metals

Article

Hybrid-Input FCN-CNN-SE for Industrial Applications: Classification of Longitudinal Cracks during Continuous Casting

Davi Alberto Sala [1,*], Andy Van Yperen-De Deyne [2], Erik Mannens [3] and Azarakhsh Jalalvand [4]

[1] Faculty of Engineering and Architecture, imec–UGhent–IDLab, 9000 Gent, Belgium
[2] Digital Solutions & Innovation, ArcelorMittal Belgium, 9042 Gent, Belgium
[3] Faculty of Sciences, Department of Computer Science, imec–UAntwerp–IDLab, 2000 Antwerp, Belgium; erik.mannens@uantwerpen.be
[4] Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, USA; azarakhsh.jalalvand@princeton.edu
* Correspondence: davialberto.sala@ugent.be

Abstract: In the presented research, machine learning methods were applied to the prediction of longitudinal cracks in steel slabs during continuous casting. We employ a deep learning approach to process 68 thermocouple signals as a multivariate time series (MTS) along with 32 static features, which encompass both chemical composition and process information. Our deep learning approach integrates two distinct parallel modules, followed by an aggregation block; a Convolutional Neural Network (CNN) processes the thermocouple MTS, while in parallel, the static data undergo processing via a Fully Connected Network (FCN). To enhance the performance of the CNN, we incorporate two Squeeze and Excitation (SE) blocks, which act as an attention mechanism across different channels. By integrating chemical information with MTS in the detection system, we improve the performance of defect detection by 15% relatively.

Keywords: continuous caster; data-driven prediction models; multivariate time series; neural networks; longitudinal crack

Citation: Sala, D.A.; Van Yperen-De Deyne, A.; Mannens, E.; Jalalvand A. Hybrid-Input FCN-CNN-SE for Industrial Applications: Classification of Longitudinal Cracks during Continuous Casting. *Metals* **2023**, *13*, 1699. https://doi.org/10.3390/met13101699

Academic Editor: Noé Cheung

Received: 6 September 2023
Revised: 27 September 2023
Accepted: 3 October 2023
Published: 6 October 2023

1. Introduction

For many years, the formation of defects during the continuous casting of steel slabs has been the subject of extensive research, and our comprehension of their genesis continues to deepen. The ideal scenario is to pre-emptively prevent defects by establishing standard operating conditions and control mechanisms. However, due to unpredictable or overlooked factors, this is not always feasible. As a result, it is crucial to oversee the process to identify quality issues and implement timely corrective measures, ensuring further material is not compromised. To ensure rigorous quality management, the procedure is observed using an array of thermocouple sensors. These sensors' readings are integrated into a control strategy, which then modulates parameters like casting speeds, flow regulation, electromagnetic forces, and nozzle immersion using mathematical models and expert systems. Some steel manufacturing facilities have begun to incorporate these systems to a certain degree [1]. Their efficacy hinges on the comprehensive understanding and integration of the interplay between the sensor readings, control interventions, and the defects within the control software algorithms.

Alongside these computational models, which are predominantly derived from a physical comprehension of the process, there exist clusters of stored data recorded over years from sensor logs. Historically, in an industrial setting, these data have been primarily referenced for historical analysis. However, with the rapidly increasing volume of stored data, there is a growing inclination towards employing big-data methodologies in industrial scenarios [2]. This paper unveils the outcomes of a collaborative research endeavor aimed at developing an automated, data-centric detection system for longitudinal cracks in steel slabs during continuous casting, conducted at the ArcelorMittal Belgium facility in Ghent.

Given that steel slabs frequently serve as precursors for sheet steel rolling, surface imperfections like cracks (as illustrated in Figure 1) can lead to extended segments of flawed products. As a result, inspecting steel slabs before rolling is vital in the industry, as it pre-empts potential issues in subsequent stages. However, most inspection systems within the area are still manually operated and certain steel grades with lower carbon contents cannot be visually inspected for longitudinal cracks since they are not visible on the surface [3]. A good detection system should efficiently flag potentially problematic cracks while keeping the number of false positives (non-cracked slabs identified as cracked) at a minimum.

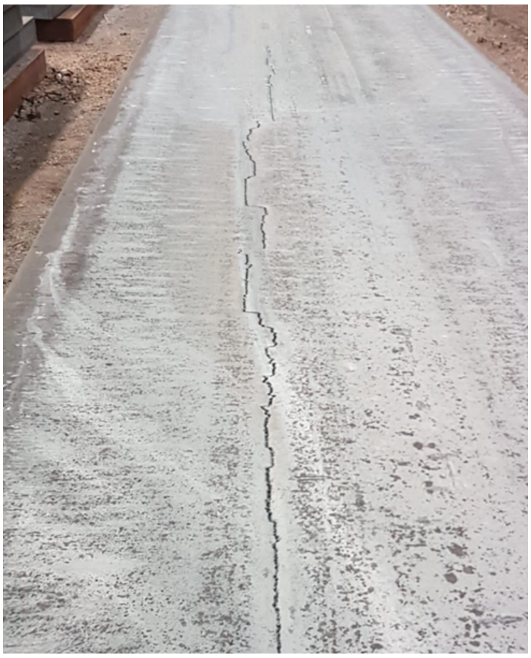

Figure 1. Example of a visible longitudinal crack.

The focus of this work is a data-driven approach for longitudinal crack detection using both thermocouple information alongside chemical composition of the steel and process features (e.g., casted width and type of nozzle and powder). We apply a hybrid Deep Learning model where the multivariate time series are processed by a one-dimensional Convolutional Neural Network (CNN) alongside Squeeze and Excitation blocks for channel-wise attention; in parallel, a Fully Connected Network is used for processing the chemical and process information.

As a baseline, we use the current model in production alongside a feature extraction technique based on classic machine learning; the development of such a baseline, the initial results, and a comparison with other feature extraction approaches were published in [4]. Based on previous results, we chose Gradient Boosted Regression Trees (GBRTs) as our baseline model.

Longitudinal Cracks

A segment of the continuous casting procedure is depicted in Figure 2: molten steel is channeled from the ladle (1) into the tundish (2). From there, it descends (3) via a ceramic nozzle, entering the mold (4). Numerous thermocouples are strategically positioned on the mold’s surface (5). Within the mold, the molten steel circulates in the liquid pool, carrying with it superheat, inclusion particles, and turbulence, factors that can influence the top surface’s level [3]. The molten steel then begins to solidify against the mold walls, forming a slender solid shell. This shell is consistently extracted from the mold's base at a casting speed synchronized with the inflow of the incoming metal. The nozzle's flow is propelled by gravity, driven by the pressure differential between the tundish's liquid level and the mold's top free surface. This flow rate is meticulously regulated to ensure a consistent liquid level within the mold. Mould powder is introduced to the top surface, where it melts, forming a protective layer that shields the molten steel from both thermal and chemical interactions with the environment above. This flux also permeates the space between the mold and the forming shell, serving as a lubricant to prevent adhesion and modulating heat transfer across this gap. The solidification process is initiated at the meniscus, the juncture where molten steel, liquid flux, and the mold wall converge. Surface imperfections in the forming shell originate at this point, influencing heat transfer further down the mold. Longitudinal cracks are one of the multiple types of surface defects that can happen during this process; in extreme scenarios, a localized thinning can result in a costly breakout, causing molten steel to spill over the lower sections of the casting machine.

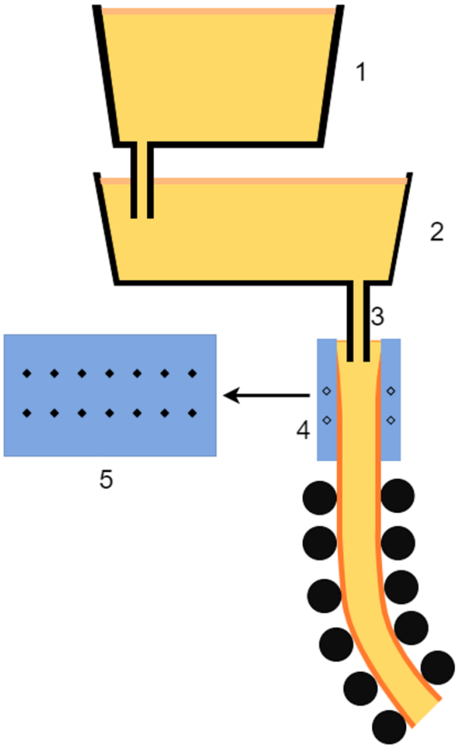

Figure 2. Simplified diagram of the casting process and thermocouple location. The molten steel is poured from the ladle (**1**) into the tundish (**2**), and through a nozzle (**3**) it enters the mold (**4**). The mold walls are equipped with thermocouples (**5**).

Figure 3 shows the time signals of the center-most thermocouples in the mold walls of four different cases: (a) a clear dip on the top row of thermocouples that is propagated on the second row after a few seconds and resulted in a longitudinal crack; (b) signals when a longitudinal crack happened but nothing was detected during casting and was only discovered during hot rolling; (c) an erratic behavior that resulted in a false positive—an alarm was given but no crack was present—and (d) the normal behavior that is observed and expected most of the time.

A comprehensive solution for detecting these irregularities is a rule-based system, exemplified in [5]. In their work, the authors suggest employing a temperature differential thermograph for the identification of slab cracks. A system bearing resemblance to this is already operational at the ArcelorMittal Belgium facility. The primary merit of such a system lies in its transparent, or "white box", architecture, enabling any triggered alarm to be visually verified. While this serves as our initial benchmark, there is room for enhancement by introducing a context-sensitive model that considers the properties of the material set to be cast. This enhancement can be achieved through a feature extraction methodology, which constitutes our secondary baseline [4]. However, in this study, we advocate for the adoption of a deep learning model.

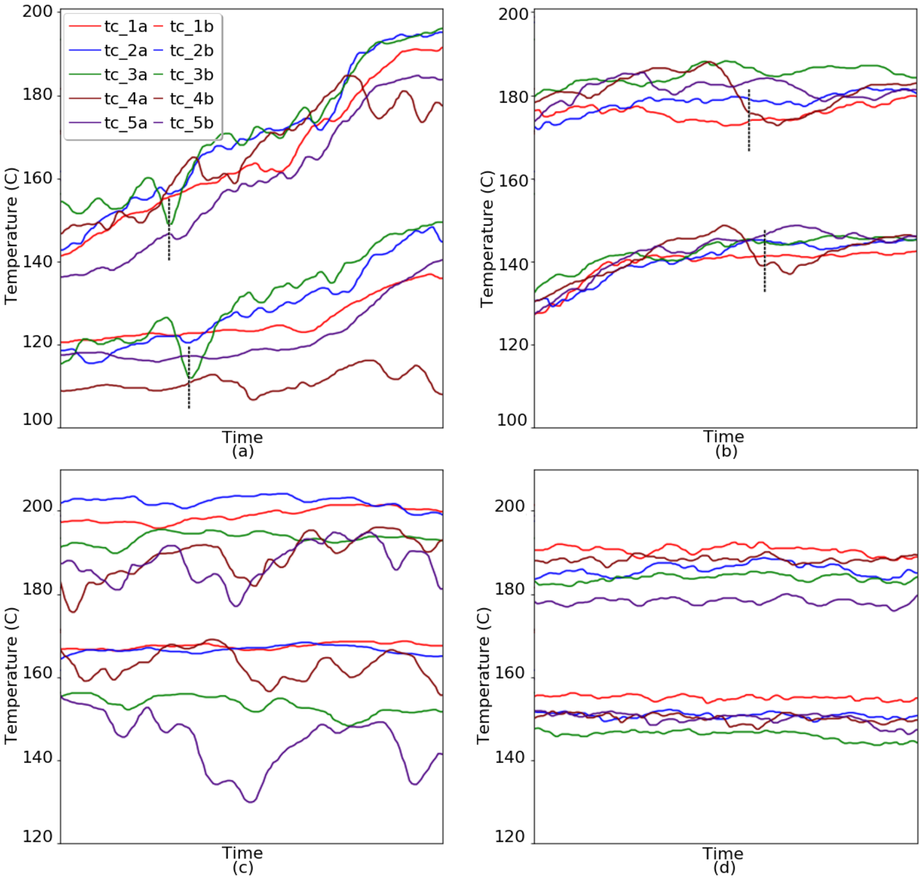

Figure 3. Time series of 10 thermocouple signals. Regarding longitudinal cracks on these samples: (**a**) shows a clear true positive; (**b**) shows an undetected defect; (**c**) a false positive where an alarm was given but no defect was present; and (**d**) normal behavior.

2. The Literature

In the continuous casting of steel, quality issues can also be pinpointed by observing mold signals, such as readings from level sensors, thermocouples embedded in mold walls, friction measurements, and more. Based on these observations, corrective measures can be taken, like reducing the casting speed or conducting subsequent visual inspections of the surface to determine potential downgrades or the need for scarfing. Distinctive thermocouple patterns associated with various defects have been recognized [1], encompassing sticker breakouts, transverse depressions, pronounced oscillation marks, narrow-face bleeds, transverse corner cracks, longitudinal cracks, and mold level variation defects, among others. This study narrows its focus specifically to the issue of longitudinal cracks.

In order to continuously monitor the quality, non-destructive testing methods have been widely developed, such as image analysis [3], eddy currents [6], magnetic powder [7], ultrasound [8], or sulfur prints [9]. Most of these systems require the installation of sensitive sensors in the continuous caster line, which is a harsh environment. Longitudinal cracks are less visible on certain grades of steel (usually slabs casted with low carbon content—also known as soft grades), which renders visual inspection systems ineffective [10].

The authors of [11] present a precision inspection system tailored to autonomously identify cracks in "as-cast" steel slabs. The task of identifying cracks is further complicated by the oxidized layer on the slab's exterior. A custom-designed laser triangulation mechanism has been introduced. Cracks are detected using a combination of morphological detection and an SVM classifier. A different technique for forecasting visual breakouts during mold monitoring is detailed in [12]. This method utilizes temperatures captured by thermocouples inside the mold and combines them with computer vision methodologies. Dubbed as mold temperature rate thermography, this strategy aids in pinpointing and extracting distinct characteristics of temperature zones, including the pace of temperature fluctuations over time, their geometric shape, and propagation speed.

Another non-destructive approach, as proposed in this paper, is to work with data-driven models based on process parameters already captured for process steering. Such data-driven models have proven to be efficient in several sectors of steel production, like end-point prediction at the basic oxygen furnace [13] and fault detection at the hot strip mill [14]. Currently, the available literature on longitudinal crack detection lacks contributions related to big-data applications and is limited to process knowledge systems.

The impact of chemical components, primarily carbon and sulfur, on longitudinal surface cracks has been comprehensively studied. As demonstrated by the researchers in [15], the carbon concentration at which the longitudinal surface cracking peaks diminishes as the sulfur content rises. Conversely, for a fixed carbon concentration, the likelihood of surface cracking escalates with an increase in sulfur content. Separate research [16] delves into the influence of the mold powder utilized, particularly its SiO_2 concentration, on the surface quality of slabs with a high aluminum composition. This study also reveals that, under identical processing conditions, wider slabs require a more extended period for the powder to achieve equilibrium. This pinpoints the importance of aggregating both chemical and process information into the input space.

Utilizing the thermocouple information, a temperature differential thermograph has been developed to identify slab cracks [5]. To pinpoint the anomalous temperature zones resulting from longitudinal cracks, potentially problematic regions are isolated and segmented using computer image processing techniques.

In the study presented in [17], a model is presented for longitudinal crack detection, leveraging principal component analysis (PCA) and support vector machine (SVM). The temperature patterns associated with the longitudinal crack defect are isolated. These encompass the standard casting temperature exhibiting minor and major oscillations, as well as the specific temperature of the longitudinal crack. PCA is employed to diminish the dimensionality of these features and an SVM is utilized as a classifier.

Novel machine learning techniques are typically introduced to industrial applications only after they have gathered substantial validation in academic research. Time-series data find utility across various industrial domains, including quality assurance, soft sensing, and predictive maintenance. Although numerous advanced methodologies have been employed for longitudinal crack detection, the primary focus has been on temperature and rolling attributes, without incorporating chemical components into the analysis. Moreover, no study was found where several years of historical data were used for analysis.

2.1. Contribution

As stated in the previous section, inspection of steel surfaces can be challenging; in particular, rule-based systems that are only based on thermocouple information may present a high false positive rate while still missing defects that do not present a clear visual sign on the column of thermocouples. For instance, a crack that is located between columns of thermocouples generates a different signal pattern than one located exactly at a thermocouple column. Moreover, it is clear from the literature that chemical composition, powder, and other process features have a high impact on defect occurrence.

We proposed improving upon the notion that thermocouples and chemical and process information are of importance for thte detection of longitudinal cracks on casted steel slabs. In this work, we present a hybrid data-driven strategy for crack detection. The model takes as input the multivariate time series from 68 thermocouples, along with chemical composition and process information. Our models will be trained and validated on three years of data, respecting temporal order. For baselines, we used the current model in production and a Gradient Boosted Regression Trees (GBRTs) model trained on extracted features from the input data. Furthermore, we analyze the impact of adding chemical and process information to the input space. The varying effect of chemical composition, as observed through different evaluation metrics, underscores the nuanced impact it has on prediction accuracy and the potential for uncovering more complex relationships between process variables and output quality.

2.2. The CNN-SE Network

In this section, we provide an overview of the Convolutional Neural Network (CNN) framework. Additionally, we discuss the Squeeze and Excitation (SE) blocks, as detailed in [18], adjusted for one-dimensional convolutions to align with our time-series dataset.

The CNN block is represented by the convolution transformation $\mathbf{F}_{tr} : \mathbf{X} \rightarrow \mathbf{U}$. This transformation maps an input $\mathbf{X} \in \mathbb{R}^{T' \times C'}$ to feature maps $\mathbf{U} \in \mathbb{R}^{T \times C}$. Here, T and T' denote the time dimensions, while C and C' represent the number of channels,

$$\mathbf{u}_c = \mathbf{v}_c * \mathbf{X} = \sum_{s=1}^{C'} \mathbf{v}_c^s * \mathbf{x}^s \tag{1}$$

In this context, $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_C]$ is derived from the convolution of the input vector $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{C'}]$ with spatial kernels $\mathbf{v}_c = [\mathbf{v}_c^1, \mathbf{v}_c^2, \ldots, \mathbf{v}_c^{C'}]$. Here, $\mathbf{v}_c^s$ denotes a spatial kernel that represents an individual channel of $\mathbf{v}_c$, interacting with the corresponding channel of $\mathbf{X}$.

The SE architecture is illustrated in Figure 4; the operation takes as input the channels originating from the convolutional transformation $\mathbf{F}_{tr}$. Overlaid on this signal are two distinct processes: Squeeze, where information is globally aggregated to distil essential features, and Excitation, which shows the dynamic modulation of these features, emphasizing certain channels of the input for enhanced representation and understanding. This representation is then used to scale the original input.

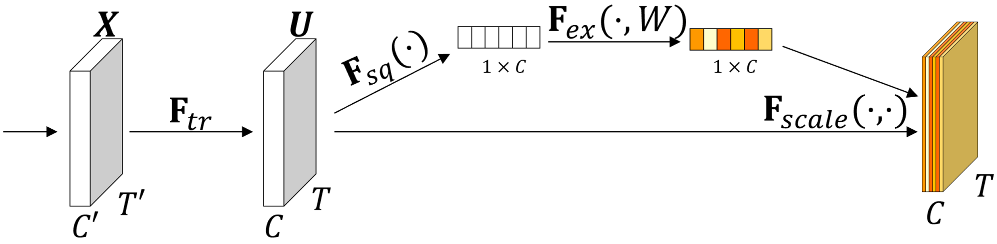

Figure 4. Computation of the Squeeze and Excitation block following a CNN.

After the convolutional unit, the SE block starts with the Squeeze function, which employs global average pooling to produce channel-wise statistics. Statistic information $\mathbf{z} \in \mathbb{R}^C$ is derived by squeezing $\mathbf{U}$ across its temporal dimension T. The c-th element of $\mathbf{z}$ is computed as:

$$z_c = \mathbf{F}_{sq}(\mathbf{u}_c) = \frac{1}{T} \sum_{t=1}^{T} u_c(t) \tag{2}$$

Following the Squeeze computation, the aggregated information undergoes an Excite operation. This step is designed to discern and capture dependencies specific to each channel.

$$\mathbf{s} = \mathbf{F}_{ex}(\mathbf{z}, \mathbf{W}) = \sigma(\mathbf{W}_2 \delta(\mathbf{W}_1 \mathbf{z})) \tag{3}$$

where weights $\mathbf{W}_1 \in \mathbb{R}^{\frac{C}{r} \times C}$ and $\mathbf{W}_2 \in \mathbb{R}^{C \times \frac{C}{r}}$ are parameters that can be trained. Here, σ denotes the sigmoid activation function, while δ represents the gating function ReLU. The factor r is introduced to reduce dimensionality, serving to curb the model's complexity. This factor can be fine-tuned as a hyper-parameter. Research presented in [18] indicates that setting $r = 16$ strikes a balance between model complexity and accuracy, especially when layers consist of 128 to 512 filters. The output block's c-th channel undergoes a rescaling process, defined as:

$$\tilde{\mathbf{x}}_c = \mathbf{F}_{scale}(\mathbf{u}_c, s_c) = s_c \mathbf{u}_c, \tag{4}$$

where $\tilde{\mathbf{X}} = [\tilde{\mathbf{x}}_1, \tilde{\mathbf{x}}_2, \ldots, \tilde{\mathbf{x}}_C]$, and $\mathbf{F}_{scale}(\mathbf{u}_c, s_c)$ refers to channel-wise multiplication between the scalar s_c and feature map $\mathbf{u}_c \in \mathbb{R}^T$.

The SE block can be incorporated into conventional convolutional frameworks, enabling adaptive adjustments to the input feature maps. This mechanism is similar to a self-attention module, where the input values are used to determine their own significance. Studies have demonstrated that by integrating SE blocks into ResNet-50 [18], a performance approaching that of ResNet-101 can be achieved. This is impressive for a model requiring only half of the computational costs. The quantity of additional parameters P needed to learn these SE maps can be determined as follows:

$$P = \frac{2}{r} \sum_{s=1}^{S} R_s \cdot G_s^2 \tag{5}$$

where S represents the total number of stages, with a stage being defined as a sequence of consecutive layers sharing the same kernel size. R_s indicates the count of blocks that are repeated within stage s, while G_s indicates the number of feature maps associated with stage s.

3. Methodology

Our approach comprises three stages: Firstly, static and temporal data from multiple databases are joined and indexed for each slab (sample), then corrupted samples are filtered out. In the second stage, an extensive number of features from each signal are extracted and the most relevant are selected through model selection. In the last step, classification models are trained and evaluated on different levels of imbalance, shuffling or maintaining time order. The impact of adding the process information and chemical composition in the dataset is also evaluated. In what follows, we describe the model architectures, dataset, and evaluation metrics used.

Two baselines are considered, namely, (1) an analytical model based on the physical behavior of the process which is currently being used in the factory and (2) a GBRT method trained on the static data and features extracted from the multivariate time series. The first one is purely the detections from a differential model similar to the one presented in [5], based solely on historical data. The second is an ensemble model following the methodology proposed in [4], where a GBRT model was trained using automated time series feature extraction and selection along with chemical features. The GBRT model combines multiple decision trees to make predictions. It leverages the strengths of decision trees, such as non-linearity and ability to handle complex relationships, while mitigating their weaknesses by boosting the performance of individual trees through iterative training.

Our proposed CNN-SE-FCN model is depicted in Figure 5, where a fully convolutional block processes the temporal data and an FCN handles the static features. The outputs of both blocks are then concatenated and passed to the last activation layer.

Regarding the FCN model, two input tensors are defined: one for the time series and another for the static data. The time series tensor adopts the shape of (N, T, M). Here, N represents the maximum count of samples (batches) within the dataset, T is the number of time steps, and M corresponds to the count of time signals in our MTS dataset. Conversely, the tensor for static data is structured as (N, K), with K denoting the cumulative number of static features employed.

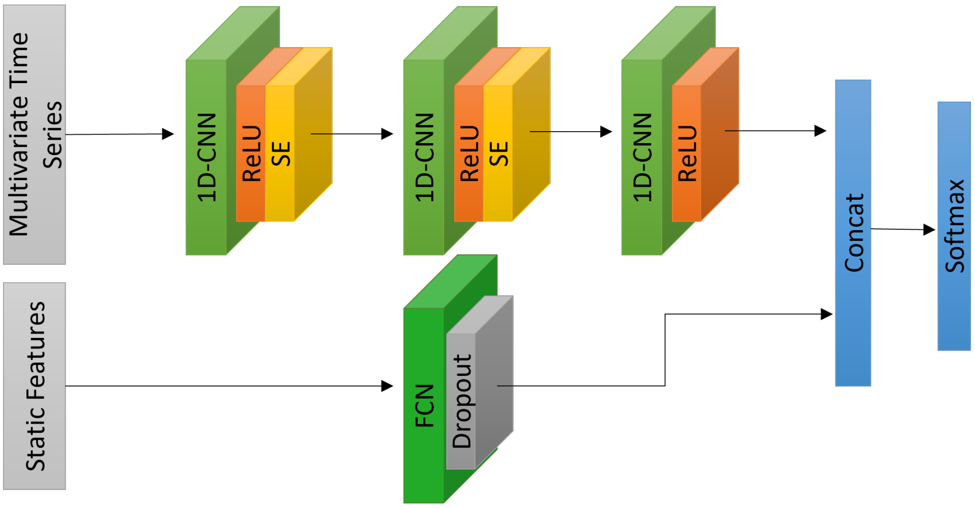

Figure 5. In the FCN-CNN-SE model, both static and time-series data are processed in parallel. Once processed, the outputs from both blocks are concatenated. This combined output then passes through a final activation layer.

In our architectural design, both the time series and static data are processed concurrently by the CNN and FCN blocks, respectively. The convolutional block comprises three convolutional layers, serving as feature extractors. These layers have kernel sizes of 8, 5, and 3, with the corresponding number of filters set at 128, 256, and 128. Each of these layers is succeeded by batch normalization and a ReLU activation function. Additionally, the initial two blocks end with SE blocks, where the reduction ratio is designated as $r = 16$, in line with the recommendations from the original paper [18]. This inclusion augments the model's complexity by $P = \frac{2}{16}(128^2 + 256^2) = 10,240$ parameters, which translates to a roughly 5% relative increase in the trainable parameters of our model. The SE mechanism bolsters performance on multivariate datasets, given that each feature map can influence the outcome to varying extents. This autonomously acquired form of channel-specific attention integrates the inter-correlation data among multiple variables.

The FCN block consists of a hidden layer featuring 64 neurons, accompanied by a dropout rate of 50% to mitigate overfitting. This is succeeded by a ReLU activation layer. Outputs from both the CNN and FCN blocks are subsequently concatenated and fed to the concluding activation layer, which employs a sigmoid activation function for our classification task.

3.1. Data Preparation

The continuous casting process, also known as strand casting, is an essential phase in steel production, where liquid steel is solidified into a semi-finished slab for subsequent rolling in the hot strip mill. The quality of the final product is strongly dependent on numerous variables, such as the temperature, composition of the liquid steel, and casting speed, among others.

Predictive models for parameters such as temperature and solidification rates are utilized to determine these specifications. These models enable precise control and management of the continuous casting process, thus ensuring that the semi-finished product aligns with the desired quality standards. Based on these predictions, further adjustments can be made during the casting process to achieve the desired properties in the steel slab. The casting phase duration can vary, but is a relatively continuous process, hence the name.

One key aspect of managing the continuous casting process is the collection and analysis of time-series signals, which provide invaluable data about the casting process. In this project, we will focus on the temperature data collected from thermocouple sensors situated around the mold. In addition to this, chemical composition and other process features will be used to further improve the precision of our predictive models.

Time-series signals can be conceptualized as univariate time series, symbolized by $\mathbf{x}$, which is a one-dimensional signal, sampled across a time domain. To be more specific, $\mathbf{x}$ is a sequentially arranged set of real values $[x_1, x_2, \ldots, x_{T'}]$, where T' denotes the comprehensive length of the signal. This sequence typically arises from the output of a sensor that is sampling while overseeing a process. When a process is under the surveillance of multiple sensors, it is characterized as a multivariate time series (MTS) because it encompasses multiple time-dependent variables. An MTS, represented by $\mathbf{X}$, is composed of various univariate time series $[\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{C'}]$, with C' indicating the total number of signals (or channels). The casting procedure is extensively instrumented, with several sensor readings and measurements available throughout its duration. In terms of data preparation, the time signals undergo rescaling to fit within the range of $[0, 1]$, and the static features are normalized to achieve a zero mean and a standard deviation of one (it is noteworthy that both the training and test sets are rescaled and normalized based solely on the training set's values). Figure 3 displays the signals for a specific slab before being rescaled.

The accessible recorded data consist of the latest 85,000 slabs produced. After filtering out corrupted or incomplete entries, the usable dataset comprises approximately $N = 80{,}000$ samples. Each of these samples is equipped with $K = 23$ static features and $M = 68$ thermocouple time signals, each spanning roughly $T = 700$ timesteps, contingent on the casting speed. This dataset was partitioned, with 70% allocated for training and validation, and the remaining 30% reserved for testing. To maintain the chronological sequence, the samples were not randomized prior to this division. This approach mirrors a realistic scenario where a model would be integrated into a production environment to forecast outcomes for subsequent heat batches. It is crucial to highlight that, during training, the training set undergoes shuffling at the start of each epoch.

Regarding the static data, they are composed of process variables and chemical compositions. The chemical composition pertains to the concentrations of chemicals identified prior to the pouring of steel into the tundish, with 18 features such as C, Mn, S, P, Cu, and Al, among others. On the other hand, the remaining five static features from the process variables consist of categorical features like the type of powder, nozzle, and stopper, as well as numerical features such as casting width and the average speed recorded over the last 3 min.

Time stamp labels were created at the time-of-crack. These labels can be of two different origins; if detected at the continuous caster by the current alarm system, a time stamp is recorded, similar to how they are marked in Figure 3a,b. The slab will later be visually verified as to whether if the defect exists or not. The defects that are not detected by the system will be automatically identified after the slab is hot-rolled. The longitudinal cracks are then measured by an automated system and the time-of-crack for each defect is estimated from these measurements.

3.2. Windowing

The process of the sliding window is shown in Figure 6. The method accumulates the historical time series data over L time steps (or data points) [19], data contained inside this window are used as input. The window is moved with a step s in time for the next prediction. The process will be continued until time series data are exhausted. The final configuration used was a window size of $L = 150$ with steps of $s = 30$. The window size of 150 was empirically chosen as both the Gradient Boosted Regression Trees (GBRTs) and our deep learning approach exhibited no significant gains when the window size was increased to 200. Conversely, a reduction in performance was observed when the window size was decreased to 100 steps. Given that our signal frequency is 1 Hz, a window of 150 time steps encapsulates two and a half minutes of data, providing sufficient time to capture the full anomalous signal associated with a longitudinal crack. These anomalies are typically shorter than 90 s in 98% of cases.

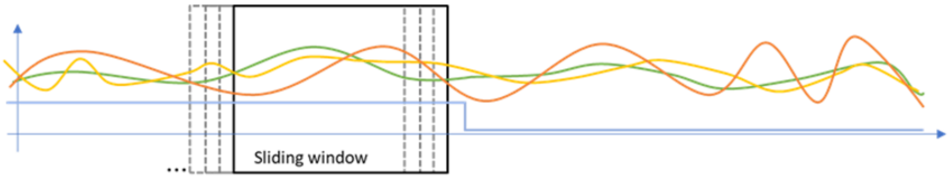

Figure 6. Using the sliding window method to segment the original Multivariate Time Series into fixed-length segments. Each data segment corresponds to the time span highlighted by the solid black box, with intervals between sampled time steps represented by the dotted black boxes.

Due to the large imbalance in our dataset, where the occurrence of cracks is approximately 0.8%, we adjusted the sampling for our training set. Specifically, we down-sampled the non-cracked sample windows by half and over-sampled the cracked ones by reducing the sliding window step to $s = 10$ on the windows containing longitudinal cracks. Following this sampling adjustment, the imbalance in our training set was reduced to about 3%, while maintaining the original imbalance of approximately 0.8% in the test set.

3.3. Evaluation Setup

We evaluated our proposed multivariate time series model against two baselines based on numerical features, reflecting a more traditional approach. The initial baseline is the extant mathematical model employed in production, which relies solely on historical data. For our second baseline, in order to evaluate the impact of the feature learning component (specifically, the CNN-SE-FCN), we also employed a feature extraction technique as detailed in [4]. In this approach, several features are extracted from multivariate time-series signals using TSFresh [20], and the top 500 most pertinent features are used alongside the process and chemical features. A Gradient Boosted Regression Tree (GBRT) is utilized in the classification task.

The F_1 score is the harmonic mean of the precision and recall,

$$F_1 = 2 \times \frac{P \times R}{P + R} \tag{6}$$

where precision (P) is defined as the ratio of true positive results to the sum of true positives and false positives. It represents the accuracy of positive predictions. Recall (R), on the other hand, is the ratio of true positive results to the sum of true positives and false negatives, indicating the proportion of actual positives that were correctly classified. A more encompassing metric, F_β, incorporates a positive real factor β. This factor allows for differential weighting, favoring either precision or recall depending on its value:

$$F_\beta = \frac{(1 + \beta^2)(P \times R)}{(\beta^2 \times P) + R} \tag{7}$$

The F_β score is more interesting in our case, since false negatives have a greater impact on production costs than false positives; for steel grades where a crack can be identified via visual inspection, the cost of a false negative can be 20 times larger, while for grades where no visual inspection is possible, scarfing 10 slabs (all false positives) would cost as much as having one false negative further down the production line. Using internal cost calculations, we have established that $F_\beta = 2.4$ better weights the costs between precision and recall in our case. Therefore, given this context, we will use the F_β score as our primary metric for evaluating model performance. Furthermore, we will be optimizing model hyper-parameters in order to maximize F_β.

We will also evaluate the Receiver Operating Characteristic (ROC) curve, as it assesses the prediction ability our classifiers as the discrimination threshold varies. It plots the true positive rate against the false positive rate. However, in highly imbalanced datasets, where the minority class is significantly less frequent, the ROC curve may be misleading.

This is due to the fact that it treats false positives and false negatives equally, potentially giving an overly optimistic measure of model performance when the number of true negatives outweighs the false positives. In contrast, the precision–recall (PR) curve is more informative for imbalanced dataset scenarios [21]. A high area under the PR curve indicates that the model has good precision and recall simultaneously, providing a more accurate indication of the model's performance on the less frequent class. Therefore, it is often the preferred method for evaluating models in imbalanced data contexts.

4. Results

In this section, we evaluate the described classification models with five different training configurations in order to measure the impact of different training techniques and the influence of different features. The evaluation metric is the $F_\beta = 2.4$ score as discussed in the previous section. Precision–recall and Receiver Operating Characteristic (ROC) [22] curves are also shown for some experiments.

Table 1 presents a comparison of different models on four metrics: precision, recall, F1 score, and F-beta score (with beta = 2.7). The models compared are rule-based, GBRT, CNN-SE-FCN without chemical features (−chem), CNN-FCN without Squeeze and Excitation (−SE), and CNN-SE-FCN with chemical features (+chem) models. The last one is the model with our proposed architecture.

Table 1. Comparison of precision, recall, F_1, and $F_{\beta=2.4}$ scores across rule-based, GBRT, and CNN-SE-FCN models with the following variations: no chemical data (−chem), no Squeeze and Excitation (−SE), and the complete model (+chem).

	Baseline		**CNN-SE-FCN**		
	Rule-Based	**GBRT**	**−chem**	**−SE**	**+chem**
Precision	0.118	0.074	0.070	0.158	0.123
Recall	0.271	0.351	0.303	0.291	0.331
$F1$	0.165	0.122	0.113	0.205	0.179
$F_{\beta=2.4}$	0.228	0.226	0.203	0.259	0.265

Looking at the precision, the highest score is achieved by the CNN-SE-FCN model without SE, with a score of 0.158. This suggests that this model has the highest proportion of true positive results among all the models but has a lower recall when compared to others. In terms of recall, the GBRT model performs the best with a score of 0.351, but shows very low precision.

The F-beta score, our main metric since it reflects production costs, is also highest for the CNN-SE-FCN model with chemical features (+chem), with a score of 0.265. This indicates that when giving more importance to recall, this model outperforms the others. Furthermore, we see a large drop in performance when no chemical features are used (−chem), highlighting the importance of including chemical and process features into time series models for industrial applications.

Figure 7 shows the Receiver Operating Curve; the area under the curve for feature extraction was 0.66, 0.70 for CNN-SE-FCN without chemical features (−chem), 0.72 for CNN-SE-FCN, and 0.71 for CNN-FCN. For lower false positive rates (<0.15), which are desired in our case given the high imbalance, GBRT, CNN, and FCN-CNN show similar performances, while FCN-CNN-SE provides a higher true positive rate. Although the ROC can be quite informative for classification tasks, in cases with a high class imbalance, it can be quite misleading [22]. In our case, it enables a cost analysis, since at the factory the real cost of false positives and false negatives is known; the optimal threshold for a model can be selected using a cost function based on these values.

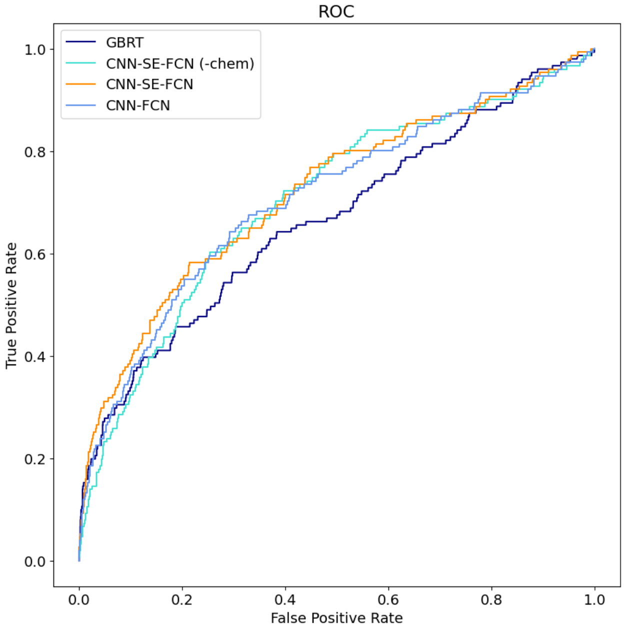

Figure 7. Receiver Operating Characteristic (ROC) curves for models: feature extraction (GBRT), CNN-SE-FCN without chemical features (−chem), CNN-SE-FCN with chemical features, and CNN-FCN (no Squeeze and Excitation block). The curves illustrate the trade-off between the true positive rate and the false positive rate for each model.

Figure 8 shows the precision–recall curve, as it better visualizes the highly imbalanced nature of this classification task. Here, we can verify that the GBRT has a higher precision when the recall is low (<0.1), but fails to maintain the performance at higher recall rates, its performance dropping. Meanwhile, CNN-SE-FCN shows a better performance with a higher recall.

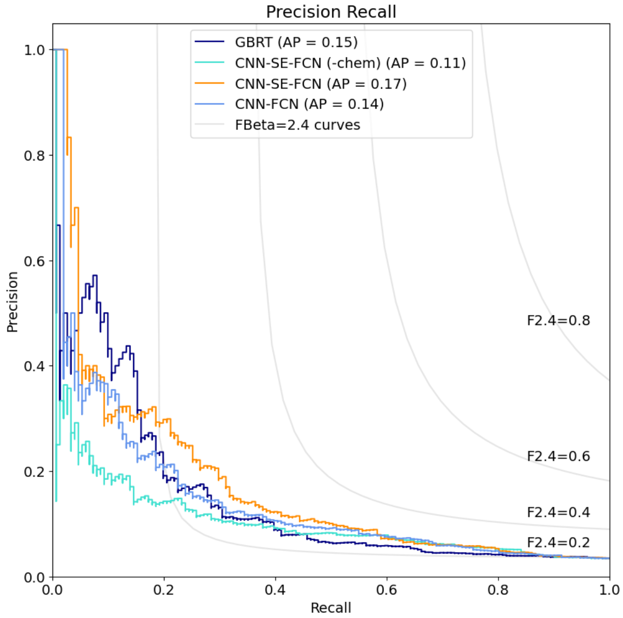

Figure 8. Precision–recall curves for various models: feature extraction, GBRT, CNN-SE-FCN without chemical features (−chem), CNN-SE-FCN with chemical features, and CNN-FCN (no Squeeze and Excitation block). The curves depict the trade-off between precision and recall for each model. The standard lines for $F_\beta = 2.4$ are also shown.

4.1. Processing Time

Processing time is an important factor in industrial systems, specially for quality control during continuous processes. Surface defects are expected to be detected in real time so operators can take measures towards reducing their impact, for instance, by reducing the casting speed when a defect occurs. Computations are intended to be executed within a fraction of a second, given that the system operates with a response time of 1 Hz (sampling interval). We evaluated the preprocessing and inference duration of FCN-CNN-SE using an AMD Ryzen 5 six-core processor, with no reliance on a GPU for the inference process. The deep learning model takes approx. 120 ms for inference while the preprocessing retrieval and preprocessing step of the time series and static data takes on average 270 ms, which is within the time requirements for real-time prediction.

For the feature extraction approach, the preprocessing time was on average 2320 ms due to the high amount of features to be calculated—which includes Fourier transforms and other computationally demanding calculations. The average inference time was 73 ms.

4.2. Robustness

To assess the temporal robustness of our model, we partitioned the training and test data based on the chronological order of batches in production. Figure 9 shows the true positives and false negatives spread across 14,000 samples. For ease of visualization, data points are consolidated into groups of 100 steel slabs each. The dotted line within the figure illustrates the linear trend. It is notable that there is no discernible increase in false negatives over time. Meanwhile, a marginal decline is observed in the linear trend of true positives, this decline can be attributed to the anomalous peak seen at the start of the timeline.

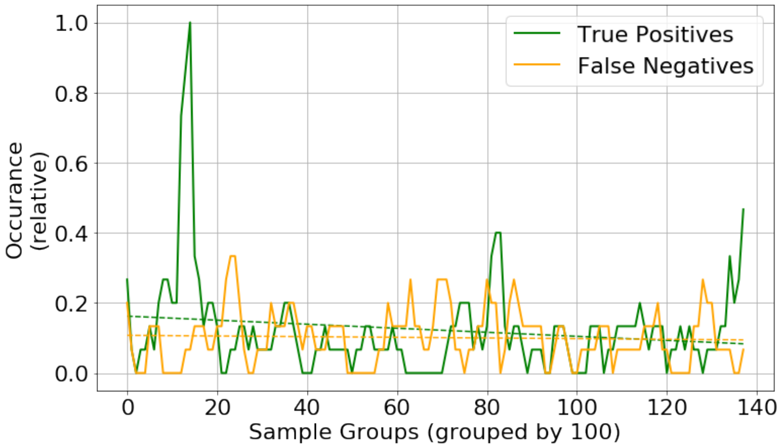

Figure 9. Relative indication of true positive (TP) and false negative (FN) detection over time, represented by full lines. Each data point represents 100 steel slabs. The dotted lines indicate the trend of TPs and FNs over the same period, providing a visual representation of the overall direction of the detections.

During the initial testing phase, we see a higher occurrence of true positives, while a consistent level of false negatives is maintained, suggesting a higher incidence of production defects during the time period, which our model accurately predicted without additional false negatives. Furthermore, the dataset contains over 60 varieties of casted steel, each with distinct chemical compositions. Notably, there was no substantial discrepancy in error across the different grades.

5. Summary and Discussion

A data-driven approach for multivariate time series combined with numerical static data was presented in this paper; the goal was to detect longitudinal cracks in steel slabs at the continuous caster. Thermocouple data from 68 sensors were combined with chemical and process information. We evaluated the performance of our models compared to a more classic industrial approach, and we also evaluated the performance gain when adding process information alongside the multivariate time series data.

Our two baseline models for comparison were the one currently in use at ArcelorMittal Belgium and a rule-based differential model that was inspired by the current literature. GBRT was used as the classifier because it meets several criteria required by industry, mainly in its ability to handle large input spaces and highly imbalanced data while being a gray box model, where feature importance can be inferred after the training process for a better understanding of process behavior.

Our choices of the baseline model, to not shuffle the data, and the resampling techniques for the training set were based on our previous work [4], where, based only on feature extraction methods, we demonstrated the potential pitfalls of shuffling data when creating training/testing splits. Shuffling can inadvertently lead to data leakage, where information from the test set leaks into the training set, thereby overestimating the model's performance. This is particularly problematic in time-series data, where shuffling can disrupt the inherent temporal order of observations. Furthermore, we explored the impact of data imbalance on model training. Imbalanced datasets, where one class significantly outnumbers the other, can bias the model towards the majority class, resulting in poor predictive performance for the minority class. We highlighted the importance of using appropriate techniques, such as resampling or using class weights, to mitigate the effects of data imbalance and ensure robust model training.

The experiments conducted here have shown that more complex models, such as CNN-SE-FCN, can outperform classic machine learning techniques, while the addition of the Squeeze and Excitation blocks has slightly improved model performance via an additional 5% more trainable parameters, and the addition of SE blocks also improved training times. Our model showed a 15% performance gain (higher F_β) over our baseline models. We also showed that removing the chemical features had the largest negative impact.

We have shown that our model is robust in an industrial setting, showing little decay in detection numbers over time. The recommended model updating period would be every 6 months; this process involves retraining the model with new data, either by including new data in the existing dataset or by adjusting the model based on the new data alone.

Optimizing the detection of defects early in the production chain is essential to enhance steel quality and guarantee the supply chain within the factory, as the cost of a defect significantly grows the longer it stays in production, wasting time and material. Evidence of which variables are more significant for detection is valuable as it compels the search for more complex approaches, which can further improve prediction. Additionally, our work can serve as a stepping stone for future research that may explore adaptable metric frameworks or tackle the challenges posed by class imbalance and cost considerations in predictive modeling in industrial domains. Based on the outcome of this research, the factory aims to further enhance the currently used model with the most important features. Feature engineering is also a topic to be discussed, as several signals have a larger influence, e.g., sensors at the center of the mold are more relevant since longitudinal cracks are more common in that region.

Author Contributions: Conceptualization, D.A.S. and A.V.Y.-D.D.; methodology, D.A.S.; software, D.A.S.; validation, D.A.S., A.V.Y.-D.D. and A.J.; formal analysis, D.A.S.; investigation, D.A.S.; resources, D.A.S. and A.V.Y.-D.D.; data curation, D.A.S. and A.V.Y.-D.D.; writing—original draft preparation, D.A.S.; writing—review and editing, A.V.Y.-D.D. and A.J.; visualization, D.A.S.; supervision, A.J. and E.M.; project administration, A.J. and E.M.; funding acquisition, D.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Flanders Innovation & Entrepreneurship (VLAIO), in cooperation with ArcelorMittal and Ghent University–imec under grant HBC.2019.2173.

Data Availability Statement: The data are not publicly available due to legal or ethical reasons.

Acknowledgments: We would like to thank Bart Soenen from ArcelorMittal Belgium for their valuable and constructive suggestions during the planning and development of this research work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thomas, B.G. Review on Modeling and Simulation of Continuous Casting. *Steel Res. Int.* **2018**, *89*, 1700312. [CrossRef]
2. Laha, D.; Ren, Y.; Suganthan, P.N. Modeling of steelmaking process with effective machine learning techniques. *Expert Syst. Appl.* **2015**, *42*, 4687–4696. [CrossRef]
3. Landstrom, A.; Thurley, M.J. Morphology-based crack detection for steel slabs. *IEEE J. Sel. Top. Signal Process.* **2012**, *6*, 866–875. [CrossRef]
4. Sala, D.A.; Jalalvand, A.; Van Yperen-De Deyne, A.; Mannens, E. Combining Time Series and Chemical Composition Data for Early Longitudinal Crack Detection at the Continuous Caster. In Proceedings of the 2022 Iron and Steel Technology Conference, Pittsburgh, PA, USA, 16–18 May 2022, pp. 711–718.
5. Gao, Y.; Wang, X.; Liu, Y.; Sun, Y.; Du, F.; Wang, F.; Wang, J. A Visual Detection Method of Longitudinal Crack Based on Computer Image Processing During Slab Continuous Casting. *Arch. Metall. Mater.* **2018**, *63*, 673–680.
6. Meilland, P.; Metz, M. Novel multiplexed eddy-current array for surface crack detection on rough steel surface. In Proceedings of the 9th ECNDT, Berlin, Germany, 25–29 September 2006, pp. 1–9.
7. Nishimine, T.; Tsuyama, O.; Tanaka, T.; Fujiwara, H. Automatic magnetic particle testing system for square billets. In Proceedings of the IAS'95, Conference Record of the 1995 IEEE Industry Applications Conference Thirtieth IAS Annual Meeting, Orlando, FL, USA, 8–12 October 1995; Volume 2, pp. 1585–1590.
8. Allazadeh, M.; Garcia, C.; Alderson, K.; Deardo, A. Ultrasonic Image Analysis of Steel Slabs. *Adv. Mater. Process.* **2008**, *166*, 26–27.
9. Sirgo, J.A.; Campo, R.; Lopez, A.; Diaz, A.; Sancho, L. Measurement of centerline segregation in steel slabs. In Proceedings of the Conference Record of the 2006 IEEE Industry Applications Conference Forty-First IAS Annual Meeting, Tampa, FL, USA, 8–12 October 2006; Volume 1, pp. 516–520.
10. Jin, C.; Kong, X.; Chang, J.; Cheng, H.; Liu, X. Internal crack detection of castings: A study based on relief algorithm and Adaboost-SVM. *Int. J. Adv. Manuf. Technol.* **2020**, *108*, 3313–3322. [CrossRef]
11. Veitch-Michaelis, J.; Tao, Y.; Walton, D.; Muller, J.; Crutchley, B.; Storey, J.; Paterson, C.; Chown, A. Crack detection in "as-cast" steel using laser triangulation and machine learning. In Proceedings of the 2016 13th Conference on Computer and Robot Vision (CRV), Victoria, BC, Canada, 1–3 June 2016; pp. 342–349.
12. Liu, Y.; Wang, X.; Du, F.; Yao, M.; Gao, Y.; Wang, F.; Wang, J. Computer vision detection of mold breakout in slab continuous casting using an optimized neural network. *Int. J. Adv. Manuf. Technol.* **2017**, *88*, 557–564. [CrossRef]
13. Sala, D.A.; Jalalvand, A.; Van Yperen-De Deyne, A.; Mannens, E. Multivariate time series for data-driven endpoint prediction in the basic oxygen furnace. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 1419–1426.
14. Ma, L.; Dong, J.; Peng, K.; Zhang, K. A novel data-based quality-related fault diagnosis scheme for fault detection and root cause diagnosis with application to hot strip mill process. *Control Eng. Pract.* **2017**, *67*, 43–51. [CrossRef]
15. Kim, K.H.; Yeo, T.J.; Oh, K.H.; Lee, D.N. Effect of carbon and sulfur in continuously cast strand on longitudinal surface cracks. *ISIJ Int.* **1996**, *36*, 284–289. [CrossRef]
16. Ji, C.X.; Cui, Y.; Zeng, Z.; Tian, Z.H.; Zhao, C.L.; Zhu, G.S. Continuous casting of high-Al steel in Shougang Jingtang steel works. *J. Iron Steel Res. Int.* **2015**, *22*, 53–56. [CrossRef]
17. Duan, H.; Wei, J.; Qi, L.; Wang, X.; Liu, Y.; Yao, M. Longitudinal Crack Detection Approach Based on Principal Component Analysis and Support Vector Machine for Slab Continuous Casting. *Steel Res. Int.* **2021**, *92*, 2100168. [CrossRef]
18. Hu, J.; Shen, L.; Sun, G. Squeeze-and-Excitation Networks. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June **2017**; pp. 7132–7141.
19. Hota, H.; Handa, R.; Shrivas, A. Time series data prediction using sliding window based RBF neural network. *Int. J. Comput. Intell. Res.* **2017**, *13*, 1145–1156.
20. Christ, M.; Braun, N.; Neuffer, J.; Kempa-Liehr, A.W. Time Series FeatuRe Extraction on basis of Scalable Hypothesis tests (tsfresh—A Python package). *Neurocomputing* **2018**, *307*, 72–77. [CrossRef]
21. Flach, P.; Kull, M. Precision-Recall-Gain Curves: PR Analysis Done Right. In *Proceedings of the Advances in Neural Information Processing Systems*; Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, **2015**; Volume 28.
22. Branco, P.; Torgo, L.; Ribeiro, R.P. A survey of predictive modeling on imbalanced domains. *ACM Comput. Surv. (CSUR)* **2016**, *49*, 1–50. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.