively host the data related to the production process while the LES system focus on the inspection and testing data of the steel.

Thus, this layer takes stock of the data for an in-depth analysis with domain knowledge. It is anticipated to generate a well-designed dataset for the neural network model training based on the data characteristics. Once the dataset is curated, the neural network model will be employed to learn the latent relationship from the dataset.

4.2.5 Application Layer

The application layer are concluded as the deployment of obtained neural network model as the top upper layers of the intelligent steel manufacturing system. It will use the incoming new data processed by model layer for prediction to provide fast and accurate services, such as an identification of potential cracks of the longitudinal surface. It will be connected with one of the process management system to finalise the data life cycle, which starts from data acquisition, to storage and model training

4.3 Analysis of Influencing Factors on Longitudinal Cracks of Continuous Casting Slab

The quality of continuous casting slab is the key factor to determine the quality of finished steel products, and the longitudinal cracks of continuous casting slab can always be kept in the rolling process. There are many factors which may affect the occurrence of cracks on the continuous casting slab, but in terms of essential factors, the most fundamental factors affecting the quality of continuous casting slab are continuous casting process, solidification principle of molten steel and inherent parameters of continuous casting machine. At present, during the actual production, the common surface defects of continuous casting slab mainly include the following types: pit, cut, scratch, crush, inclusion, peeling, burr, foreign matter pressing in on the upper surface, foreign matter pressing in and longitudinal crack.

Longitudinal cracks in all slab defects will not only directly affect the quality of continuous casting slab, but also affect the quality of rolled products. From the metallurgical process principle, the longitudinal cracks on the surface of continuous casting slab are mainly formed in the crystallizer, and then spread into longitudinal cracks on the surface after cooling. The chemical composition in peritectic steel can affect the high-temperature mechanical properties of steel, which indirectly affects the crack of peritectic steel by affecting the high-temperature mechanical properties of steel. In other words, any chemical composition that can change the high temperature mechanical properties of steel can cause longitudinal cracks on the surface of continuous casting s-lab, and all chemical compositions that can improve the mechanical properties can also reduce the probability of longitudinal cracks on the surface of continuous casting slab.

Using the process data collected and preprocessed by the industrial IoT system described in the previous section, this section first analyzes the chemical elements and process parameters affecting the formation of longitudinal cracks on the slab surface. The results of data analysis appeared hereinafter are the results of statistics through a large number of data. The equation for calculating the probability of longitudinal cracks on the slab surface under the condition that one element is a specific value is as follows:

$$p(z|e = m) = \frac{n(z|e = m)}{n(e = m)} \tag{Equation 4.1}$$

where, $n(z|e = m)$ indicates the total number of samples labeled as longitudinal crack defects when one influencing factor is a certain value in the data samples; $n(e = m)$ indicates the total number of samples when one influencing factor is a certain value in the data samples.

4.3.1 Influence of Chemical Elements

During steelmaking, chemical elements have a very important influence on the properties of steel. Some chemical elements are beneficial, having a positive influence on the properties of steel, while others are harmful, having a negative influence on the properties of steel. This chapter mainly studies the chemical elements that produce longitudinal cracks on the slab surface, that is, only the elements that have negative influence on the continuous casting process are studied without considering the elements that have positive influence on the continuous casting process. This chapter discusses how the contents of carbon, copper, phosphorus and sulfur influence the formation of longitudinal cracks on the slab surface.

4.3.1.1 Influence of Carbon (C) Content

During continuous casting, when molten steel solidifies in peritectic zone, it will change phase from body-centered cubic to face-centered cubic, so the volume will shrink. The volume shrinkage caused by phase transformation promotes the air gap and separation between the primary shell of continuous casting slab and the copper wall of crystallizer. The uneven distribution of air gap leads to uneven heat transfer, and the heat transfer between solidified shell and copper wall decreases, resulting in uneven temperature gradient on the surface of the shell, uneven thickness of the shell as well as depression. The more serious the surface depression is, the more sensitive the shell is to cracks. Under the action of static pressure, thermal stress and friction force of molten steel, cracks are caused by pressure concentration in depressed parts. After the crack is formed and cooled in the secondary cooling region, the crack will expand, and then obvious longitudinal crack defects will be formed. Figure 4.3 is the relationship between the C content in molten steel and the probability of longitudinal cracks during the actual production:

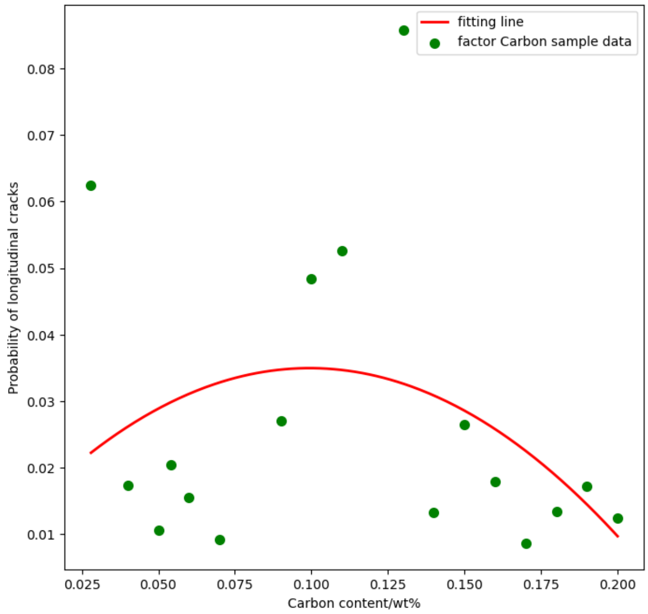


Figure 4.3: Relationship between the carbon content and the probability of longitudinal cracks.

It can be concluded from Figure 4.3 that for the continuous casting slab produced by the iron and steel manufacturer, those with high longitudinal crack index are concentrated in the range with carbon content of $0.05\% - 0.15\%$. When the carbon content is 0.1%, the probability of longitudinal cracks on the surface of continuous casting slab is the highest. When there is no special requirement for carbon content in steel products, the carbon content may be adjusted downward or upward in order to ensure the quality of the products produced.

4.3.1.2 Influence of Copper Content

Cu is another important factor causing longitudinal cracks on the continuous casting slabs. The Cu element in steel is particularly easy to accumulate at grain boundaries and crystal defects. If the alloy of Cu and low melting point elements accumulated at grain boundaries exists in the form of liquid film, the grain boundary strength will be reduced, and cracks will occur under the action of stress. Figure 4.4 is the relationship between the Cu content in molten steel and the probability of longitudinal cracks during the actual production:

It can be concluded from Figure 4.4 that in the continuous casting slab data of the

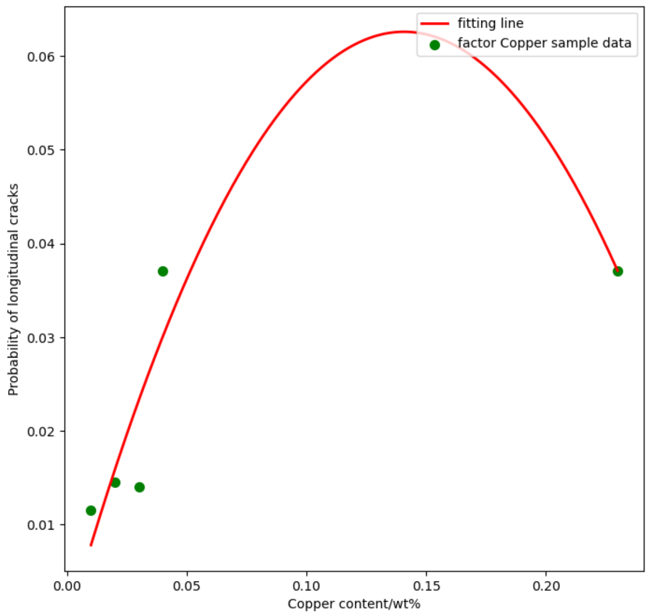


Figure 4.4: Relationship between the copper content and the probability of longitudinal cracks.

iron and steel manufacturer, the copper content is between $0.01\% - 0.05\%$, and the longitudinal crack index increases with the increase of copper content. In order to avoid longitudinal cracks on the surface of continuous casting slab caused by copper content, the copper content shall be controlled below 0.05%.

4.3.1.3 Influence of Phosphorus Content

Generally, phosphorus is a harmful element in steel. Under normal circumstances, the phosphorus in steel can be totally dissolved in ferrite which has a strong solution strengthening effect, increasing the strength and hardness of the steel, but significantly reducing the plasticity and toughness. In the crystallization process of phosphorus, it is easy to produce intracrystalline segregation, which leads to high phosphorus content in local parts, resulting in cold brittleness. In addition, the segregation of phosphorus also causes the steel to form banded structure after hot rolling, so the phosphorus content in steel shall be strictly controlled. Figure 4.5 is the relationship between the phosphorus content in molten steel and the probability of longitudinal cracks during the actual production:

It can be seen from Figure 4.5 that with the increase of phosphorus content, the probability of longitudinal cracks on the surface of continuous casting slab decreases first and then increases, and the decrease and increase range are very large. When the

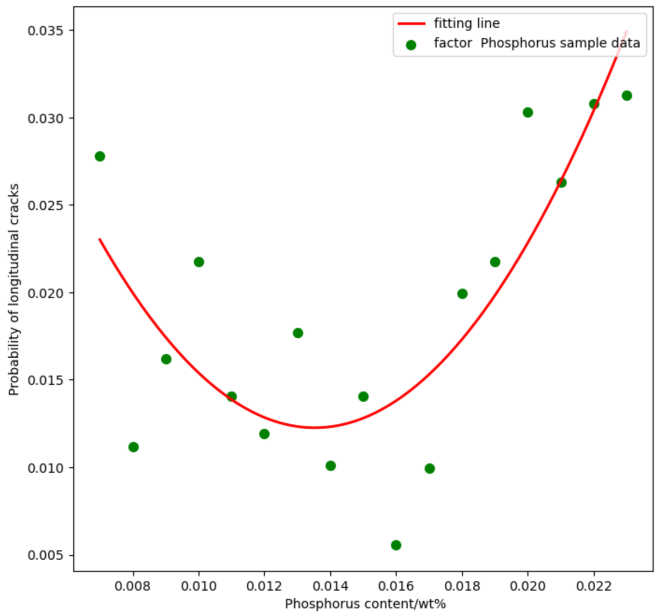


Figure 4.5: Relationship between the phosphorus content and the probability of longitudinal cracks.

phosphorus content is about 0.0135%, the probability of longitudinal cracks on the slab surface is the lowest. In order to reduce the probability of longitudinal cracks on the continuous casting slab during the production, the phosphorus content in molten steel shall be about 0.0135%.

4.3.1.4 Influence of Sulfur Content

Sulfur is a common harmful element in steel, and the sulfur content in steel has obvious influence on the quality of continuous casting slab. The solubility of sulfur in iron in solid state is very small, but it exists in steel in the form of FeS. Because of the poor plasticity of FeS, the steel containing more sulfur is brittle. More seriously, FeS and Fe can form eutectic with low melting point, which is distributed on grain boundary of austenite. When the steel is hot pressed, the eutectic on the grain boundary has melted, and the bond between grains is broken, making the steel crack along the grain boundary during the processing. Therefore, the sulfur content in the steel shall be strictly limited. Figure 4.6 is the relationship between the S content in molten steel and the probability of longitudinal cracks during the actual production:

It can be seen from Figure 4.6 that when the sulfur content is 0.0052%, the probability of longitudinal cracks is the lowest; when the sulfur content is greater than 0.0075%, the probability of longitudinal cracks on the surface of continuous casting slab increases

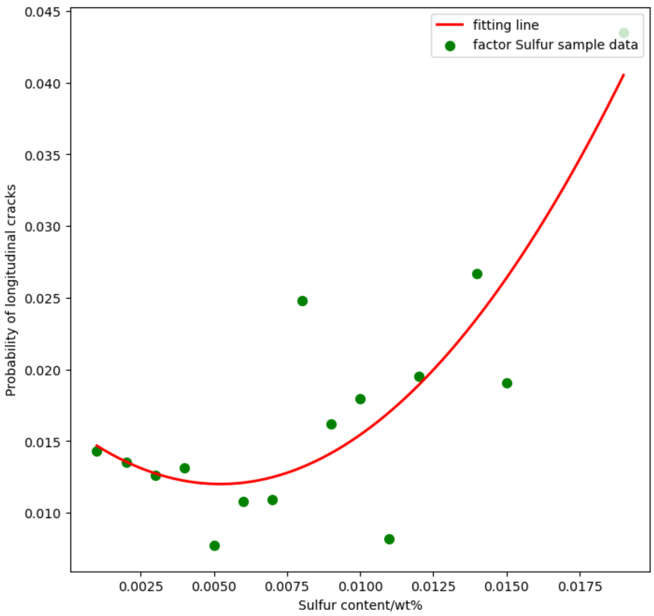


Figure 4.6: Relationship between sulfur content and longitudinal crack probability.

rapidly with the increase of the sulfur content. Segregation of S in steel greatly reduces the zero plasticity temperature of steel, and the reduction of binding ability between S precipitation and machine body forces grain boundary slip and induces stress. Under the action of stress, S and grain boundary form pores and further develop into cracks. Therefore, in order to ensure the quality of continuous casting slab in the actual production, the sulfur content shall be controlled at about 0.0052%.

4.3.2 Influence of Process Parameters

During continuous casting of slab, the main purpose is to produce defect-free slab, and longitudinal cracks will seriously affect the quality of continuous casting slab. Some other defects can be eliminated by finishing and grinding process, and longitudinal cracks are generally not easy to be eliminated, and will still exist on the finished product after rolling. The occurrence of longitudinal cracks on the surface of two groups of slabs is not caused by a single factor, but by the interaction of various influencing factors. Although the chemical composition of molten steel is an important factor affecting the longitudinal crack of continuous casting slab, it is not the only type of influencing factor. Generally, the longitudinal cracks on the slab surface are initially induced in the crystallizer, and expand into longitudinal cracks on the surface after cooling in the secondary cooling zone. Therefore, it can be seen that all unreasonable process factors except chemical composition are possible factors for the formation of longitudinal cracks. This chapter analyzes form the aspect of the factors of casting speed and crystallizer.

4.3.2.1 Influence of Crystallizer

Crystallizer, as one of the most important parts of continuous casting machine, is mainly used to ensure that molten steel is uniformly solidified into slab shell with required specifications and shapes, and continuous casting slab will not produce defects such as longitudinal cracks by adjusting crystallizer parameters. The longitudinal cracks on the slab surface initially occurred in the crystallizer, and spread through the secondary cooling zone to form obvious longitudinal cracks. In this section, the influence of crystallizer on longitudinal cracks on slab surface is analyzed from three aspects: crystallizer taper, heat flux at wide side and heat flux at narrow side.

1). Influence of Crystallizer

During continuous casting, molten steel is injected into the crystallizer and cooled, solidified and contracted to form a primary slab shell, which then will be separated from the crystallizer copper plate, resulting in an air gap with the crystallizer copper plate. In order to make up for the air gap, the crystallizer will adjust the taper as contraction compensation. Because the cooling strength of the corner is greater than that of the middle part, the contraction of the corner is faster than that of the middle part, and the air gap hinders the heat transfer, causing the rise of the corner temperature and the occurrence of longitudinal cracks.

If the taper of the crystallizer is too small, compensation cannot be implemented. When the volume of the slab shell shrinks, the air gap between the crystallizer wall and the casting slab still exists, so that the surface temperature of the slab shell increases and becomes thinner. Under the action of the static pressure of molten steel, the narrow-faced slab shell expands outward and cracks easily occur. If the taper is too large, the wear of the lower opening of the crystallizer will be aggravated, the heat transfer at the narrow side will be faster, and longitudinal cracks are easy to occur under the action of the tensile force at the wide side. Inconsistent taper at two narrow sides of crystallizer leads to uneven heat transfer and different pulling force at the wide side, causing cracks at the weak part at the wide side.

2). Influence of Crystallizer Heat Flux

During continuous casting production, the heat flux distribution is asymmetric, that is, the thermal behavior of the slab is uneven, which will affect the normal continuous casting production. The solidification control of molten steel in crystallizer is the key factor among many factors affecting the surface quality of slab, and its main purpose is to obtain dense solidification structure without surface defects. Generally, the uniform heat flux is formed by adjusting the cooling water quantity at the wide and narrow sides of the crystallizer, so that the uniform solidification elongation and stress effect of the slab shell are minimized. Therefore, the heat flux at the wide and narrow sides has great influence on the longitudinal cracks on the slab surface. Figure 4.7 shows the relationship between wide side heat flux and longitudinal crack probability during actual production, and Figure 4.8 shows the relationship between heat flux at narrow side and longitudinal crack probability during actual production:

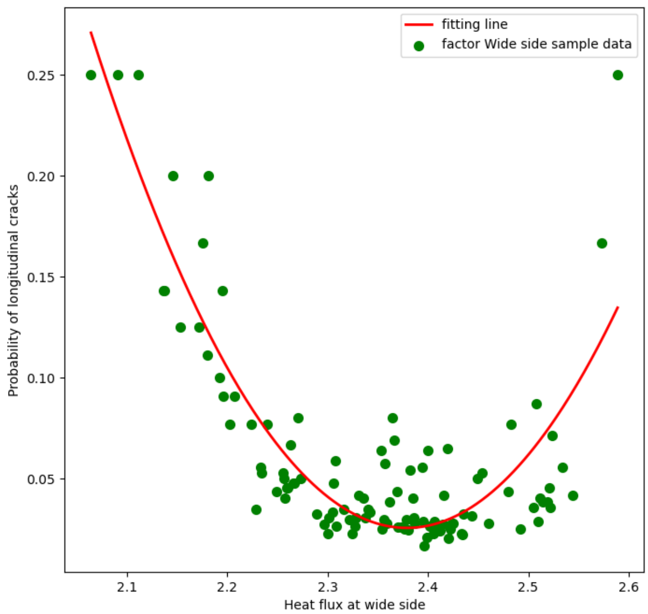


Figure 4.7: Relationship between heat flux at wide side and longitudinal crack probability.

It can be seen from Figure 4.7 that with the increase of heat flux at wide side, the probability of longitudinal cracks on the surface of continuous casting slab first decreases and then increases. When the heat flux at wide side is about 2.375, the probability of longitudinal cracks on the slab surface is the lowest. When the heat flux at wide side is between $2.3-2.45$, the probability of longitudinal cracks on the slab surface is in a relatively low and stable phase. Therefore, in order to reduce the probability of longitudinal cracks during the production, the heat flux at wide side in the continuous casting process is kept between $2.3 - 2.45$.

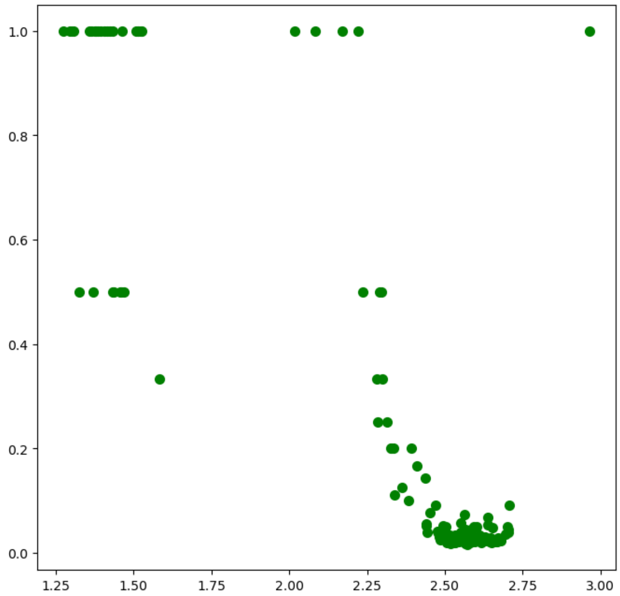


Figure 4.8: Relationship between heat flux at narrow side and longitudinal crack probability.

It can be seen from Figure 4.8 that when the heat flux at narrow side is less than 2.25, the probability of longitudinal cracks on the surface of continuous casting slab is very high; while it is between $2.4 - 2.6$, the probability of longitudinal cracks on the surface of slab is very low. In order to reduce the probability of longitudinal cracks in the production, the heat flux at narrow side in continuous casting process shall be kept between $2.4 - 2.6$.

4.3.2.2 Influence of Casting Speed

Among many continuous casting technologies, the casting speed of slab is very important, and it is important to keep the casting speed consistent in multi-furnace continuous casting. However, when special circumstances such as changing nozzle and blocking nozzle occur, the casting speed of slab has to be reduced or even stop casting. Longitudinal cracks may occur on slab surface with the increase of casting speed, because the solidification of primary shell of continuous casting slab will be delayed and the temperature of casting slab surface will rise, resulting in longitudinal cracks under stress. The relationship between casting speed and longitudinal crack probability in actual production is shown in Figure 4.9:

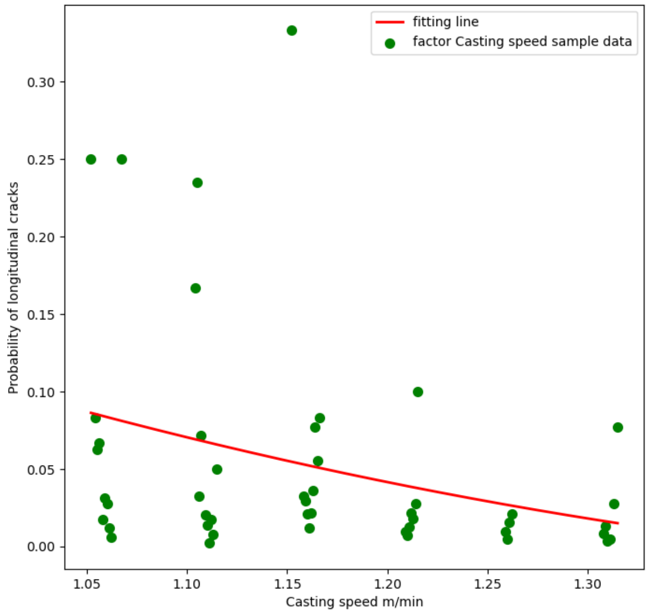


Figure 4.9: Relationship between casting speed and longitudinal crack probability.

It can be seen from Figure 4.9 that the probability of longitudinal cracks on the surface of continuous casting slab decreases with the increase of casting speed. According to the actual production data, the casting speed does not exceed $1.4m/min$, and the probability of longitudinal cracks on the slab surface may be reduced if the casting speed is kept slightly higher than $1.30m/min$ combined with the information obtained in the plot.

This section analyzes how chemical elements and process factors affect the longitudinal cracks on the surface of continuous casting slab. It can be seen from the data that the relationship between chemical elements and process parameters and the probability of longitudinal cracks on the surface of continuous casting slab is nonlinear, and as the steel sector belongs to the process industry, the numerical control of the former process will have an impact on the control of the latter process. Therefore, it is very difficult to control the probability of longitudinal cracks on the surface of continuous casting slab by controlling the value range of each process parameter independently. Next, the artificial neural network will be introduced to fit the relationship between each process (chemical elements and production process) in the production of continuous casting slab and whether longitudinal cracks on the surface of continuous casting slab occur.

4.4 Computational Model

In section 4.2, a comprehensive framework for intelligent steel manufacturing system is introduced, among which the neural network model has been included in the model layer. The computational model is critical to learn from the collect IoT sensor data. In this section, we will include the neural network based on back propagation algorithm as the evaluated model given its effectiveness and popularity. Particularly, we have proposed a novel back propagation algorithm based neural network model to enhance the performance of the computational model.

4.4.1 The Back propagation Neural Network

BP neural network (Backpropagation Neural Network) is a multi-layer feedforward neural network, whose main features are input signal forward propagation and error backward propagation. During forward propagation, the input signal is processed from the input layer through the hidden layer to the output layer, and the neuron state of each layer only affects the neuron state of the next layer. If there is an error between the predicted value and the real value in the output layer, the process of error back propagation is turned into, and the weights and thresholds of neural network are adjusted according to the prediction error, so that the predicted output of BP neural network keeps approaching the real result. The detailed calculation methods of the forward propagation of BP neural network input signal and the backward propagation of error have been introduced in Chapter 2.

Although BP neural network is one of the most mature neural network algorithms, it has good self-learning, self-adaptation, robustness and generalization ability, and the three-layer BP neural network can approximate any nonlinear function with arbitrary accuracy [183]. However, the basic BP neural network needs to be improved to a great extent. For example, the structural parameters of the network (for example, the number of nodes in the hidden layer) are usually set by domain knowledge, and there are no rules to assist the decision-making process of parameter adjustment [184]. Therefore, we put forward a new method, which is based on BP neural network with pre-optimised number of nodes in hidden layer to predict the possibility of longitudinal surface cracks, so as to realize intelligent manufacturing.

4.4.2 The Proposed Pre-optimised BP Neural Network

In this section, instead of using the traditional BP neural network model as the learning model[151], we have proposed a pre-optimised BP neural network based on the IoT sensor data, which has aimed to optimise the number of nodes of the hidden layer in the basic BP network to avoid the negative influence of the model. During the optimisation of the number of nodes of the hidden layer, the training set was divided into 50 different categories and the weights and thresholds are randomly initialised to reduce random error. The training dataset is a random pro-rata segmentation of a large data set and the test set is a fixed data set. Then, F1-score is calculated based on the average performance of 50 different models on the validation set. F1-score is one harmonic measurement of the precision and recall values, which indicates the overall performance of the training models.

Since there is one hidden layer in the BP network, the number of the nodes of the hidden significantly influence the model performance. The number of the nodes in the hidden must be less than $N-1$, or it will lose its generalisation ability, where N is the number of training samples. Base on the previous experiences, the number of the nodes in the hidden can be determined by the number of nodes in the input layer and the output layer, which can be formulated as follow:

$$\begin{aligned} m &= \sqrt{n+l} + \alpha \\ m &= \sqrt{nl} \\ m &= log_2 n \end{aligned} \qquad \text{(Equation 4.2)}$$

Here, α is an integer between 1 and 10, m is the number of the nodes in the hidden layer, n is the number of the nodes of the input layer, and l is the number of the nodes of the output layer. According to Equation 4.2, we will obtain three different m values. The biggest one will be chosen as the upper bound. In this research, the upper bound value is determined as 19 in the experiment. We iterated the m value from 1 to 19 to construct the model, respectively. Also, these models are trained based on the 50 different datasets aforementioned. The average results collected on the validation dataset F1 will be subsequently utilised as the performance measurements of each hidden layer. The parameter setting of the best model is chosen as the final model setting. The choosing procedure of the number of the nodes in the hidden layer is represented as Figure 4.10.

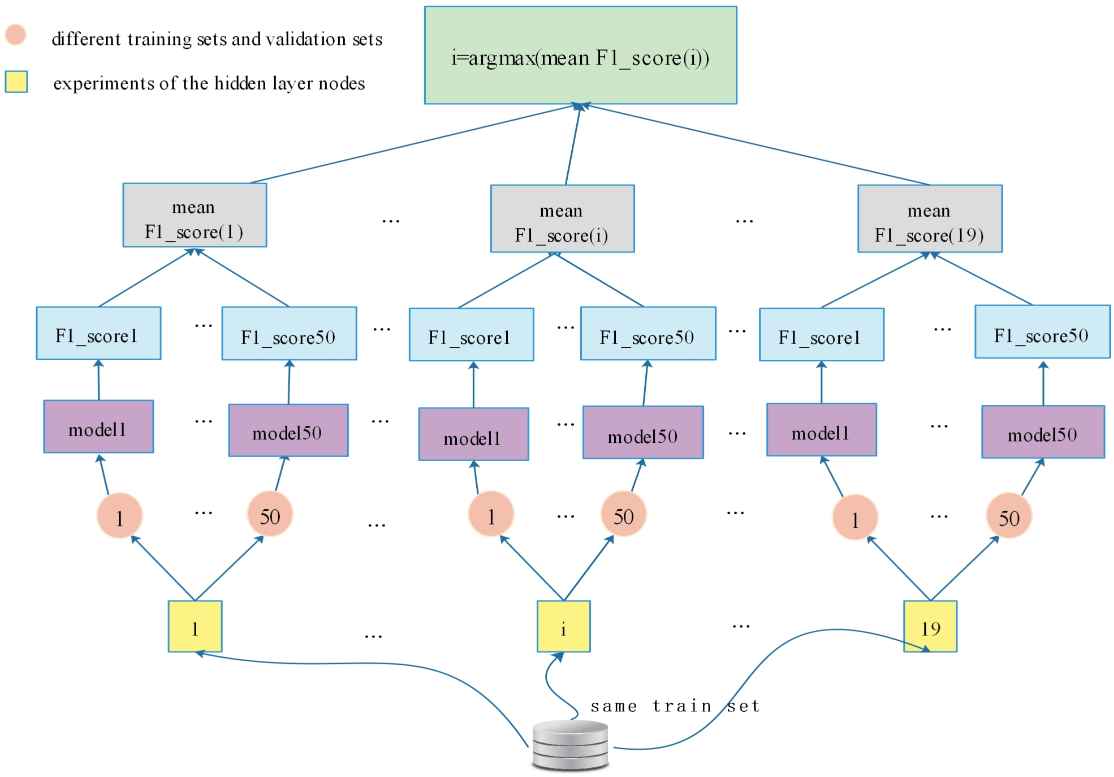


Figure 4.10: The Optimization Process of the Hidden Layer in BP Neural Network.

4.5 Experimental Verification of the Proposed Model

The IoT sensor data source, evaluation criteria, and the analysis of the experiment results are discussed in this section. More specifically, the collection of data, feature selection, and data cleaning will be discussed in the section of data source. The recall, precision, F1-score and AUC used in this study will be introduced in the section of evaluation criteria. The optimization of the hidden layer in the BP network, and the selection strategy of the number of hidden nodes will be discussed and analysed in the section about the analysis of experiment results. Moreover, this experiment involved testing pre-optimized BP network and the conventional network with the randomly selected number of hidden nodes. Lastly, the performances of optimized BP network and the network with the randomly selected number of hidden nodes are compared and analysed, the experiment workflow about the pre-optimized BP network is shown in

Figure 4.11.

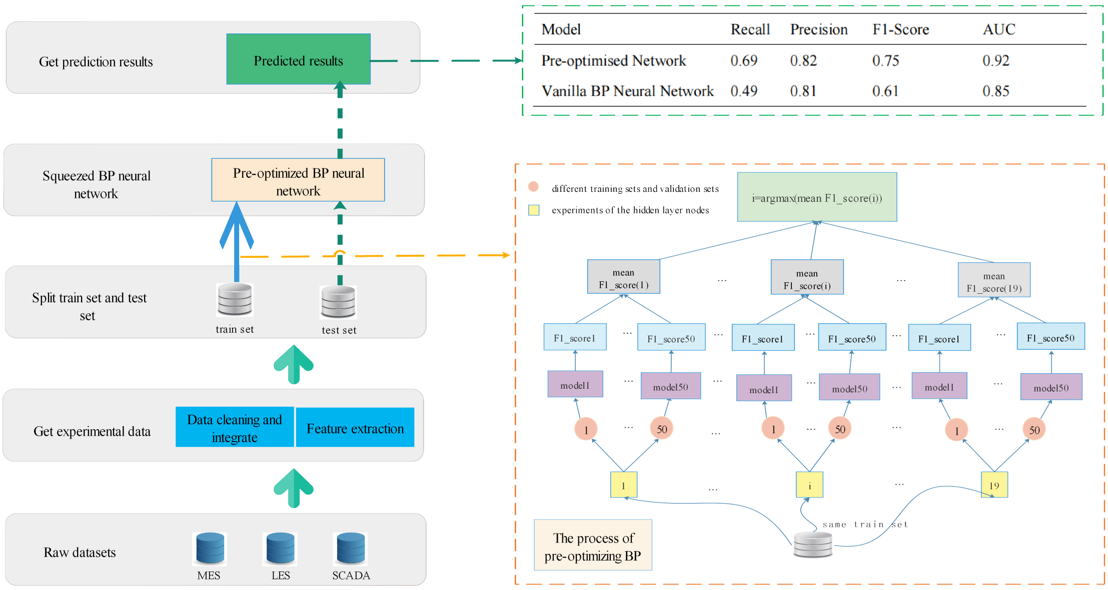


Figure 4.11: The Experimental Process of the Proposed BP Neural Network Model.

4.5.1 The IoT Sensor Data Source

The IoT sensor data is collected from the construction site and stored in different computing infrastructures, such as MES, LES, and supervisory control and data acquisition (SCADA) system of Nanjing Iron and Steel Union Co. Ltd (NISCOs) information system. The information about chemical components of plate blank is collected from MES, the relevant information about the continuous casting production process is collected from SCADA, and the quality information about plate blank (evaluations about longitudinal crack defects) is collected from LES. Based on the collected raw information, previous experiences about continuous casting, main chemical components, production process parameters, the temperature level of pouring procedure, and the sensitive factors of iron, we engineered 71 features. The 71 features include 59 variables in the continuous casting process, such as crystallizer taper, crystallizer vibration amplitude, crystallizer vibration frequency, crystallizer wide and narrow heat fluxes, continuous caster pull speed, etc., 10 chemical elements (e.g., C, P, S, Cu, Mn, Ni, etc.), as well as pouring superheat and Mn/S constructed from metallurgical knowledge. The missing values and outlier are fixed by using average values. There are 105 longitudinal cracked samples and 331 normal samples involved in this experiment. 80% of the samples are selected as the training set,20% of the samples are set as the test set, 80% of the training samples are used to construct the model; and 20% of the training samples are used to validate the model.

4.5.2 Evaluation Criteria

Recall, precision, F1-score, and AUC are selected in this experiment as the evaluation metrics, which can be formulated as following.

1. Recall
$$R = \frac{TP}{TP + FN}. \tag{Equation 4.3}$$

2. Precision
$$P = \frac{TP}{TP + FP}. \tag{Equation 4.4}$$

3. F1-Score:
$$F1 = \frac{2 \cdot P \cdot R}{P + R}. \tag{Equation 4.5}$$

4. AUC:

   Given a positive and negative sample, the AUC value indicates the possibility when the capability of the classifier correctly predicting the positive sample is higher than incorrectly predicting the negative sample as positive. Supposing there are M positive samples and N negative samples, there will be $M * N$ sample pairs (one positive sample together with one negative sample). By calculating the possibility of the $M * N$, the AUC value will be obtained. In summary, the AUC value can be calculated as following:

$$AUC = \frac{\sum I(P_{postive}, P_{negative})}{M \cdot N}. \tag{Equation 4.6}$$

   where the I is calculated by following:

$$\sum I(P_{postive}, P_{negative}) = \begin{cases} 1 & P_{postive} > P_{negative} \\ 0.5 & P_{postive} = P_{negative} \\ 0 & P_{postive} < P_{negative} \end{cases}. \tag{Equation 4.7}$$

TP refers to the true positive samples, FP represents the false positive samples, FN are the false negative samples, and TN are the true negative samples.

4.5.3 Experiment Results and Analysis

During the optimisation of the hidden layer of BP network, based on the Equation 4.2 the upper bound of the number of the nodes in the hidden layer is set to 19. Then, we iterated from 1 to 19 to construct different models. Moreover, to reduce the random error, we construct 50 models for each setting, and use average F1-score to represent the performance of each model under a specific setting. As shown in Figure 4.12, the average value of F1-score increases with the increasing number of nodes. The F1-score reaches the peak when the hidden layer contains 17 nodes. Hence, we choose 17 as the final setting of our BP network.

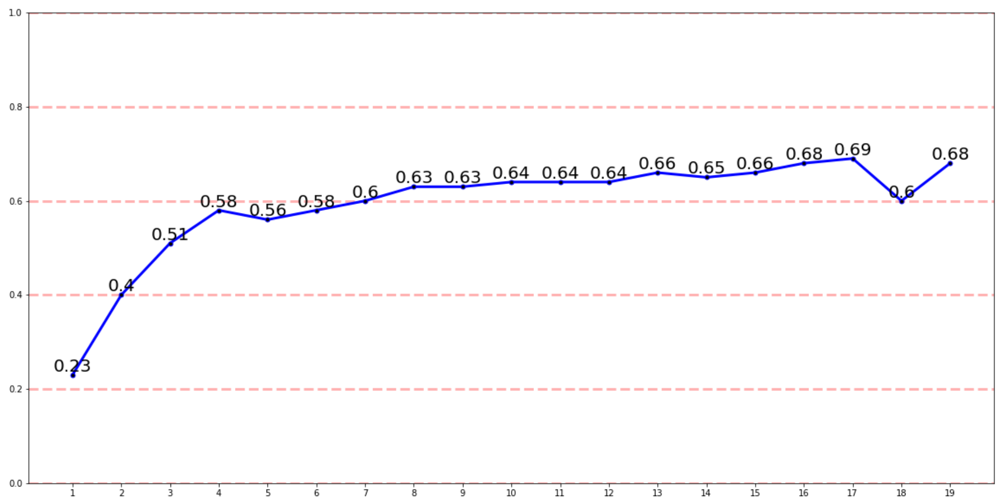


Figure 4.12: The Average F1-Score on Different Hidden Layers.

The performance of the pre-optimised BP network and the BP network with randomised node number is shown in Table 4.1. The performance of the BP networkis calculated by averaging the evaluation results of 1 to 19 models.

Table 4.1: Performance Comparison between Proposed Pre-optimised and Vanilla BP Neural Network.

model	recall	precision	F1-score	auc
Pre-optimised Network	0.69	0.82	0.75	0.92
Vanilla BP Neural Network	0.49	0.81	0.61	0.85

From the Table 4.1, we can see that the recall of pre-optimised BP networkis 0.69, the precision value is 0.82, the F1-score is 0.75, and the AUC value is 0.92; while the network with randomised node number only has 0.49 recall value, 0.81 precision, 0.61 F1-sore, and 0.85 AUC. Based on these evaluation criteria, we can conclude that pre-optimised BP network outperformed that conventional BP with randomised node number in the hidden layer, especially for the F1-score and recall value. The recall of the pre-optimised BP network is 20% higher than the one of conventional BP network; the F1-score of the pre-optimised BP networkis 14% higher; the precision of the pre-optimised BP network is 1% higher; and the AUC of the pre-optimised BP networkis 7% higher. The 0.92 AUC value of the pre-optimised BP network indicates that our model has high accuracy when predicting longitudinal crack of the slab surface. Based on the above results, we can say that our model outperforms the conventional BP network with randomised node number. The proposed model can handle the task of prediction of longitudinal cracks on the slab surface.

The ROC curve, which is defined as a receiver operating characteristic curve, is further examined for our models. The ROC curve defines the ability of a model via a varied discrimination threshold. The ROC curves of the pre-optimised BP networkand the network with randomly chosen node number of the hidden layer are shown in Figure 4.13. The average value of different models is used to construct the ROC curve.

In Figure 4.13, the yellow line is the ROC curve of the preoptimised BP network, the green line is the ROC curve of the BP network with randomly selected hidden node number. By comparing these two lines, we can clearly see that the most portion of the yellow line is above the green one, which indicates the pre-optimised BP network has greater AUC value than the BP network with randomised hidden nodes number. This result is consistent with the results in Table 4.1, which demonstrates that the pre-optimised BP network has higher accuracy in the task of predicting the longitudinal crack of slab surface.

4.6 Summary

In this chapter, firstly, some factors affecting longitudinal cracks on the surface of continuous casting slab are analyzed, and some results that may avoid longitudinal cracks on the surface of casting slab are obtained through analysis. The results are as follows:

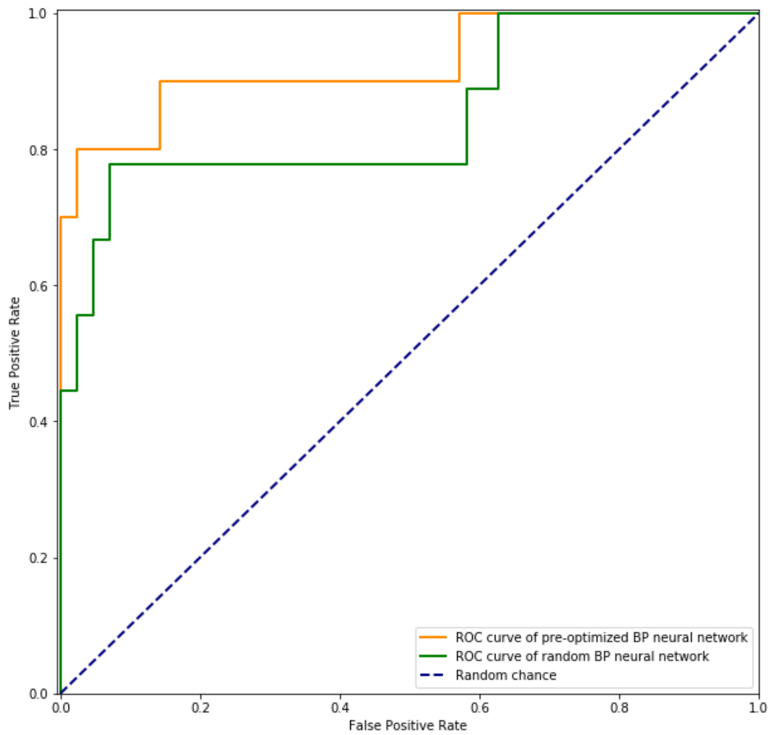


Figure 4.13: The ROC Curves for Different Models.

(1) When the carbon content is in the range of $0.05\% - 0.15\%$, the probability of longitudinal cracks in the continuous casting slab is relatively high, and when the carbon content is 0.1%, the probability of longitudinal cracks on the surface of the continuous casting slab is the highest. When there is no special requirement for carbon content in steel products, in order to ensure the quality of produced products, the carbon content can be adjusted downward or upward.

(2) When the copper content is in the range of $0.01\% - 0.05\%$, the longitudinal crack index increases with the increase of the copper content. In order to avoid longitudinal cracks on the surface of continuous casting slab caused by the copper content, the copper content shall be controlled below 0.05%.

(3) With the increase of phosphorus content, the probability of longitudinal cracks on the surface of continuous casting slab first decreases and then increases, and the range of decrease and increase is very large. When the phosphorus content is about 0.0135%, the probability of longitudinal crack on the surface of the slab is the lowest. In order to reduce the probability of longitudinal cracks in continuous casting slab, the phosphorus content in molten steel shall be about 0.0135%.

(4) When the sulfur content is 0.0052%, the probability of longitudinal cracks is the lowest; when the sulfur content is greater than 0.0075%, the probability of longitudinal cracks on the surface of continuous casting slab increases rapidly with the increase of the sulfur content. Therefore, in order to ensure the quality of continuous casting slab in the actual production process, the sulfur content shall be controlled at about 0.0052%.

(5) With the increase of heat flux at wide side, the probability of longitudinal cracks on the surface of continuous casting slab first decreases and then increases. When the heat flux at wide side is about $2.375MW/m^2$, the probability of longitudinal cracks on the slab surface is the lowest. When the heat flux at wide side is between $2.3MW/m^2$ and $2.45MW/m^2$, the probability of longitudinal cracks on the slab surface is in a plateau period. Therefore, in order to reduce the probability of longitudinal cracks during production, the heat flux at wide side in the continuous casting process is kept between $2.3MW/m^2$ and $2.45MW/m^2$.

(6) When the heat flux at narrow side is less than $2.25MW/m^2$, the heat flux rate at narrow side of longitudinal cracks on the surface of continuous casting slab is very high; while it is between $2.4MW/m^2$ and $2.6MW/m^2$, the probability of longitudinal cracks on the slab surface is very low. In order to reduce the probability of longitudinal cracks in the production, the heat flux at narrow side in continuous casting process shall be kept between $2.4MW/m^2$ and $2.6MW/m^2$.

(7) The probability of longitudinal cracks on the surface of continuous casting slab decreases with the increase of casting speed. According to the actual production data, the casting speed does not exceed $1.4m/min$, and the probability of longitudinal cracks on the slab surface may be reduced if the casting speed is kept slightly higher than $1.30m/min$ combined with the information obtained in the plot.

Through the above analysis, we have a certain understanding of the influence of a single factor on the longitudinal cracks on the slab surface. However, continuous casting is a complicated procedure, and the longitudinal cracks on the slab surface are the result from many factors. Therefore, we use the pre-optimized BP neural network to fit the possible factors with whether the slab surface is cracked or not to solve the pain point.

To achieve this goal, we have systematically analysed the industrial service-oriented architecture in steel manufacturing area. Particularly, our filed experience and IoT based infrastructure have greatly improved our system design. Moreover, an five-layer intelligent steel manufacturing system has been demonstrated in Section 4.2. Taking the system into practice, we have also investigated the task of detecting longitudinal surface crack. Numerous challenges from the five layer system have been discussed and tackled for the task of detecting crack on continuous casting slabs. The goal is to well harness the IoT data and computational intelligence techniques to address the problems occurred by traditional approaches, which mostly cost a huge amount of time and human resources. Meanwhile, the detecting machines are costly and sensitive to the environments causing a low accuracy of the results.

Aiming at these disadvantages, our designed system has well served the intelligent steel manufacturing environment by providing efficient and effective solutions. We have also integrated the five-layer system with a novel neural network model based on the pre-optimised backpropagation algorithm for the industrial prediction task. The multi-sources IoT data of continuous casting slabs from Nanjing Steel Corporation are incorporated in the model. Lastly, we have evaluated the model on the collected IoT sensor data and the performance has demonstrated its superior capability and efficiency.

Although we have achieved satifactory results in the prediction of longitudinal cracks on the surface of the steel plate, it is feasible to further optimize in terms of the diversity of influencing factors, the number of data and the algorithm. In terms of the diversity of influencing factors, we intend to deploy some sensors in the production environment to collect more data on the variables affecting the surface longitudinal crack during production. In the following research, we will investigate more layers of neural network, integrated machine learning algorithm, fuzzy algorithm and other algorithms, as well as the combination of multiple algorithms, to seek for a algorithm with higher precision. At the same time, some algorithms used to optimize the parameters of machine learning and neural networks will also be included in the next stage.

Chapter 5

Ensemble Machine Learning Systems for the Estimation of Steel Quality Control

Following the introduction of production process of steel products in the Chapter 1, two applications of intelligent quality prediction in steelmaking and continuous casting are introduced in the Chapters 3 and 4 respectively. And the intelligent prediction application of mechanical properties of rolled steel finished products will be introduced in this chapter. The successful application of this method will replace the sampling-experiment detection method, reduce the cost and time required by experiment, and quickly deliver qualified products to customers.

In this chapter, the main contents are as follows: Section 5.1 mainly expounds the present situation and challenges of steel plate mechanical properties, Section 5.2 introduces the ensemble machine learning method for steel plate quality control proposed in this chapter, Section 5.3 introduces the data set used in this chapter in detail, Section 5.4 evaluates the performance of the model established in this chapter, and Section 5.5 summarizes the contents of this chapter and looks forward to future work.

5.1 Problem Description

According to the basic knowledge of iron and steel smelting introduced in the Chapter 1, iron and steel smelting is a process-oriented industry, so each production line is of continuous production during the production. In addition, the production scale of each production line is relatively large, which means in case quality problems occur, the whole batch of products will also suffer similar quality problems, in which case, it will lead to serious economic losses. Accurately predicting the physical properties of steel has become an important research issue for iron and steel industry. So there must be a method to test the quality of products to avoid more economic losses.

For traditional methods, laboratory equipment is mainly used to verify quality data. It means that products are sampled during the production, and the samples are sent to the laboratory and processed by laboratory instruments for analysis. For example, a tensile machine is used to determine the tensile strength of finished products. After obtaining the test result, if it meets the user's requirements, the production will be carried out according to the plan. If not, the product will need to be reprocessed in the subsequent procedure. As a result, these conventional test methods are not only costly, but also time consuming, which will affect the efficiency and benefit of factory production. Another test method is to deploy SPC (Statistical Process Control) system, which has been widely used in iron and steel industry. However, SPC system can only give parameter warnings during the production, but cannot predict the actual values related to product quality.

In view of the limitations of the above issues, this chapter puts forward the method of regression analysis in machine learning to predict the physical properties of steel products. Regression analysis is used to estimate the relationship between the input variables (dependent variables) and result variables (dependent variables) in the combination; its function is to understand how the typical value (actual value) of the dependent variable changes when the independent variable (predicted value) changes[185]. In this chapter, the prediction of physical properties during steel production is mainly explored via using continuous variables (information such as temperature and pressure during rolling). The forecasting system is implemented by defining the task as a traditional regression problem, which involves applying one or more continuous input(s) to predict the required output[186].

5.2 Ensemble Machine Learning for Steel Quality Control

The ensemble algorithm as well as the associated data flow has been implemented in a comprehensive system of Nanjing Iron and Steel Company. Figure 5.1 demonstrates the framework of this machine learning-based performance prediction system, which consists of four major layers from bottom to the top: raw datasets, data extraction and preprocessing, data modelling and analysis platform, and steel quality control integrated systems.

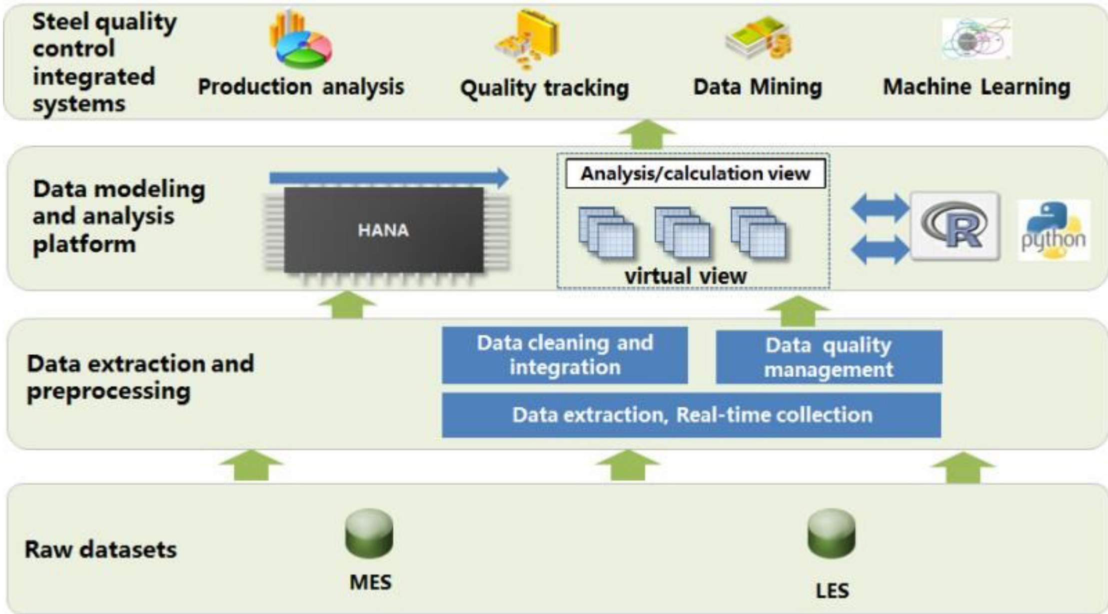


Figure 5.1: System Framework.

Raw datasets contain historical observations, a manufacturing execution system (MES) and Lab Execution System (LES) information.

Data extraction and preprocessing include analysis and processing of the abnormal data, filling of the missing values, and duplicate data removal. Due to the difference between the feature variables, to accelerate the convergence of the model, the processing of the datasets should be scaled their feature variables down to a range between 0 and 1.

Data modelling and analysis platform use Extract Transform Load (ETL) tools and real-time acquisition function to extract the information from the historical data set, which is imported into the system applications and products high-performance analytic appliance (SAP HANA) based data warehouse. The HANA built-in analysis tools, R language, and Python are also included to establish the platform.

Finally, based on the data modelling and analysis platform, production process analysis, quality tracking, data mining, and machine learning are developed as several major components to form the complete steel quality control layer.

5.2.1 Working Principle and Data Flow in System

As the steel quality control system using ensemble learning has been developed and implemented to solve the regression problem, Figure 5.2 shows a flowchart of the working principle of the ensemble learning system. After obtaining the MES and LES data from the steel quality management system, data analysis, data processing, and feature engineering are conducted to select 57 features as an input to the estimation models. As can be seen from Figure 5.3, we also apply a heatmap to rank the top 10 correlation features. A high value demonstrates a strong relationships between features.

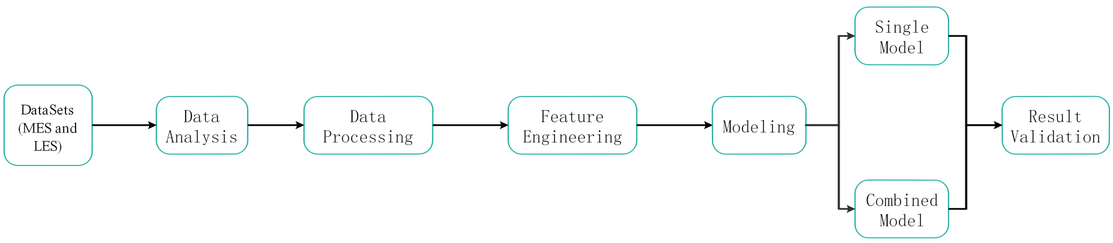


Figure 5.2: Flow Chart of Ensemble Learning System.

Provided that the data extraction and preprocessing have been well conducted, sorted data are then fed into the layer above it. Specifically, the data analysis phase is to understand the relationship between raw data and to explore their correlations and distribution. At the data processing phase, removing all duplicate records and replacing missing data process with valuable data are executed accordingly, which can be then utilized to provide accurate predictions for the steel company. Finally, based on a series of domain knowledge, we select 57 features such as the thickness of steel plates, the rolling temperature, the amount of cooling water, and chemical elements.

Moreover, as the estimation of steel quality control is a supervised regression learning, we start with the core machine learning concepts to generate a set of T base learners $\{h_1, \cdots, h_T\}$ to tackle the problem. In other words, the original training data is divided in a manner of T-fold. Then, the training data is processed for T times and each time it only produces one prediction.

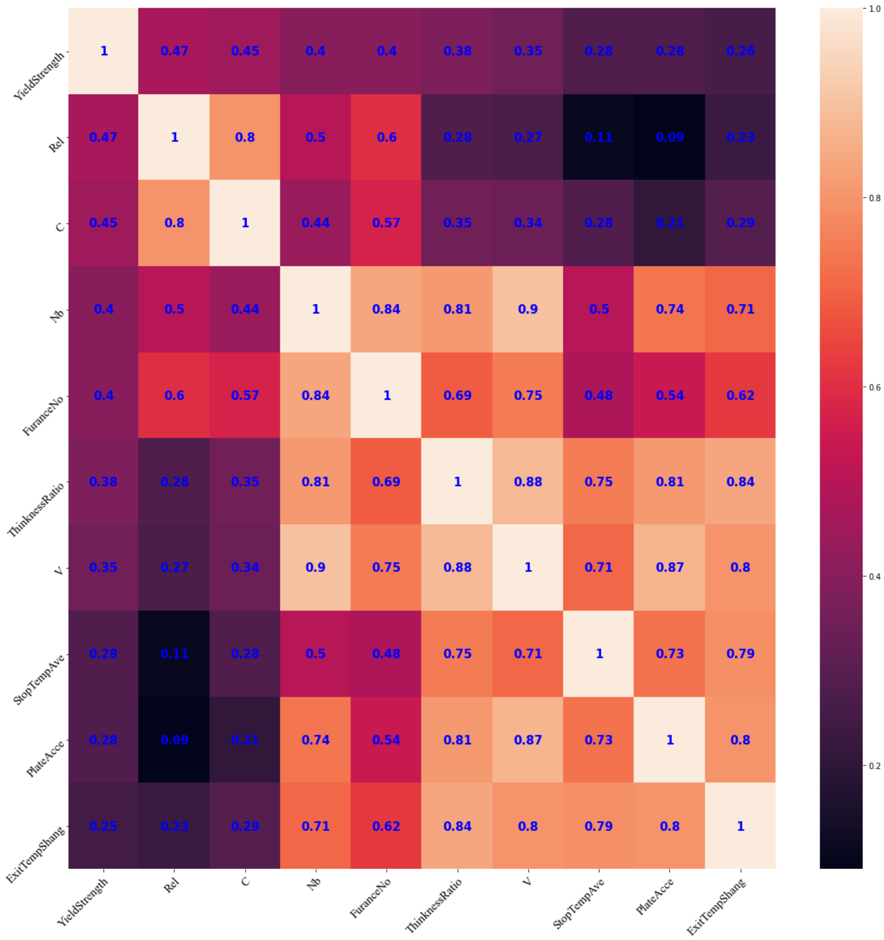


Figure 5.3: Rank Top 10 Features in the Correlation Heatmap.

In the top layer of the system, averaging method and stacking method works together with the initial learners in order to form a combined model, which functions in obtaining the corresponding combined output $H(x)$ for the dependent variable x respectively. Furthermore, for the result validation stage, we measure our proposed model combinations by R-squared(R^2), Root Mean Square Error (RMSE), and Percentage of Error (PE).

5.2.2 Data Availability

Most steel companies have established information systems at the earlier stage, such as MES, LES, Enterprise Resource Planning (ERP), Energy Management System (EM-S), and supervisory control and data acquisition (SCADA). The systems house the data generated during the whole product lifecycle. It includes order data, quality information in product manufacturing processes, price and quality information of the raw materials, process parameter information, as well as the energy consumption information from the production process. The datasets describe the entire life cycle of corresponding products. In this study, the data sources are primarily collected from the existing information systems of the Nanjing Iron and Steel company, which are MES and LES.

5.3 Empirical Evaluation

5.3.1 Datasets

We use two different sets of data as demonstrated in Table. The detailed explanation is as follows.

MES dataset: It is an information management system for the production process execution layer of the manufacturing enterprise. We mainly obtained production process data of steel plates and process data from the MES system of the Nanjing Iron and Steel Company, including rolling performance, cooling performance, continuous casting performance and chemical composition.

LES dataset: LES refers to the inspection and testing system of a manufacturing enterprise. We mainly obtained the performance evaluation data of the steel plate from the LES system, including yield strength, tensile strength, elongation and impact work.

5.3.2 Baselines

In this chapter, 11 basic models will be used for training and learning of data sets, and these basic models will be upgraded into ensemble learning models. The basic principles of 11 basic models and the principle of ensemble learning algorithm, which have been introduced in the machine learning algorithm section of Chapter 2, will not be repeated in this section.

Table 5.1: Datasets.

Dataset	MES	LES
Data source	Enerprise producion process execuion system	Inspection and testing system
Time Span	2014/4-2017/8	2014/4-2017/8
Granular Data	Rolling performance parameter, cooling performance parameter, continuous casting performance parameter, and chemical composition parameter.	Sample serial number
Data detailed information	The rolling performance of steel plates (Obtained by the steel number) includes 17 characteristics such as rolling mode, number of passes, rolling temperature, water volume and reddening temperature.	The inspection and testing data include four indexes of yield strength, tensile strength, elongation and impact energy of the steel plate.
	The cooling performance of steel plates(Obtained by the steel number) includes 13 characteristics such as water temperature, water pressure, water inlet temperature, water volume and reddening temperature.	
	The continuous casting performance and chemical composition (Obtained by slab number)include 25 essential characteristics, such as chemical composition: C, Mn, P, S, Si and other main chemical composites. The continuous casting performance: medium temperature, drawing speed and average liquid level.	

5.4 Model Performance

5.4.1 Evaluation Measurements

To evaluate the performance of our prediction system and furthermore to compare with our baselines, R-Square and root mean square error are included as the main measurements.

In details, we firstly define that y_i is the observed value, $\bar{y}$ is the average of the observed values, $\hat{y}$ is the predicted value, and n is the number of all available ground truths. $\bar{y}$ is derived from Equation 5.1, relying on the availability of the observed value y_i:

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i. \qquad \text{(Equation 5.1)}$$

Several evaluation measurements are defined as below, Equation 5.2, Equation 5.3 and Equation 5.4:

R-Square: the improvement in prediction from the regression model compared to the mean model. This value ranges from 0 to 1, while a closer value to 1 indicates a better prediction results.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_l)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_l)^2}. \qquad \text{(Equation 5.2)}$$

Root Mean Square Error (RMSE): the standard deviation of the residuals, measures the differences between the values which predicted by a model and the values observed. The lower the RMSE value, the better prediction results we get.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \qquad \text{(Equation 5.3)}$$

Another metrics, percentage of error (PE), is defined in EquationEquation 5-4 to illustrate the percentage of error comparing with the average of the observed values. As shown in Equation 5.4, it stays synchronous with the RMSE level.

$$RMSE = \frac{RMSE}{\bar{y}}. \qquad \text{(Equation 5.4)}$$

The experiment results in the execution of our 11 baselines are shown in Table 5.2. Without any doubts, we can find that algorithms involve boosting strategy (GBDT, L-GBM, and XGBoost) perform better than conventional single regression models (Linear Regression, Ridge Regression, and Lasso Regression). This result further indicates that the ensemble model can have the potential to outperform a single regression model.

Table 5.2: Comparisons with baselines on MES and LES.

Model	R^2	RMSE	Percentage of error (%)
Linear Regression	0.431	17.774	3.22
Ridge Regression	0.443	17.576	3.18
Lasso Regression	0.444	17.580	3.18
Elastic Net	0.443	17.588	3.18
SVM(RBF)	0.522	16.288	2.95
KRR(RBF)	0.539	15.993	2.90
KNN(distance)	0.530	16.158	2.93
RF	0.530	16.151	2.92
GBDT	0.539	15.996	2.90
LGBM	0.540	15.973	2.89
XGBoost	0.545	15.890	2.88

The implementation of the ensemble learning system is carried out in the following sequence: training the prediction of a set of single models (base learners) at the first-level learner stage, the prediction of combined model (meta-learner) at second-level learner stage, and then the performance evaluation[71].

5.4.2 Main findings

To improve the accuracy of the model and reduce the over-fitting, we adopt the methods of averaging and stacking sequentially. Based on the many could be better than all theorem, we only include some single learners to compose an ensemble model instead of applying all of them, which obtains superior performance[69, 187].

Additionally, to reach a desired ensemble model, the selected base learners should be as diverse as possible between each other, while they can also report good performance independently[69, 188]. In this work, the combination strategy is as follows:

1) Combine the models with the Averaging method and combine the following models: SVM, KRR GBDT, XGBoost.

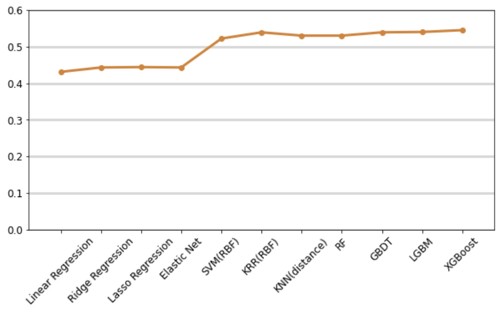
2) Combine the models with the Stacking method and combine the following models SVM, KRR, lightGBM, XGBoost.

The results of the Averaging ensemble model and Stacking ensemble model are illustrated in Table 5.3 and Figure 5.4. We reported all results regarding the base models as well as the ensemble models in Figure 5.4 and Figure 5.5.

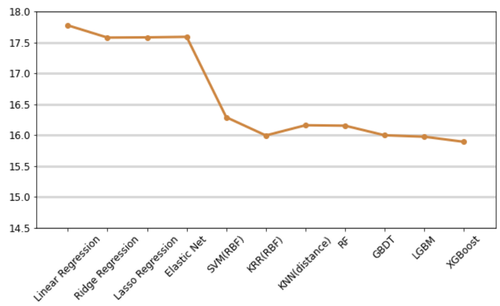
Table 5.3: Comparisons with averaging model and stacking model.

Model	Averaging Model	Stacking Model
R^2	0.550	0.548
RMSE	15.790	15.847
Percentage of error(%)	2.86	2.87

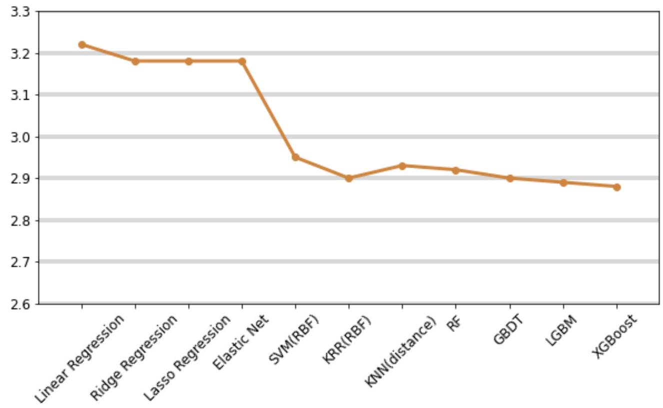
From Table 5.3 and Figure 5.4, we can see that the two assembled models outperform the single ones from all the evaluation metrics, including R^2, RMSE and percentage of error (PE). A lower RMSE and a lower PE indicate a better model whilst vice versa for R^2.



(a) R-squared



(b) RMSE



(c) Percentage of Error(%)

Figure 5.4: Model Results on MES and LES.

From the statistics and plots, it is easy to see that the assembled models, averaging ensemble model and stacking ensemble model, have lower RMSE and PE values and higher R-square values.

To better measure the performance of the models, the prediction errors of the models are collected in ten folds random evaluation. L, M, N and O is reported as the averaging ensemble and stacking ensemble models in Figure 5.5. As we can see from Figure 5.5, the averaging combination of SVM, KRR, LGBM, XGBoost (M) is better than other stacking method and single models.

By far, 4905 real-world samples from the on-site collection are used in our models for predicting steel quality. It is worth to consider that the model accuracy will be further improved when a larger and more comprehensive dataset is continuously generated and fed into our models in the near future.

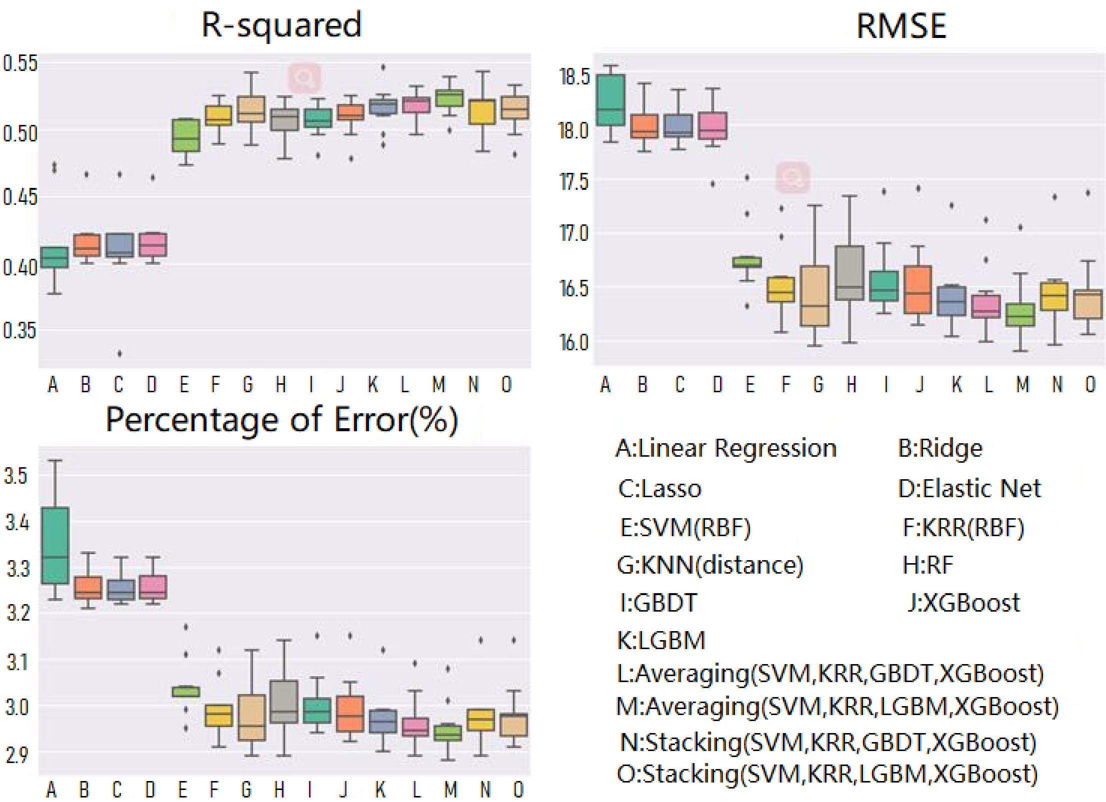


Figure 5.5: Models Ten Folds Random Evaluation.

5.5 Summary

In this chapter, we propose a practical machine learning enabled prediction system for forecasting steel quality control, based on historical observation data. We evaluate the system, specifically concerning the performance of the deployed ensemble models on two types of datasets: MES and LES. The ensemble machine learning systems for steel quality control achieves performances which are significantly higher than other 11 baseline approaches, confirming that our prediction system is better and more robust to the steel quality prediction.

So far, this successfully deployed steel quality prediction system lays a foundation to incorporate machine learning and data analytics technologies in the real-world production process. The design of this system has taken the advantages from the data collection, domain knowledge building as well as the machine learning technologies. Therefore, there is great potential for discovering and exploiting a more sophisticated machine learning model for improving the accuracy of the steel quality prediction when data is accumulated dramatically.

In the future, we will consider other model combination strategies together with other types of base learners, such as neural networks with the fuzzy system[189, 190]. Meanwhile, with the massive data sources, some machine learning strategies, such as deep learning, will be incorporated in the next stage. Notably, using deep neural networks can not only have the potential for mining latent information, but also relieving the workload of feature engineering.

Chapter 6

Summary and Prospect

6.1 Summary

Iron and steel are the basic raw material for all industries and widely used in various fields, including buildings, bridges, vessels, containers, medical devices and automobiles. However, the production procedure of steel is very complicated, which leads to quite complicated quality control procedure of steel. Therefore, the steel quality control is regarded as a great and huge challenge for the whole steel industry.

With the development of social economy, consumers are increasingly pursuing high-quality products and services, and enterprises are more committed to improving the quality of products and services. Under such background, the theory of total quality management has been put forward. It has attracted great attention from Chinese and international enterprises and theoretical circles since it is put forward. It is a great success in implementation since its introduction in China. For a long time, Chinese iron and steel enterprises have carried out a lot of research and application work in total quality management, and achieved remarkable results, but there are still some shortcomings. Under the new development situation, followed higher and new requirements put forward for the product and service quality of enterprises, iron and steel enterprises also put forward the requirements for further development of total quality management.

This research studies the quality management of Nanjing Iron and Steel Company, which has established the quality management system; however, the system still has some defects. The quality management system has a high demand on the statistical basis of business personnel, which leads to the limitation of the utilization rate of the system. At present, the quality management system is mainly focused on the post-analysis and processing of the data, and the quality problems of the products were mainly tested by sampling-experiment method, so that the quality of the products could not be tested in time or predicted in advance. In the quality control system, with different responsibilities and functions, each information system is only responsible for storing data with corresponding functions. The data between different information systems are independent from each other to form data islands. The iron and steel production process belongs to the process industry, and the data in multiple information systems are combined with each other to conduct in-depth analysis, prediction and early warning of product quality. Therefore, it is necessary to introduce some new product quality control methods in the iron and steel industry.

In recent years, based on Industry 4.0 and intelligent manufacturing plans some countries have already released, intelligent manufacturing has become the core of all industries in the world, and it will also affect the future of a country. The implementation of intelligent manufacturing will also bring huge profits to enterprises. It can not only improve the automation and information level of enterprises, but also help reduce the number of personnel and improve efficiency. Most importantly, it helps to improve the quality and reduce the cost of products[41].

Quality control is an important part of intelligent manufacturing and also the most critical link in enterprise production[42], especially in the iron and steel industry. With the promotion of Industry 4.0 and intelligent manufacturing, the iron and steel enterprises have also introduced big data analysis-based technology and AI-based technology to improve their competitiveness in the industry. The application of big data analysis methods and AI technology in quality control will enable enterprises to generate accurate quality analysis and good predictions based on the data distributed throughout the factory, realizing online adjustment of production process. This is not only conducive to the positive significance and effect of enterprises product quality control, but also conducive to reducing product costs.

Based on the shortcomings and pain points of traditional quality control systems and the advantages of new methods based on big data analysis in the context of intelligent manufacturing, and on the basis of a large amount of data related to the product life cycle, this thesis, by taking the iron and steel manufacturing industry as the research object, discusses the application of machine learning and deep learning in three quality control aspects: scrap steel intelligent rating, prediction of longitudinal crack on continuous casting slab surface and mechanical properties prediction of steel plate.

In the iron and steel smelting process, scrap steel is an important raw material for steelmaking in electric furnace and in converter; so iron and steel production companies will buy a large number of scrap steel for steel-making every year. The quality evaluation of purchased scrap steel mainly depends on manual experience, which is subjective, so different experienced masters may give different grades for the same scrap steel. Relying on manual experience to rate scrap steel may also result in "private relationship scrap steel", which makes the rating standard not opaque, standardized and open. At the same time, due to the large amount of dust contained in scrap steel and the high decibel noise generated during the acceptance of scrap steel, it is liable to damage the health of the acceptance rating personnel.

Based on the above problems, this thesis proposes a scrap steel intelligent rating system, which combines the existing ERP system, metering system, image acquisition equipment and artificial intelligence platform of iron and steel enterprises to rate the scrap steel. The most important core of the scrap steel intelligent rating system is the computer vision algorithm based on deep learning in the artificial intelligence platform, which combines object detection and image classification algorithms to judge the material type of scrap steel. The average accuracy rate between the material type predicted on-line by the scrap steel intelligent rating system proposed in this research and the on-site acceptance result of the inspector reached about 83% and that between the amount of impurities deducted by the system from the verification set and the on-site acceptance result of the inspector reached about 85%. Therefore, the accuracy rate of the scrap steel intelligent rating system proposed in this thesis is very good. The successful application of the intelligent system avoids the possible problems in the original manual acceptance method, making the rating standard transparent, standardized and open, and saving a lot for Nanjing Iron and Steel Company.

After steelmaking, it will enter the stage of continuous casting. The surface quality control of continuous casting slab is very important, especially the detection of longitudinal cracks on the surface of continuous casting slab. The longitudinal cracks on the surface of continuous casting slab reduce the comprehensive yield of continuous casting, and even cause the whole slab to be scrapped in severe cases, thus affecting the smooth operation of other technological processes. In addition, they may be inherited into rolled products, affecting the quality of rolled products. At present, Nanjing Iron and Steel Company mainly detects the surface quality of continuous casting slab by manual way. The manual detection method requires experienced technicians, and the efficiency of manual detection is low. Visual inspection is prone to fatigue and error, and different technicians have different criteria for judging whether there are cracks on the surface.

In view of the above problems, this research firstly focuses on how some important chemical elements and process parameters in the production process affect the probability of longitudinal cracks on the surface of continuous casting slab. It is found through this study that the relationship between chemical elements and process parameters and the probability of longitudinal cracks on the surface of continuous casting slab is nonlinear. The main study results are as follows:

1. When the carbon content is in the range of $0.05\% - 0.15\%$, the probability of longitudinal cracks in the continuous casting slab is relatively high, and when the carbon content is 0.1%, the probability of longitudinal cracks on the surface of the continuous casting slab is the highest. When there is no special requirement for carbon content in steel products, in order to ensure the quality of produced products, the carbon content may be adjusted downward or upward.

2. When the copper content is in the range of $0.01\% - 0.05\%$, the longitudinal crack index increases with the increase of the copper content. In order to avoid longitudinal cracks on the surface of continuous casting slab caused by the copper content, the copper content shall be controlled below 0.05%.

3. With the increase of phosphorus content, the probability of longitudinal cracks on the surface of continuous casting slab first decreases and then increases, and the range of decrease and increase is very large. When the phosphorus content is

about 0.0135%, the probability of longitudinal crack on the surface of the slab is the lowest. In order to reduce the probability of longitudinal cracks in continuous casting slab, the phosphorus content in molten steel shall be about 0.0135%.

4. When the sulfur content is 0.0052%, the probability of longitudinal cracks is the lowest; when the sulfur content is greater than 0.0075%, the probability of longitudinal cracks on the surface of continuous casting slab increases rapidly with the increase of the sulfur content. Therefore, in order to ensure the quality of continuous casting slab in the actual production process, the sulfur content shall be controlled at about 0.0052%.

5. With the increase of heat flux at wide side, the probability of longitudinal cracks on the surface of continuous casting slab first decreases and then increases. When the heat flux at wide side is about $2.375 MW/m^2$, the probability of longitudinal cracks on the slab surface is the lowest. When the heat flux at wide side is between $2.3 MW/m^2$ and $2.45 MW/m^2$, the probability of longitudinal cracks on the slab surface is in a plateau period. Therefore, in order to reduce the probability of longitudinal cracks during the production, the heat flux at wide side in the continuous casting process is kept between $2.3 MW/m^2$ and $2.45 MW/m^2$.

6. When the heat flux at narrow side is less than $2.25 MW/m^2$, the heat flux rate at narrow side of longitudinal cracks on the surface of continuous casting slab is very high; while it is between $2.4 MW/m^2$ and $2.6 MW/m^2$, the probability of longitudinal cracks on the slab surface is very low. In order to reduce the probability of longitudinal cracks during the production, the heat flux at narrow side in continuous casting process shall be kept between $2.4 MW/m^2$ and $2.6 MW/m^2$.

7. The probability of longitudinal cracks on the surface of continuous casting slab decreases with the increase of casting speed. According to the actual production data, the casting speed does not exceed $1.4 m/min$, and the probability of longitudinal cracks on the slab surface may be reduced if the casting speed is kept slightly higher than $1.30 m/min$ combined with the information obtained in the plot.

Through the above study, it can be seen that the relationship between chemical elements and process parameters and the probability of longitudinal cracks on the surface of continuous casting slab is nonlinear, and as the field of iron and steel belongs to the process industry, the numerical control of the former process will have an impact on the control of the latter process. Therefore, it is very difficult to control the probability of longitudinal cracks on the surface of continuous casting slab by controlling the value range of each process parameter independently. Next, this thesis proposes a five-layer service-oriented intelligent steel manufacturing system based on industrial IoT, including sensor layer, network layer, database layer, model layer and application layer. In the system, the first four layers are responsible for data acquisition and preprocessing, and the last layer is responsible for modeling the preprocessed data. In the application layer, a pre-optimised BP neural network is proposed to model the pre-processed data, and the results show that this method can effectively predict the longitudinal cracks on the surface of continuous casting slab. The successful application of the intelligent system avoids the problems existing in the original manual inspection method, provides guarantee for the subsequent processes, and saves some costs for Nanjing Iron and Steel Company.

Finally, the physical properties of rolled steel plates are studied. The physical properties of steel plates represent whether the quality of the steel plates is qualified. If they are qualified, the products will be delivered; if not, they need to be reprocessed. At present, the physical properties of steel plates are tested by sampling-experiment method, which is not only costly but also time-consuming, and it will also affect the efficiency and cost of factory production.

In order to improve production efficiency, reduce production costs and speed up the delivery of products to customers, this thesis adopts the integrated learning method to fit the chemical elements, production process parameters collected during the production and the physical properties tested by experiments. The results show that the Averaging integrated model based on SVM, KRR, LGBM and XGBoost can well predict the mechanical properties of steel plates.

To sum up, under the background of Industry 4.0 and Made in China 2025, this thesis applies AI algorithm in the iron and steel quality field to solve the pain points in some traditional quality management and control systems, so as to provide new ideas and methods for quality prediction, management, control and cost control of iron and steel products, thus greatly improving the quality detection efficiency of the industry.

6.2 Future Outlook

Under the background of Industry 4.0 and Made in China 2025, enterprises in various industries are undergoing digital and intelligent transformation, and the steel industry is no exception. In this thesis, some quality monitoring problems in the field of iron and steel are studied, which has made some contributions to the quality monitoring in the field of iron and steel, but there are still many shortcomings. From a macro point of view, the future study is mainly divided into two aspects: the first is the supplement of the studied quality problems and the attempt of other better algorithms; the second is the exploration of other quality problems that have not been studied.

First of all, this section will introduce the contents that need to be supplemented for the quality problems that have been studied. According to the sequence of this study, we put forward a scrap steel intelligent rating system in the task of scrap steel quality evaluation, which is mainly divided into two modules: image acquisition and scrap steel material type identification. In the image acquisition module, the camera picture acquisition is controlled by the truck driver through swiping the card, but there is no control over the driver's swiping time and direction. In the future study and implementation, some control functions will be added to standardize the picture acquisition process. For example, the following control function will be added:

(1) Add the function of control over the swiping time to avoid swiping the card two times within a short time and increase the acquisition of effective pictures;

(2) Optimize the current total photographing time according to the loading time of each type of vehicle, so as to avoid including the picture of scrap steel loading on the next vehicle into the picture of this loading when the driver forgets to swipe the card for confirmation;

(3) Add the evaluation and warning function of the swiping direction. When the driver chooses the wrong direction, the system will evaluate the captured picture and finally give an alarm to remind the driver to choose the parking direction again;

(4) Add an access control system associated with scrap steel intelligent rating at the wharf where scrap steel loading takes place, and prohibit the scrap steel truck from driving out of the rating point when it does not swipe the card for rating during loading;

(5) Add the push function of card swiping information and loading point information, so as to avoid repeated card swiping when the driver is not sure whether the card swiping is successful, and to ensure the effective operation of the system;

(6) Add the function of auto-focusing the camera according to the type of vehicle, so that the picture of scrap steel material type in the carriage is very clear.

In the identification module of scrap steel material type, this study only focuses on three common types of scrap steels, which means that the system proposed this time cannot identify other scrap steel material types except these three ones. In the future study, the training on identification of other material types will be added to enrich the identification range of the model. In this study, the material type of each zone is given, and the granularity is relatively coarse. In order to improve the accuracy of the current algorithm, in the following research, we mainly consider the combination of machine learning algorithm and image segmentation algorithm to determine the grade of scrap steel and calculate the weight of impurities. The image segmentation algorithm is used to identify and calculate the various types of steel scrap and the area of each type of steel scrap in the effective picture, the type of impurity and its corresponding area. The machine learning algorithm is used to fit the results produced by the image segmentation algorithm and the corresponding manually-given scrap grades and the weight of impurities.

In the task of predicting longitudinal cracks on the surface of continuous castingslab, this study is divided into two levels: data acquisition and model application. At the level of data acquisition, the data of some main influencing factors in the experience are collected according to the experience of technicians, while some influencing factors are not collected. In the future study, some sensors will be added to obtain the data of more influencing factors, to guarantee the accurate prediction of the model. At the level of model application, this study mainly uses pre-optimised BP neural network to fit the data. In the following research, we will investigate more layers of neural network, integrated machine learning algorithm, fuzzy algorithm and other algorithms, as well as the combination of algorithms, to seek for higher precision algorithm. At the same time, some optimization algorithms used to optimize the parameters of machine learning and neural networks will also be included in the next stage.

In the task of predicting the physical properties of steel plates, this thesis proposes to make the predication by applying the integrated learning method and analyzing the chemical elements and process variables in the production, and try to apply other machine learning methods to predict them in future study, so that the predicted values are close to the true values, thus providing a strong guarantee for the quality of steel plates.

Secondly, as iron and steel smelting is a complex process, every small link involves intelligent control such as quality management, energy balance and process optimization. In the future study, the contents of quality management shall be expanded and enriched to ensure the quality of products and improve the market competitiveness of the company.

In the field of quality control, some variables in the available data are taken as a representative value in studies on the prediction of mechanical properties of steel plates. For example, the temperature. In the process of steel plate smelting, the temperature values at the head, tail and middle of the steel plate are different, and it is a bit rough to apply one temperature value to represent the overall temperature of a steel plate, later we will increase the intensity of data collection, and according to the speed of data collection, different temperature values will correspond to different locations of the steel plate. Of course, the data of other smelting process variables will also be collected in this way, and to ensure the synchronization of different variables, namely different variables correspond to the same position of the steel plate at the same time, so as to provide a strong guarantee for the detailed prediction of the mechanical properties of the steel plate.

Steel plate surface quality inspection is also an important part of steel plate quality control. In future research, we will investigate the location and types of steel plate surface defects and apply automated, intelligent and standardized methods to assist or replace the existing manual inspection methods in the production line. This is to solve the problems of missed inspections, quality objections and high labor intensity, while being able to quickly trace the causes of defects.

The two examples of quality control mentioned above are the quality control of finished steel plates. Quality control can also be refined to the production process, such as the prediction of temperature, C, P, and S content at the ending point of the converter, which can apply the chemical element content of the molten iron, the ironmaking process, and the production process of converter steelmaking to predict the temperature, C, P, and S content at the end of the converter.

In the process of steel product production, there is quality control at every step. In the future, we intend to collect all the production data of a steel product together to monitor the quality of the product in real time, immediately alerting if there is a quality problem and tracing the cause of the quality problem. If it is a critical quality problem, production is stopped to reduce unnecessary production costs as well as to make plans for re-production and to improve the speed of product delivery.

Meanwhile, we will try to investigate the application of intelligent methods in other aspects of the steel field to achieve digital and intelligent steel smelting. For example, in the area of material tracking, computer vision methods are applied to identify steel slab numbers and steel plate numbers. This is combined with other systems of steel production in order to quickly and accurately determine the location of billets and steel plates. In terms of planning and scheduling, energy dispatching, etc., we will also try to apply intelligent means to optimize them, realize the effective and full utilization of resources, reduce unnecessary waste of resources, lower production costs, and provide a strong guarantee for the digital and intelligent transformation of the steel industry.

Bibliography

[1] N. Guelton, C Lops, and H. Sordini. Cross coating weight control by electromagnetic strip stabilization at the continuous galvanizing line of arcelormittal florange. Metallurgical and Materials Transactions B, 47(4):2666–2680, 2016.

[2] J. Guo, MD Seo, C. B. Shi, J. W. Cho, and S. H. Kim. Control of crystal morphology for mold flux during high-aluminum ahss continuous casting process. Metallurgical & Materials Transactions B, 47(4):2211–2221, 2016.

[3] A. Adema, Y. Yang, and R. Boom. DEM-CFD modelling of the ironmaking blast furnace:. Mechanical Maritime & Materials Engineering, 16(5):1081–1085, 2009.

[4] Jean-Philippe Harvey and Aimen E Gheribi. Process simulation and control optimization of a blast furnace using classical thermodynamics combined to a direct search algorithm. Metallurgical and Materials Transactions B, 45(1):307–327, 2014.

[5] Heikki Jalkanen and Lauri Holappa. Converter steelmaking. In Treatise on process metallurgy, volume 3.

[6] Jeremy AT Jones, Ben Bowman, and Peter A Lefrank. Electric furnace steelmaking. The making, shaping and treating of steel, 1, 1998.

[7] H. Yang, X. Zhang, K. Deng, W. Li, G. Yong, and L. Zhao. Mathematical simulation on coupled flow, heat, and solute transport in slab continuous casting process. Metallurgical & Materials Transactions B, 29(6):1345–1356, 1998.

[8] He, LIANG, 1), Zhiyong, XUE, 2), Chunjing, WU, 1), and Qing. Research on continuous core-filling casting forming process of copper-clad aluminum bimetal

composite material. Acta Metallurgica Sinica(English Letters), 23(3):206–214, 2010.

[9] William L Roberts. Hot rolling of steel. CRC Press, 1983.

[10] Zhongkai Ren, Hong Xiao, Xiao Liu, and Zhichao Yan. An analysis of the metal transverse flow in the roll gap for ultra-thin strip rolling using the energy method. ISIJ International, 58(2):309–315, 2018.

[11] William L Roberts. Cold rolling of steel. Routledge, 2017.

[12] Sanjay L Ahire, Robert Landeros, and Damodar Y Golhar. Total quality management: a literature review and an agenda for future research. Production and Operations Management, 4(3):277–306, 1995.

[13] I Raine. Total quality management for sustainable development. Business Process Management Journal, 12:632, 2006.

[14] Barrie G Dale, Ton Van Der Wiele, and Jos Van Iwaarden. Managing quality. John Wiley & Sons, 2007.

[15] Caulcutt and Roland. Statistical process control (spc). Assembly Automation, 16(4):10–14, 1996.

[16] R. M. Fortuna. Beyond quality: taking spc upstream. Quality Progress, 21(6):23–28, 1988.

[17] A. V. Feigenbaum. Total quality control / a.v. feigenbaum. Harvard business review, 34(6), 1991.

[18] B. Chen, H. U. Ning, Y. Wang, H. Yang, and CEPREI. The practice of total quality management in a corporation. Electronic Product Reliability and Environmental Testing, 2018.

[19] L. H. Sui and Z. X. Shu. New model of total quality management in the 21 century. Policy-making Reference, 2003.

[20] Tang Xiaofen. Investigation on quality management maturity of shanghai enterprises. The TQM Journal, 2013.

[21] T. Mady and M. Quality management practices. International Journal of Quality & Reliability Management, 26(3):214–233, 2009.

[22] Armand V Feigenbaum. Total quality-control. Harvard business review, 34(6):93–101, 1956.

[23] Harry Roberts. Quality is personal: a foundation for total quality management. Simon and Schuster, 2010.

[24] R Ashley Rawlins. Total quality management (TQM). AuthorHouse, 2008.

[25] Joseph M Juran et al. Juran on quality by design: the new steps for planning quality into goods and services. Simon and Schuster, 1992.

[26] M. Zairi and M. A. Youssef. Quality function deployment. International Journal of Quality & Reliability Management, 12(6):9–23, 1995.

[27] Pall and GabrielA. Quality process management. Quality process management, 1987.

[28] James W Dean Jr and David E Bowen. Management theory and total quality: improving research and practice through theory development. Academy of management review, 19(3):392–418, 1994.

[29] M. Kano. Statistical process control using process chemometrics. Systems Control & Information, 48(3):39–50, 2004.

[30] Hiroyasu Shigemori. Quality and operation management system for steel products through multivariate statistical process control. In Proceedings of the World Congress on Engineering, volume 1, pages 119–124. IAENG London, 2013.

[31] P Kaushik, M Lowry, H Yin, and H Pielet. Inclusion characterisation for clean steelmaking and quality control. Ironmaking & Steelmaking, 39(4):284–300, 2012.

[32] Alan M Turing and J Haugeland. Computing machinery and intelligence. MIT Press Cambridge, MA, 1950.

[33] Warren S McCulloch and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. The bulletin of mathematical biophysics, 5(4):115–133, 1943.

[34] Lei ZOU and Xian-feng ZHANG. Application and development of the artificial intelligence. Netinfo Security, 2:11–13, 2012.

[35] Tom M Mitchell et al. Machine learning. 1997.

[36] Ethem Alpaydin. Introduction to machine learning. MIT press, 2020.

[37] Petia Koprinkova-Hristova, Valeri Mladenov, and Nikola K Kasabov. Artificial neural networks. Eur. Urol, 40(1):245, 2015.

[38] Yu-chen Wu and Jun-wen Feng. Development and application of artificial neural network. Wireless Personal Communications, 102(2):1645–1656, 2018.

[39] HB Wang, LY Zhang, and ZHJ Hu. Theory on artificial neural network and its application. Electron Technol Shanxi, 2:41–43, 2006.

[40] Shiyong Wang, Jiafu Wan, Daqiang Zhang, Di Li, and Chunhua Zhang. Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Computer networks, 101:158–168, 2016.

[41] Maqbool Khan, Xiaotong Wu, Xiaolong Xu, and Wanchun Dou. Big data challenges and opportunities in the hype of industry 4.0. In 2017 IEEE International Conference on Communications (ICC), pages 1–6, 2017.

[42] X. Zhang, J. Zhao, W. Wei, L. Cong, and W. Feng. An optimal method for prediction and adjustment on byproduct gas holder in steel industry. Expert Systems with Applications, 38(4):4588–4599, 2011.

[43] J Zhao, Q Liu, and W Pedrycz. Effective noise estimation-based online prediction for byproduct gas system in steel industry. IEEE Transactions on Industrial Informatics, 8(4):953–963, 2012.

[44] A. Kusiak and A. Verma. A data-mining approach to monitoring wind turbines. IEEE Transactions on Sustainable Energy, 3(1):150–157, 2011.

[45] F. Castellani, D. Astolfi, P. Sdringola, S. Proietti, and L. Terzi. Analyzing wind turbine directional behavior: Scada data mining techniques for efficiency and power assessment. Applied Energy, 185(pt.2):1076–1086, 2015.

[46] G. A. Susto, A. Schirru, S. Pampuri, S. Mcloone, and A. Beghi. Machine learning for predictive maintenance: A multiple classifier approach. IEEE transactions on industrial informatics, 11(3):812–820, 2015.

[47] S. Pan, T. Morris, and U. Adhikari. Developing a hybrid intrusion detection system using data mining for power systems. IEEE Transactions on Smart Grid, 6(6):3104–3113, 2015.

[48] LIU Jing-wen. Application of artificial intelligence in metallurgical electric automation control. World Nonferrous Metals, page 01, 2018.

[49] LN An and Shi-jie Zhang. Progress and prospects of expert system. Application Research of Computers, 24(12):1–4, 2007.

[50] D. T. Pham and A. A. Afify. Machine-learning techniques and their applications in manufacturing. Proceeding of the Institute of Mechanical Engineers .part B.journal of Engineering Manufacture, 219(5):395–412, 2005.

[51] MH Fazel Zarandi and P Ahmadpour. Fuzzy agent-based expert system for steel making process. Expert systems with applications, 36(5):9539–9547, 2009.

[52] Purusottam Narayana Mishra, SK Kak, and SC Srivastava. An expert system for ld steel making. IETE Journal of Research, 47(1-2):85–89, 2001.

[53] Somnath Mitra, Rajeev Kumar Singh, and Ashis Kumar Mondal. An expert system based process control system for silicon steel mill furnace of rourkela steel plant. In 2014 Fourth International Conference of Emerging Applications of Information Technology, pages 29–33. IEEE, 2014.

[54] LI Ze-hui. A soft-measuring method for pulverized coal particle size of blast furnace grinding system based on cbr. Metallurgical Industry Automation, page 03, 2013.

[55] Sheena Angra and Sachin Ahuja. Machine learning and its applications: A review. In 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), pages 57–60. IEEE, 2017.

[56] Hui Zou. The adaptive lasso and its oracle properties. Journal of the American statistical association, 101(476):1418–1429, 2006.

[57] Arthur E Hoerl and Robert W Kennard. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1):55–67, 1970.

[58] Hui Zou and Trevor Hastie. Regularization and variable selection via the elastic net. Journal of the royal statistical society: series B (statistical methodology), 67(2):301–320, 2005.

[59] Corinna Cortes and Vladimir Vapnik. Support-vector networks. Machine learning, 20(3):273–297, 1995.

[60] Zhihua Ye and Honglian Li. Based on radial basis kernel function of support vector machines for speaker recognition. In 2012 5th International Congress on Image and Signal Processing, pages 1584–1587. IEEE, 2012.

[61] F. Vermet. Statistical Learning Methods. Big Data for Insurance Companies, 2018.

[62] Bernhard Scholkopf and Alexander J Smola. Learning with kernels: support vector machines, regularization, optimization, and beyond. Adaptive Computation and Machine Learning Series, 2018.

[63] Erin J Bredensteiner and Kristin P Bennett. Multicategory classification by support vector machines. In Computational Optimization, pages 53–79. Springer, 1999.

[64] Vladimir Vovk. Kernel ridge regression. In Empirical inference, pages 105–116. Springer, 2013.

[65] Thierry Denoeux. A k-nearest neighbor classification rule based on dempster-shafer theory. In Classic works of the Dempster-Shafer theory of belief functions, pages 737–760. Springer, 2008.

[66] David Adedayo Adeniyi, Zhaoqiang Wei, and Y Yongquan. Automated web usage data mining and recommendation system using k-nearest neighbor (knn) classification method. Applied Computing and Informatics, 12(1):90–108, 2016.

[67] Ankit Dixit. Ensemble Machine Learning: A beginner's guide that combines powerful machine learning algorithms to build optimized models. Packt Publishing Ltd, 2017.

[68] Thomas G Dietterich. Ensemble methods in machine learning. In International workshop on multiple classifier systems, pages 1–15. Springer, 2000.

[69] Zhi-Hua Zhou. Ensemble learning, encyclopedia of biometrics, 2009.

[70] TG Ditterrich. Machine learning research: four current direction. Artificial intelligence magzine, 4:97–136, 1997.

[71] Zhi-Hua Zhou. Ensemble methods: foundations and algorithms. Chapman and Hall/CRC, 2019.

[72] Funda Güneş, Russ Wolfinger, and Pei-Yi Tan. Stacked ensemble models for improved prediction accuracy. In Proc. Static Anal. Symp., pages 1–19, 2017.

[73] Silke Janitza, Gerhard Tutz, and Anne-Laure Boulesteix. Random forest for ordinal responses: prediction and variable selection. Computational Statistics & Data Analysis, 96:57–73, 2016.

[74] Real Carbonneau, Kevin Laframboise, and Rustam Vahidov. Application of machine learning techniques for supply chain demand forecasting. European Journal of Operational Research, 184(3):1140–1154, 2008.

[75] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. Annals of statistics, pages 1189–1232, 2001.

[76] Ron Appel, Thomas Fuchs, Piotr Dollár, and Pietro Perona. Quickly boosting decision trees–pruning underachieving features early. In International conference on machine learning, pages 594–602. PMLR, 2013.

[77] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pages 785–794, 2016.

[78] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. Advances in neural information processing systems, 30:3146–3154, 2017.

[79] Thorsten Wuest, Daniel Weimer, Christopher Irgens, and Klaus-Dieter Thoben. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research, 4(1):23–45, 2016.

[80] László Monostori, András Márkus, Hendrik Van Brussel, and E Westkämpfer. Machine learning approaches to manufacturing. CIRP annals, 45(2):675–712, 1996.

[81] Der-Chang Li and Yao-San Lin. Using virtual sample generation to build up management knowledge in the early manufacturing stages. European Journal of Operational Research, 175(1):413–434, 2006.

[82] Jean-Loup Loyer, Elsa Henriques, Mihail Fontul, and Steve Wiseall. Comparison of machine learning methods applied to the estimation of manufacturing cost of jet engine components. International Journal of Production Economics, 178:109–119, 2016.

[83] Paolo Priore, David de la Fuente, Javier Puente, and José Parreño. A comparison of machine-learning algorithms for dynamic scheduling of flexible manufacturing systems. Engineering Applications of Artificial Intelligence, 19(3):247–255, 2006.

[84] Der-Chiang Li, Chih-Sen Wu, Tung-I Tsai, and Yao-San Lina. Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge. Computers & Operations Research, 34(4):966–982, 2007.

[85] Susana Fernández, Ricardo Aler, and Daniel Borrajo. Machine learning in hybrid hierarchical and partial-order planners for manufacturing domains. Applied Artificial Intelligence, 19(8):783–809, 2005.

[86] Joonki Hong and Vittaldas V Prabhu. Distributed reinforcement learning control for batch sequencing and sizing in just-in-time manufacturing systems. Applied Intelligence, 20(1):71–87, 2004.

[87] Min Kong, Jun Pei, Jin Xu, Xinbao Liu, Xiaoyu Yu, and Panos M Pardalos. A robust optimization approach for integrated steel production and batch delivery scheduling with uncertain rolling times and deterioration effect. International Journal of Production Research, 58(17):5132–5154, 2020.

[88] Davide Armellini, Paolo Borzone, Sara Ceschia, Luca Di Gaspero, and Andrea Schaerf. Modeling and solving the steelmaking and casting scheduling problem. International Transactions in Operational Research, 27(1):57–90, 2020.

[89] Kunkun Peng, Quanke Pan, and Biao Zhang. An improved artificial bee colony algorithm for steelmaking–refining–continuous casting scheduling problem. Chinese journal of chemical engineering, 26(8):1727–1735, 2018.

[90] MH Fazel Zarandi and F Dorry. A hybrid fuzzy pso algorithm for solving steelmaking-continuous casting scheduling problem. International Journal of Fuzzy Systems, 20(1):219–235, 2018.

[91] Jian-ping Yang, Bai-lin Wang, Qian Liu, Min Guan, Tie-ke Li, Shan Gao, da Guo Wei, and Qing Liu. Scheduling model for the practical steelmaking-continuous casting production and heuristic algorithm based on the optimization of furnace-caster matching mode. ISIJ International, 60(6):1213–1224, 2020.

[92] Thorsten Wuest, Christopher Irgens, and Klaus-Dieter Thoben. An approach to monitoring quality in manufacturing using supervised machine learning on product state data. Journal of Intelligent Manufacturing, 25(5):1167–1180, 2014.

[93] Hongbin Jia, Yi Lu Murphey, Jinajun Shi, and Tzyy-Shuh Chang. An intelligent real-time vision system for surface defect detection. In Proceedings of the 17th

International Conference on Pattern Recognition, 2004. ICPR 2004., volume 3, pages 239–242. IEEE, 2004.

[94] Chunguang Shen, Chenchong Wang, Xiaolu Wei, Yong Li, Sybrand van der Zwaag, and Wei Xu. Physical metallurgy-guided machine learning and artificial intelligent design of ultrahigh-strength stainless steel. Acta Materialia, 179:201–214, 2019.

[95] Fucun Li, Jianqing Wu, Fang Dong, Jiayin Lin, Geng Sun, Huaming Chen, and Jun Shen. Ensemble machine learning systems for the estimation of steel quality control. In 2018 IEEE International Conference on Big Data (Big Data), pages 2245–2252. IEEE, 2018.

[96] Qian Xie, Manu Suvarna, Jiali Li, Xinzhe Zhu, Jiajia Cai, and Xiaonan Wang. Online prediction of mechanical properties of hot rolled steel plate using machine learning. Materials & Design, 197:109201, 2021.

[97] Tianqi Cheng and Guici Chen. Prediction of mechanical properties of hot-rolled strip steel based on pca-gbdt method. In Journal of Physics: Conference Series, volume 1774, page 012002. IOP Publishing, 2021.

[98] Miguel Cuartas, Estela Ruiz, Diego Ferreño, Jesús Setién, Valentín Arroyo, and Federico Gutiérrez-Solana. Machine learning algorithms for the prediction of non-metallic inclusions in steel wires for tire reinforcement. Journal of Intelligent Manufacturing, pages 1–13, 2020.

[99] Gülser Köksal, Inci Batmaz, and Murat Caner Testik. A review of data mining applications for quality improvement in manufacturing industry. Expert systems with Applications, 38(10):13448–13467, 2011.

[100] Alok Kumar Choudhary, Jenny A Harding, and Manoj Kumar Tiwari. Data mining in manufacturing: a review based on the kind of knowledge. Journal of Intelligent Manufacturing, 20(5):501–521, 2009.

[101] Dazhong Wu, Connor Jennings, Janis Terpenny, Robert X Gao, and Soundar Kumara. A comparative study on machine learning algorithms for smart manu-

facturing: tool wear prediction using random forests. Journal of Manufacturing Science and Engineering, 139(7), 2017.

[102] Ratna Babu Chinnam. Support vector machines for recognizing shifts in correlated and other manufacturing processes. International journal of production research, 40(17):4449–4466, 2002.

[103] Achmad Widodo and Bo-Suk Yang. Support vector machine in machine condition monitoring and fault diagnosis. Mechanical systems and signal processing, 21(6):2560–2574, 2007.

[104] ES Grishin. Development of intelligent algorithms for the continuous diagnostics and condition monitoring subsystem of the equipment as part of the process control system of a stainless steel pipe production enterprise. In IOP Conference Series: Materials Science and Engineering, volume 939, page 012027. IOP Publishing, 2020.

[105] Matthias Karner, Robert Glawar, Wilfried Sihn, and Kurt Matyas. An industry-oriented approach for machine condition-based production scheduling. Procedia CIRP, 81:938–943, 2019.

[106] Gian Antonio Susto, Simone Pampuri, Andrea Schirru, Guiseppe De Nicolao, Sean F McLoone, and Alessandro Beghi. Automatic control and machine learning for semiconductor manufacturing: Review and challenges. In Proceedings of the 10th European Workshop on Advanced Control and Diagnosis (ACD 2012), 2012.

[107] David G Lowe. Distinctive image features from scale-invariant keypoints. International journal of computer vision, 60(2):91–110, 2004.

[108] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), volume 1, pages 886–893. Ieee, 2005.

[109] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. Computer Science, 2014.

[110] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.

[111] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, 25:1097–1105, 2012.

[112] Matthew D Zeiler and Rob Fergus. Visualizing and understanding convolutional networks. In European conference on computer vision, pages 818–833. Springer, 2014.

[113] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence, 37(9):1904–1916, 2015.

[114] Robert Hecht-Nielsen. Theory of the backpropagation neural network. In Neural networks for perception, pages 65–93. Elsevier, 1992.

[115] Martin A Fischler and Robert A Elschlager. The representation and matching of pictorial structures. IEEE Transactions on computers, 100(1):67–92, 1973.

[116] Joseph L Mundy. Object recognition in the geometric era: A retrospective. Toward category-level object recognition, pages 3–28, 2006.

[117] Jean Ponce, Martial Hebert, Cordelia Schmid, and Andrew Zisserman. Toward category-level object recognition, volume 4170. Springer, 2007.

[118] Henry A Rowley, Shumeet Baluja, and Takeo Kanade. Neural network-based face detection. IEEE Transactions on pattern analysis and machine intelligence, 20(1):23–38, 1998.

[119] Edgar Osuna, Robert Freund, and Federico Girosit. Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition, pages 130–136. IEEE, 1997.

[120] Rong Xiao, Long Zhu, and Hong-Jiang Zhang. Boosting chain learning for object detection. In Proceedings Ninth IEEE International Conference on Computer Vision, pages 709–715. IEEE, 2003.

[121] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, volume 1, pages I–I. Ieee, 2001.

[122] Li Liu, Wanli Ouyang, Xiaogang Wang, Paul Fieguth, Jie Chen, Xinwang Liu, and Matti Pietikäinen. Deep learning for generic object detection: A survey. International journal of computer vision, 128(2):261–318, 2020.

[123] Hiroshi Murase and Shree K Nayar. Visual learning and recognition of 3-d objects from appearance. International journal of computer vision, 14(1):5–24, 1995.

[124] David G Lowe. Object recognition from local scale-invariant features. In Proceedings of the seventh IEEE international conference on computer vision, volume 2, pages 1150–1157. Ieee, 1999.

[125] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. IEEE transactions on pattern analysis and machine intelligence, 27(10):1615–1630, 2005.

[126] Serge Belongie, Jitendra Malik, and Jan Puzicha. Shape matching and object recognition using shape contexts. IEEE transactions on pattern analysis and machine intelligence, 24(4):509–522, 2002.

[127] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence, 24(7):971–987, 2002.

[128] Oncel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. In European conference on computer vision, pages 589–600. Springer, 2006.

[129] Josef Sivic and Andrew Zisserman. Video google: A text retrieval approach to object matching in videos. In Computer Vision, IEEE International Conference on, volume 3, pages 1470–1470. IEEE Computer Society, 2003.

[130] Gabriella Csurka, Christopher Dance, Lixin Fan, Jutta Willamowski, and Cédric Bray. Visual categorization with bags of keypoints. In Workshop on statistical learning in computer vision, ECCV, volume 1, pages 1–2. Prague, 2004.

[131] Paul Viola and Michael J Jones. Robust real-time face detection. International journal of computer vision, 57(2):137–154, 2004.

[132] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[133] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013.

[134] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[135] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28:91–99, 2015.

[136] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd: Single shot multibox detector. In European conference on computer vision, pages 21–37. Springer, 2016.

[137] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[138] Pedro Felzenszwalb, David McAllester, and Deva Ramanan. A discriminative-

ly trained, multiscale, deformable part model. In 2008 IEEE conference on computer vision and pattern recognition, pages 1–8. Ieee, 2008.

[139] Pedro F Felzenszwalb, Ross B Girshick, and David McAllester. Cascade object detection with deformable part models. In 2010 IEEE Computer society conference on computer vision and pattern recognition, pages 2241–2248. Ieee, 2010.

[140] Tomasz Malisiewicz, Abhinav Gupta, and Alexei A Efros. Ensemble of exemplar-svms for object detection and beyond. In 2011 International conference on computer vision, pages 89–96. IEEE, 2011.

[141] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. IEEE transactions on pattern analysis and machine intelligence, 32(9):1627–1645, 2009.

[142] Ross Brook Girshick. From rigid templates to grammars: Object detection with structured models. The University of Chicago, 2012.

[143] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. International journal of computer vision, 104(2):154–171, 2013.

[144] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 7263–7271, 2017.

[145] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767, 2018.

[146] Shiyi Xing, Jianguo Ju, and Jinsheng Xing. Research on hot-rolling steel products quality control based on bp neural network inverse model. Neural Computing and Applications, 31(5):1577–1584, 2019.

[147] ZM Cai, HC Ji, WC Pei, XM Huang, WD Li, and YM Li. Constitutive model of

3cr23ni8mn3n heat-resistant steel based on back propagation (bp) neural network (nn). Metalurgija, 58(3-4):191–195, 2019.

[148] H Ahmadi, HR Rezaei Ashtiani, and M Heidari. A comparative study of phenomenological, physically-based and artificial neural network models to predict the hot flow behavior of api 5ct-l80 steel. Materials Today Communications, 25:101528, 2020.

[149] Qiang Li, Dongxu Wang, Mingyu Zhao, Minghao Yang, Jianfeng Tang, and Kai Zhou. Modeling the corrosion rate of carbon steel in carbonated mixtures of mdea-based solutions using artificial neural network. Process Safety and Environmental Protection, 147:300–310, 2021.

[150] Ling Qiao, Yong Deng, Yuan Wang, and Jingchuan Zhu. A comparative study on arrhenius equations and bp neural network models to predict hot deformation behaviors of a hypereutectoid steel. IEEE Access, 8:68083–68090, 2020.

[151] Qiang Li, Gang Cao, Jiang Li, and Ning Wang. Process estimated temperature model of molten steel in lf based on bp neural network combined with expert system. In Applied Mechanics and Materials, volume 48, pages 853–857. Trans Tech Publ, 2011.

[152] Xunqian Xu, Yuwen Gu, Wei Huang, Dakai Chen, Chen Zhang, and Xiao Yang. Structural optimization of steelepoxy asphalt pavement based on orthogonal design and gabp algorithm. Crystals, 11(4):417, 2021.

[153] Xianlin Wang, Yuqing Liu, and Haohui Xin. Bond strength prediction of concrete-encased steel structures using hybrid machine learning method. In Structures, volume 32, pages 2279–2292. Elsevier, 2021.

[154] Xiaofeng Yue, Guoyuan Ma, Fuqiuxuan Liu, and Xueliang Gao. Research on image classification method of strip steel surface defects based on improved bat algorithm optimized bp neural network. Journal of Intelligent & Fuzzy Systems, (Preprint):1–13, 2021.

[155] Ke Xu, Yang Xu, Peng Zhou, and Lei Wang. Application of rnamlet to surface

defect identification of steels. Optics and Lasers in Engineering, 105:110–117, 2018.

[156] Guizhong Fu, Peize Sun, Wenbin Zhu, Jiangxin Yang, Yanlong Cao, Michael Ying Yang, and Yanpeng Cao. A deep-learning-based approach for fast and robust steel surface defects classification. Optics and Lasers in Engineering, 121:397–405, 2019.

[157] Yan Liu, Jiahui Geng, Zhenfeng Su, Weicun Zhang, and Jiangyun Li. Real-time classification of steel strip surface defects based on deep cnns. In Proceedings of 2018 Chinese Intelligent Systems Conference, pages 257–266. Springer, 2019.

[158] Jiawei Zhang, Heying Wang, Ying Tian, and Kun Liu. An accurate fuzzy measure-based detection method for various types of defects on strip steel surfaces. Computers in Industry, 122:103231, 2020.

[159] Jie Liu, Kechen Song, Mingzheng Feng, Yunhui Yan, Zhibiao Tu, and Liu Zhu. Semi-supervised anomaly detection with dual prototypes autoencoder for industrial surface inspection. Optics and Lasers in Engineering, 136:106324, 2021.

[160] Yu He, Kechen Song, Qinggang Meng, and Yunhui Yan. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 69(4):1493–1504, 2019.

[161] Hongwen Dong, Kechen Song, Yu He, Jing Xu, Yunhui Yan, and Qinggang Meng. Pga-net: Pyramid feature fusion and global context attention network for automated surface defect detection. IEEE Transactions on Industrial Informatics, 16(12):7448–7458, 2019.

[162] Yu He, Kechen Song, Hongwen Dong, and Yunhui Yan. Semi-supervised defect classification of steel surface based on multi-training and generative adversarial network. Optics and Lasers in Engineering, 122:294–302, 2019.

[163] Jiaqiao Zhang, Xin Kang, Hongjun Ni, and Fuji Ren. Surface defect detection of steel strips based on classification priority yolov3-dense network. Ironmaking & Steelmaking, 48(5):547–558, 2021.

[164] Xinglong Feng, Xianwen Gao, and Ling Luo. X-sdd: A new benchmark for hot rolled steel strip surface defects detection. Symmetry, 13(4):706, 2021.

[165] Praveen Damacharla, Achuth Rao, Jordan Ringenberg, and Ahmad Y Javaid. Tlu-net: A deep learning approach for automatic steel surface defect detection. In 2021 International Conference on Applied Artificial Intelligence (ICAPAI), pages 1–6. IEEE, 2021.

[166] Guo-hua Jiao, Guang-liang Wu, and Yan-hui Sun. Reason and control of edge cracking of csp hot-rolled strip. IRON AND STEEL-BEIJING-, 41(6):27, 2006.

[167] Hermann Riedel, Florence Andrieux, Tom Walde, and Kai-Friedrich Karhausen. The formation of edge cracks during rolling of metal sheet. steel research international, 78(10-11):818–824, 2007.

[168] Zhijie Jiao, Xiaotong Meng, Chunyu He, and Jun Wang. Data communication of process control system for plate mill. In 2016 5th International Conference on Measurement, Instrumentation and Automation (ICMIA 2016). Atlantis Press, 2016.

[169] Aitor Moreno, Gorka Velez, Aitor Ardanza, Iñigo Barandiaran, Álvaro Ruíz de Infante, and Raúl Chopitea. Virtualisation process of a sheet metal punching machine within the industry 4.0 vision. International Journal on Interactive Design and Manufacturing (IJIDeM), 11(2):365–373, 2017.

[170] Serena Zheng, Noah Apthorpe, Marshini Chetty, and Nick Feamster. User perceptions of smart home iot privacy. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–20, 2018.

[171] Abdul-Rahman Al-Ali, Imran A Zualkernan, Mohammed Rashid, Ragini Gupta, and Mazin AliKarar. A smart home energy management system using iot and big data analytics approach. IEEE Transactions on Consumer Electronics, 63(4):426–434, 2017.

[172] Tie Qiu, Zhao Zhao, Tong Zhang, Chen Chen, and CL Philip Chen. Underwater internet of things in smart ocean: System architecture and open issues. IEEE transactions on industrial informatics, 16(7):4297–4307, 2019.

[173] Athman Bouguettaya, Munindar Singh, Michael Huhns, Quan Z Sheng, Hai Dong, Qi Yu, Azadeh Ghari Neiat, Sajib Mistry, Boualem Benatallah, Brahim Medjahed, et al. A service computing manifesto: the next 10 years. Communications of the ACM, 60(4):64–72, 2017.

[174] Yang Lu. Industry 4.0: A survey on technologies, applications and open research issues. Journal of industrial information integration, 6:1–10, 2017.

[175] Wattana Viriyasitavat, Li Da Xu, Zhuming Bi, and Assadaporn Sapsomboon. Blockchain-based business process management (bpm) framework for service composition in industry 4.0. Journal of Intelligent Manufacturing, 31(7):1737–1748, 2020.

[176] Fei Tao, Ying Cheng, Li Da Xu, Lin Zhang, and Bo Hu Li. Cciot-cmfg: cloud computing and internet of things-based cloud manufacturing service system. IEEE Transactions on industrial informatics, 10(2):1435–1442, 2014.

[177] Zhaohui Wu, Jianwei Yin, Shuiguang Deng, Jian Wu, Ying Li, and Liang Chen. Modern service industry and crossover services: Development and trends in china. IEEE Transactions on Services Computing, 9(5):664–671, 2015.

[178] Lanfranco Urli and Sergio Murgia. Use of ethernet communications for real-time control systems in the metals industry. In 2011 IEEE International Conference on Automation Science and Engineering, pages 6–11. IEEE, 2011.

[179] Zhilun Cheng, Yuqing Fan, and Riaz Ahmad. Implementation of order oriented mes in iron and steel industry. In 2006 4th IEEE International Conference on Industrial Informatics, pages 504–508. IEEE, 2006.

[180] Heeje Lee and Joongyoon Lee. Development concepts of smart service system-based smart factory (4sf). In INCOSE international symposium, volume 28, pages 1153–1169. Wiley Online Library, 2018.

[181] Gianpiero Brunetti. Architecture and implementation of a mes system in a large scale steel plant: Severstal cherepovets success story. In 2016 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pages 1077–1081. IEEE, 2016.

[182] Francisco Almada-Lobo. The industry 4.0 revolution and the future of manufacturing execution systems (mes). Journal of innovation management, 3(4):16–21, 2015.

[183] Shifei Ding, Chunyang Su, and Junzhao Yu. An optimizing bp neural network algorithm based on genetic algorithm. Artificial intelligence review, 36(2):153–162, 2011.

[184] LI Xiao Feng, XU Jiu Ping, and HE Chang Zheng. The establishment of self-adapting algorithm ofbp neural network and its application. Systems Engineertheory &Practice, 5(24):1–8, 2004.

[185] Evangelos C Alexopoulos. Introduction to multivariate regression analysis. Hippokratia, 14(Suppl 1):23, 2010.

[186] Zhiqiang Ge, Zhihuan Song, Steven X Ding, and Biao Huang. Data mining and analytics in the process industry: The role of machine learning. Ieee Access, 5:20590–20616, 2017.

[187] Zhi-Hua Zhou, Jianxin Wu, and Wei Tang. Ensembling neural networks: many could be better than all. Artificial intelligence, 137(1-2):239–263, 2002.

[188] Anders Krogh, Jesper Vedelsby, et al. Neural network ensembles, cross validation, and active learning. Advances in neural information processing systems, 7:231–238, 1995.

[189] Dipak Laha, Ye Ren, and Ponnuthurai N Suganthan. Modeling of steelmaking process with effective machine learning techniques. Expert systems with applications, 42(10):4687–4696, 2015.

[190] Yun Bai, Zhenzhong Sun, Jun Deng, Lin Li, Jianyu Long, and Chuan Li. Manufacturing quality prediction using intelligent learning approaches: A comparative study. Sustainability, 10(1):85, 2018.