Temperature distribution prediction in control cooling process with recurrent neural network for variable-velocity hot rolling strips

Dong Chen1 · Rui Zhang1 · Zhenlei Li1 · Yunjie Li1 · Guo Yuan1

Received: 17 January 2022 / Accepted: 15 March 2022 / Published online: 29 April 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
Control cooling is an essential method for microstructure and mechanical property control in hot rolling strip making. Therefore, it is vital to realize high-precision temperature distribution prediction and control in cooling process to ensure the industrial production. In this paper, a traditional mechanism model based on finite-difference method combined with online cycle velocity calculation strategy was introduced as one of the baseline methods estimating temperature distribution. However, considering calculation time, variable-velocity rolling makes it difficult to rapidly realize temperature and modifying water distribution of all segments in cooling zone. Herein, a temperature distribution prediction method based on recurrent neural network was proposed, by fully considering the variable-velocity rolling dynamic characteristics. And the temperature distribution prediction performance of the model with different recurrent cell and time steps was evaluated. The results indicated that the proposed model could realize temperature distribution prediction, and the model based on bi-LSTM and 48 timesteps has the highest determination coefficient value of 0.976, the lowest root mean square error of 8.03, and a mean absolute error of 5.7. Furthermore, compared with baseline model, the proposed model retained lower computational cost, making it applicable in industrial application by providing real-time temperature distribution prediction.

Keywords Hot rolling strip · Variable-velocity rolling · Temperature distribution prediction · Recurrent neural network

1 Introduction
As an important fundamental raw material, hot rolling strips are widely used in industries, automobile, shipbuilding, and other fields. In hot rolling strip making, the control cooling process on run-out table is vital for steel product pursuing excellent and stable performance [1]. After finishing rolling, the strip will be cooled by water from a start cooling temperature to target temperature with a given cooling rate before being coiled, which promotes obtaining more fine grain size and improving strength and toughness [2, 3]. In control cooling process, high-precision temperature distribution prediction and cooling path control are key factors for ensuring the mechanical property [4].

In cooling process, as the strip delivers from the finishing mill to the down coiler, water distribution must be adjusted according to variable rolling velocity to meet the target coil temperature (CT). Water cooling of strip contains complicated heat exchange process, affected by many factors, such as strip temperature, water flow, strip thickness, chemical composition, and strip velocity etc. [5, 6]. Once the strip enters the cooling zone, strip thickness, width, and composition in strip length direction have little difference; the start cooling temperature and velocity are the main factors influencing temperature control. Currently, to increase the productivity and decrease heat loss on delay of transfer bar, the finishing rolling process always adopts the variable-velocity rolling strategy, which leads the strip velocity varying in a large scale [7]. Meanwhile, the adjustment of rolling velocity is also used to control finishing mill delivery temperature (FDT, herein also is start cooling temperature), by which the rolling acceleration/deceleration and velocity are dynamically changed to meet the target FDT and also influence coil temperature control accuracy, significantly [8]. Under variable-velocity rolling, water distribution is expressed as open/closed cooling headers, and header flow is the only adjustment method to guarantee target coil temperature. Therefore, precise water distribution calculation

1 State Key Laboratory of Rolling and Automation, Northeastern University, No. 11, Lane 3, Wenhua Road, HePing District, Shenyang 110819, China

✉ Guo Yuan
yuanguoneu@163.com

© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

The International Journal of Advanced Manufacturing Technology (2022) 120:7533–7546
https://doi.org/10.1007/s00170-022-09065-8
Based on accuracy velocity prediction is crucial for cooling path and temperature distribution control in cooling process [9].

To build a high-accuracy temperature distribution prediction and control model, many scholars have been paying attention to related research. So far, three main kinds of models have been developed and applied. Index model is an important scheme to deal with convective heat transfer, considering convective between cooling water and strip surface as the main heat exchange in the cooling process [10, 11]. However, due to simplifications and ignoring temperature gradient in strip thickness direction, the application effect of index model is not good, especially for heavy thickness plates. Another kind of model is statistical model, which was built relying on statistical regularity [12, 13]. In statistical modeling, by fully considering the physical mechanism of the cooling process, the main influence parameters of cooling process were considered and determined by statistical regression methods [14–17]. Furthermore, to improve the control precision, the influence parameters were classified to categories based on steel grades and thicknesses or other boundary conditions. Then, each category has a series of model parameters. However, the temperature control accuracy of this method was limited by the number of categories. Furthermore, some advanced methods were used to improve the control accuracy, such as fuzzy control and model prediction control [12, 18, 19]. The advanced methods in control cooling greatly rely on the statistical model or index model. The third model is built by finite difference method (FDM) and heat transfer theory, which divides strip and time into nodes and finite number of steps to solve differential equations by approximate derivatives. Thence, the spatial and sequential temperature distribution of strip could be obtained [20, 21]. During modeling, the heat transfer form is radiation and convectional heat transfer for surface node, while the heat transfer form is heat conduction for inner nodes. In control cooling process, the temperature gradient in thick direction is fully considered in FDM. Consequently, control accuracy was obviously improved, especially for the heavy thickness strip. At present, the most used method is combining FDM and statistical method; the former is used for temperature prediction model building, and the latter is used for seeking optimal model parameters.

As above-mentioned, strip velocity significantly influences cooling process. To realize high-precision prediction of velocity in cooling process, time-velocity-distance (TVD) method and real-time modification are the commonly used methods. Li proposed a velocity-controlled strategy for each strip element based on TVD [22]. An online cyclic modification and calculation strategy is designed to minimize the temperature fluctuation in variable-velocity rolling. However, consumption time of modification and calculation for each segment exceeds the requirement of the online system. Based on this, Li proposed an improved method that could realize online monitor and control of cooling temperature by modifying at setpoint and decreasing the computation time [20]. However, due to the irregular changes of hot rolling, the more times of temperature modifying are beneficial for the temperature control. Thence, a new temperature prediction method which could quickly realize prediction and repeatedly match the more times of velocity modifying is needed.

Although the introduced temperature prediction methods and models are still accepted and widely used due to practicality and stability, it is difficult to further improve control level and satisfy the requirement mentioned above. Therefore, some model based on machine learning was developed and shows prospect of high control accuracy.

As one of the well-known instance-based learning methods, KNN has been widely applied in many fields and also generally used in control cooling process. Zheng adopted a new approach to build a model for accelerated cooling process by IIR-KNN [23]. Based on the k selected similar case, locally linear reconstruction is applied to determine the best output parameters for current plate. Zhang proposed a variable scale grid model for temperature self-adaptive control, which builds a multi-dimension space system, and each dimension is determined by effect of each factor on heat transfer [24]. Every plate can find a corresponding point in space, according to its own process conditions to correct temperature calculation. The neural network is another one of most popular algorithms used in steel industry, which has stronger nonlinearity and capability of adaptive information processing. Some researchers have successfully adopted ANNs to predict heat transfer coefficient, temperature, or water flow [25–27]. Valentina introduced an ANN to find correlations between model parameters and process variables [28]. Xing developed a hybrid intelligent identification model by combining the RBF neural networks, CBR, and fuzzy logic reasoning, which can make a great contribution in improving the coil temperature precision by prediction precision of correct identification [29]. However, in the abovementioned, variable strip velocity and temperature distribution in cooling process were ignored. They only take constant velocity and final cooling temperature into consideration. Obviously, in order to keep mechanical performance fluctuation within a narrow range, those are not suitable for the variable-velocity rolling and cannot realize temperature distribution prediction.

In view of the above problems, by fully considering the ability of dealing with sequence and the variable-velocity rolling dynamic characteristic, a novel temperature prediction model based on recurrent neural network was proposed and constructed, which could accurately predict temperature distribution under variable-velocity rolling. To improve the precision and robustness, the Pauta criterion and isolate
forest algorithm were used to preprocess the industrial data collected from control cooling process. In view of particular of water distribution feature, the feature process method was introduced. The loss function was modified to improve training efficiency and accuracy of the model. In addition, the recurrent cell and length of input sequence were tested to determine optimal recurrent neural network model.

2 Control cooling process on run-out table of hot rolling strips

2.1 Description of cooling process

The schematic diagram Fig. 1 shows control cooling process of hot rolling strips on run-out table. During the cooling process, the strip will be sent into run-out table after being delivered from the finishing mill, and then it will be cooled in the following long cooling zone and finally coiled by down coiler. The cooling zone is made up by ultrafast cooling (UFC) and laminar cooling. The UFC system lies between the finishing mill and laminar cooling, contains 40 top headers and 40 bottom headers, and divided into 4 groups. The UFC header flow and pressure can be adjusted stepless for different cooling ability with range of headers, heat dissipation along the length and width directions could be ignored. Also, heat latent of phase transaction could be contained in correction of specific heat. The heat transfer equation is simply expressed as Eq. (1):

$$\rho c_p \frac{dT}{dt} - \frac{\partial}{\partial x} \left[ \lambda \frac{dT}{dx} \right] = 0$$  \hspace{1cm} (1)

The boundary conditions on top and bottom surface can be described in form like Newton’s convection Eq. (2) and equivalent to water cooling and air cooling:

$$\pm \lambda \frac{dT}{dx} = \alpha (T - T_m)$$ \hspace{1cm} (2)

where \(\rho\) is density of strip steels in \(\text{kg/m}^3\), \(c_p\) is specific heat in \(\text{J/(kg K)}\), \(\lambda\) is coefficient of thermal conductivity in \(\text{W/(m K)}\), \(T\) is temperature of strip steels in \(\text{K}\), \(\tau\) is heat transfer time in \(\text{s}\), \(x\) is thickness direction, \(T_m\) is cooling medium temperature in \(\text{K}\), and \(\alpha\) is heat transfer coefficient in \(\text{W/(m}^2\text{K)}\).

Finite difference method was used to model internal heat transfer through conduction. And finite difference calculations maintained the temperature of each node as the piece moves from FDT pyrometers to CT pyrometers. As shown in Fig. 2, the half thickness of strip was researched. With boundary condition, the node in thickness Eq. (1) can be expressed in explicit difference form as Eq. (3):

\[
\begin{bmatrix}
T_1 \\
T_2 \\
T_3 \\
\vdots \\
T_{n-2} \\
T_{n-1} \\
T_n
\end{bmatrix}^{p+1} = 
\begin{bmatrix}
1 - 2F_0 & 2F_0 & 0 & 0 & \cdots & 0 & 0 \\
F_0 & 1 - 2F_0 & F_0 & 0 & \cdots & 0 & 0 \\
F_0 & F_0 & 1 - 2F_0 & F_0 & \cdots & 0 & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
0 & \cdots & \cdots & \cdots & \cdots & \cdots & 0 \\
\end{bmatrix} 
\begin{bmatrix}
T_1 \\
T_2 \\
T_3 \\
\vdots \\
T_{n-2} \\
T_{n-1} \\
T_n
\end{bmatrix}^{p} + 
\begin{bmatrix}
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p} \\
\frac{F_0}{\rho c_p} \Delta x \frac{\lambda \Delta \tau}{\rho c_p}
\end{bmatrix}
\]

(3)

50 ~ 300m³/h and 0.3 ~ 1.0 MPa, respectively. Laminar cooling is consisted of 72 U-type top headers and 72 straight-bottom headers, divided into 16 groups. The foremost 14 groups are the main cooling section and the last two groups are vernier cooling for feedback control. In cooling process, the number of cooling headers and the water flow of each header are taken as control variables to adjust the temperature distribution of the strip.

2.2 Thermal dynamic model

Considering that heat transfer in thickness direction is much higher than width direction and rolling direction for high rolling velocity and uniform cooling capability of cooling
water convection. So, the heat transfer coefficient of water cooling in the UFC system and laminar cooling can be calculated as follows, Eqs. (6) and (7), respectively:

\[
\alpha_w = C_w \cdot A_1 \cdot \frac{Q}{B \cdot L} \cdot (H)^{A_2} \cdot \exp\left(-A_3 (T - T_w)\right) \cdot \left(\frac{T_w}{T_o}\right)^{A_4} \cdot \left(\frac{P}{P_o}\right)^{A_5} \cdot \left(\frac{V}{V_o}\right)^{A_6}
\]

(6)

\[
\alpha_w = C_w \cdot A_1 \cdot \frac{Q}{B \cdot L} \cdot (H)^{A_2} \cdot \exp\left(-A_3 (T - T_w)\right) \cdot \left(\frac{T_w}{T_o}\right)^{A_4} \cdot \left(\frac{V}{V_o}\right)^{A_6}
\]

(7)

where \(C_w\) is correction coefficient of strip width, \(A_1\)–\(A_6\) are model coefficients, \(Q\) is summation water flow in m³/h, \(T_w\) is measured water temperature in K, \(P\) is measured cooling water pressure in MPa, \(V\) is measured velocity of strip in m/s, \(T_o\) is reference cooling water temperature in K, \(P_o\) is reference water pressure in MPa, \(V_o\) is based velocity in m/s, \(B\) is width of cooling bank in m, \(L\) is effective cooling length in m, and \(H\) is strip thickness in m.

Based on the above equations, the through thickness temperature field can be obtained, once the parameters in heat transfer coefficient equations were determined. As can be seen in Eqs. (6)-(7), water flow, strip temperature, water temperature, and pressure were easier to obtain; the difficulty was to get accurate strip velocity in variable-velocity rolling.

### 3 The online cycle velocity calculation strategy

In rolling process, strip velocity combining with inter-stand cooling is regulated to keep the measured strip temperature meet target FDT. Generally, strip goes through low velocity threading, first acceleration, high acceleration, and tail out deceleration for FDT control as shown in Fig. 3, the typical time, velocity, and distance curve (TVD curve). TVD curve represents the relationship of rolling velocity, time, and rolling distance, by which the entrance and exit velocity of segment in different micro cooling could be obtained. Meanwhile, rolling acceleration or deceleration will be changed, when the measured FDT exceeds allowable tolerance. Due to variable-velocity rolling, water distribution should be dynamically adjusted to guarantee final coil temperature. Thus, accuracy velocity prediction for strip is basis for temperature prediction and water distribution control. To realize high-accuracy velocity prediction, an online cycle velocity calculation algorithm was introduced.

To achieve high-accuracy temperature control in whole strip length, the strip is divided into numbers of fixed-length segments, and each segment was separately controlled. The velocity curve and water-cooling time of each segment passing through run-out table needs to be predicted before the segment enters the cooling section. Generally, in different rolling stage, strip velocity change has certain characteristics. Then, the entire length of strip was divided into five parts based on the characteristics of TVD curve: low velocity threading, low acceleration before the strip coiled, high acceleration to max running speed, tail out deceleration, and the remaining length uncoiled, named \(S_1\)–\(S_5\) part for convenience. The online cycle velocity calculation algorithm is shown in Fig. 4.

---

**Fig. 1** Schematic diagram of cooling system on run-out table

**Fig. 2** One-dimensional region discretization
To begin with, each part length $l_i$ and velocity $v_i$ of $S_i$ ($i$ values from 1 to 5) was calculated with finishing rolling preset threading velocity and acceleration/deceleration according to kinematics formula (Eq. (8)). In this way, the overall TVD curve of strip was determined. In addition, running velocity of the segment in cooling section was calculated based on the overall TVD curve. Herein, running velocity of each segment was discerned as entrance and exit velocity of $V_{ij}$ and $V_{ij}'$ at $j$th micro cooling section. Furthermore, prediction velocity was utilized to calculate the temperature distribution on run-out table, and water distribution would be adjusted to satisfy the target temperature. Once a new segment arrives at FDT, velocity calculation algorithm will be trigged. Based on real-time rolling velocity and acceleration/deceleration, the above three steps will be repeated. After re-calculation, the velocity of all segments in cooling zone would be modified:

$$\left\{ \begin{array}{l}
V_i^2 - V_0^2 = 2as \\
t = \frac{V_i - V_0}{a}
\end{array} \right. \tag{8}$$

The following figures in Fig. 5 are calculation velocity and actual velocity curve of 4.25 mm SPA-H and 13.75 mm Q355B, respectively. The results show that the abovementioned velocity calculation algorithm can accurately predict the strip velocity.

4 Deep learning approach for temperature distribution prediction

4.1 Problem existing

With the above thermal dynamic model and online cycle velocity calculation strategy, it would be possible to achieve desired cooling process temperature control by re-calculation of temperature distribution and modifying water distribution, successively. The more times TVD curves are updating, the higher the accuracy of velocity and temperature prediction. However, it would occupy excessive computation resources of temperature re-calculation and modifying, which could not meet requirement of the online model. Despite that the improved method was researched, it still tries to reduce the times of velocity and temperature update which is not beneficial to temperature control in cooling process. Therefore, it is necessary to build a less computational model which also could achieve variable-velocity rolling temperature prediction to enhance the capability of dynamic control.
4.2 Fundamental of recurrent neural network

Recurrent neural network (RNN) is a type of recursive neural network, which takes sequence data as input and performs recursion in evolution direction of sequence, and RNN was developed to solve the issue of the information forgotten in traditional feed-forward and feedback neural networks [30]. As shown in Fig. 6, given an input sequence \( X = (x_1, x_2, ..., x_t) \), a standard RNN computes the hidden vector sequence \( S = (s_1, s_2, ..., s_t) \) and output vector sequence \( Y = (y_1, y_2, ..., y_t) \) by iterating the following equations from \( t = 1 \) to \( t \):

\[
y_t = g(V \ast S_t) \quad (9)
\]

\[
s_t = f(U \ast s_t + W \ast s_{t-1}) \quad (10)
\]

where \( V \) denotes weight matrices of hidden-output layer, \( U \) is input-hidden layer weight matrices. \( W \) expresses hidden step weight matrices. \( g \) and \( f \) are function activation function, and \( f \) is usually an element-wise application of a sigmoid function. It can be seen from Fig. 6 that the values of hidden layer depend not only on current input \( x_t \), but also on previous hidden layers. In other words, RNN has memory of the previous input content. Due to the “memory,” the RNN and its evolutions have performed well in sequence tasks and were widely used in text recognition, speech recognition, real-time translation, and image identification [31–33].

Subsequently, the long short-term memory (LSTM) unit [34, 35] and its simple version gated recurrent unit (GRU) [36, 37] are proposed to overcome vanishing gradient and capture long-term dependency. As can be seen in Fig. 7, compared with standard RNN, LSTM and GRU have more complex structure. In standard RNN, repeating module only has a very simple structure, tanh layer. While in LSTM, the repeating module has a different structure, except for tanh layer; there are also three different gates: forget gate \( f_t \), input gate \( i_t \), and output gate \( o_t \).
gate $i_t$ and output gate $o_t$. To achieve the long-term dependency, at each time step, the cell state $c_t$ is slowly updated to new unit, which is determined by $f_t$ and $i_t$. The $f_t$ controls the memory with sigmoid activation function based on input vector $x_t$, and prior hidden state $h_{t-1}$, while $i_t$ controls remembering current information. Finally, hidden state $h_t$ is calculated using $o_t$ and $c_t$ with tanh activation. Furthermore, GRU is simplified from LSTM to decrease a large amount of learning parameters. As shown in Fig. 7c, different with LSTM, $h_t$ is determined by an update gate $z_t$ and a reset gate $r_t$, while no longer a cell state exists in the unit. The former controls how much information from the previous hidden state will carry over to the current one, while the latter decides whether the previous hidden state is forgettable [37].

4.3 Temperature distribution prediction based on recurrent neural networks

As Sect. 3 mentioned, to realize high-precision temperature control of entire strip length, the strip is divided into a number of segments, and each segment is separately controlled. Correspondingly, cooling zone is also divided into several micro cooling zones, same length with segment. In the cooling process, segment passes through all micro cooling zones in sequence from FDT pyrometer to CT pyrometer; with temperature drop increasing, strip temperature finally meets the target value. Obviously, strip temperature calculation in each micro cooling zone has obvious feature of sequence data. Thence, RNN algorithm can be adopted to prediction cooling temperature in variable-velocity process based on existing models. Taking one temperature calculation in micro cooling zone as one timestep, the illustration of the proposed model is shown in Fig. 8. Except for basic parameter of strip, such as thickness, width which is almost not changed, strip velocity, and water distribution in each micro cooling zone are divided into input sequential values, and temperature distribution sequence is output. It should be noted: $T$, $V$, and $F$ shown in Fig. 8 mean temperature, velocity, and water distribution in each cooling zone.

In hot rolling strip line, automation system records almost all of process parameters in database or production logs, mainly including real-time data and model setup data. Two months of 12,998 coil production data relevant to control cooling process on run-out table have been collected. First, due to the unstable measuring of pyrometer, data of each strip head end and tail end is removed. Then, the remained data are randomly sampled and combined to obtain 99,000 experiment strip segment data. These samples are used to develop temperature distribution prediction model based on RNN. Nineteen key variables including temperature distribution data are selected for neural network modeling. Because medium temperature in cooling zone is hardly measured, and temperature distribution except for initial temperature FDT and final cooling temperature CT is measured, other temperature distribution values were padded with traditional model re-calculation values based on measured temperature data. The descriptions of dataset are shown in Table 1 and Fig. 9. In Fig. 9, fractal dimension visualization diagrams of experiment data show distribution of significant variables and output data. The enormous amount of data proves the developed model will have a strong robustness with the dataset. In particular, total flow here is used for abnormal detection.

Water distribution in each cooling zone represents cooling header opened/closed status and header water flow. Because of different cooling mechanism, UFC spray headers and laminar spray headers have different cooling efficiency even though in

![Fig. 7 Schematic diagram of recurrent units: a standard recurrent neural network, b operation of LSTM unit, c operation of GRU unit](image)
same cooling water flow. Moreover, water flow of UFC spray header can be adjusted in a wide range. In order to distinguish
UFC header, laminar main cooling header, and vernier header, cooling efficiency $E$ for different spray header was introduced.

Taking one top header opened as example, heat transfer coefficient Eqs. (6~7) can be transformed shown as Eq. (10). After
the cooling efficiency of different spray header is determined, water distribution in micro cooling zone can be indicatory as shown in Fig. 10. In addition, owing to the different number in each micro cooling zone, water distribution sequence is padding with 0, same with the closed spray headers.

$$h_w = E \cdot \frac{Q}{BL} (h)^{A_2} \cdot \exp(-A_3 (t - T_w)) \left( \frac{T_w}{T_0} \right)^{A_4} \left( \frac{V}{V_0} \right)^{A_5}$$

(11)

---

**Table 1** Description of parameter of input and output variable for RNN model

| No | Parameter | Description     | Unit | Range          | Mean  | Std  |
|----|-----------|-----------------|------|----------------|-------|------|
| 1  | $t$       | Strip thickness | mm   | [1.6, 20]      | 4.201 | 2.216|
| 2  | $W$       | Strip width     | mm   | [836, 2075]    | 1391.406 | 228.065|
| 3  | $C$       | C content       | wt%  | [0.0006, 0.495] | 0.054 | 0.049|
| 4  | $Si$      | Si content      | wt%  | [0.0001, 1.151] | 0.052 | 0.106|
| 5  | $Mn$      | Mn content      | wt%  | [0.114, 2.410] | 0.471 | 0.451|
| 6  | $Ni$      | Ni content      | wt%  | [0, 0.075]     | 0.009 | 0.007|
| 7  | $Cr$      | Cr content      | wt%  | [0, 0.547]     | 0.031 | 0.077|
| 8  | $V$       | V content       | wt%  | [0, 0.040]     | 0.001 | 0.002|
| 9  | $Nb$      | Nb content      | wt%  | [0, 0.056]     | 0.004 | 0.010|
| 10 | $Mo$      | Mo content      | wt%  | [0, 0.244]     | 0.004 | 0.024|
| 11 | $Ti$      | Ti content      | wt%  | [0, 0.174]     | 0.021 | 0.029|
| 12 | $B$       | B content       | wt%  | [0, 0.003]     | 0.000 | 0.000|
| 13 | $Cu$      | Cu content      | wt%  | [0, 0.273]     | 0.014 | 0.035|
| 14 | $T_w$     | Water temperature | °C  | [29.5, 38.7]  | 33.894 | 1.874|
| 15 | $P$       | Water pressure  | Bar  | [3, 8.5]       | 4.234 | 0.457|
| 16 | $T_F$     | Finish rolling delivery temperature | °C | [771, 945] | 880.989 | 28.073|
| 17 | $V$       | Velocity        | m/s  | [2.11, 19.63]  | 10.334 | 2.843|
| 18 | $F$       | Water distribution | –  | –              | –     | –     |
| 19 | $q$       | Total flow      | m$^3$/h | [1880, 12905] | 4615 | 1880|

**T** Temperature distribution | °C | [471, 763] | [704, 864]

---

![Fractal dimension visualization diagrams of data](image-url)
In general, the original data collected from hot rolling line always contains abnormality, noise, and outliers, which may influence modeling and lead to misleading predictions. Therefore, these data must be removed. In first, the date of missing values was directly removed. And then, noise data was eliminated by Pauta criterion. The Pauta criterion implemented in this paper was given as Eqs. (11–13) [38]; data with absolute deviation exceeding the limitation will be regarded as noise and removed. Eighteen key variables will be treated in this way:

\[
|x_i - \mu| > 3\sigma \tag{12}
\]

\[
\mu = \frac{1}{n} \sum_{i} x_i \tag{13}
\]

\[
\sigma = \sqrt{\frac{1}{n} \sum_{i} (x_i - \mu)} \tag{14}
\]

where \( \mu \) is average of \( x_i \) and \( \sigma \) is standard deviation of samples. The result of abnormal detection with Pauta criterion is shown in Fig. 11.

Due to that Pauta criterion is only effective for normal distribution or approximate normal distribution, in order to remove more outlier data, isolation forest is adopted to mining anomaly. Isolation forest algorithm, also named iforest, was proposed by Zhou [38]. Isolation forest is an efficient approach to anomaly detection that considers that most points in the dataset are not anomalous, and anomalies are typically very different from the rest of the dataset. Due to its linear time complexity and low demand for storage space, isolation forest has been used in many scenarios. This paper applies iforest to eliminate outlier of 19 features shown in Table 1. There are two hyperparameter isolation trees \( t \), and sub-sampling size \( \psi \) should be optimized for iforest. Herein, we use grid search method to select suitability for each parameter. The grid search result is shown in Fig. 12.
As can be seen in Fig. 12, the fluctuating range of proportion is higher at small isolation tree $t$ and then stabilized when $t \geq 100$. In general, train process time will be linearly increasing with isolation, and the iforest will be converged well when isolation tree $t = 100$; the higher isolation has no meaning [38]. On the other hand, with the increasing of sub-sampling size $\psi$, the proportion is gradually decreasing; when $\psi$ is bigger than 256, the proportion is stable in range of $0.14 \sim 0.16$. In summary, we select $t = 100$ and $\psi = 256$ as algorithm parameter. Base on low signal-to-noise ratio in general, the top 2% rank in dataset was removed as abnormal data, after determining parameter.

4.3.2 Data normalization and division

Different parameters often have different dimensions and units (as shown in Table 1). In order to eliminate the “eaten” situation caused by dimensions, input data must be normalized. The method used in this paper is zero-mean normalization and the formula is

$$x_i^* = \frac{(x_i - \mu)}{\sigma}$$

where $x_i^*$, $x_i$, $\mu$, and $\sigma$ are normalized data, original data, mean of all the data, and standard deviation. After normalization, the dataset is split into training/validation/testing set with proportion 80%, 10%, and 10%, respectively.

Learning parameter depending on the proposed model is trained with root mean square prop (RMSPop) algorithm for minimizing error between prediction value and actual value. The loss function of mean square error here is defined as

$$\text{Loss} = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}$$

(16)

where $\hat{y}_i$ and $y_i$ are prediction temperature distribution and actual temperature distribution in cooling zone, respectively, while $N$ is the number data in dataset. Due to only the start cooling temperature FDT and final CT in temperature distribution sequence could be measured, others are padded with traditional model re-calculation values. If prediction error of temperature distribution was directly used as criterion of gradient descent, each prediction error of temperature in distribution sequence is equivalent, while measured actual value is worth specially utilizing. Thus, an additional weight was added, and the definition of new improved loss function is as follows:

$$\text{Loss} = \frac{\sum_{i=1}^{N} w_j (\hat{y}_i - y_i)^2}{N}$$

(17)

And the weight of each error is derived from the distance with actual temperature measured as Eq. (17); the smaller the distance between temperature distribution and actual temperature, the greater the weight value of the loss corresponding:

$$w_j = \frac{1}{\min(|j - n|, |j - 1|)}$$

(18)

where $j$ is serial number in sequence, and $n$ is length of the input sequence.

5 Result and discussion

5.1 Prediction of temperature distribution

In this section, to build an optimal model, several experiments such as compassion concerning types of recurrent cells, sequence length, and hyper-parameter settings are carried out and briefly illustrated. Above all, different types of recurrent cells are compared, i.e., standard RNN, LSTM, and GRU. The recurrent cell shows a significant difference in extracting the temporal information via hidden states as introduced in Sect. 4.2. Secondly, the influence of length of input sequence, named timesteps, was examined. Literatures show that vanishing gradient due to decay as sequence length increasing makes the loss of earlier input sequence [38]. The performance of the proposed method is evaluated with prediction accuracy and computation time. Prediction accuracy is measured by three different metrics: root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$). The formulas of RMSE, MAE, and $R^2$ are as follows:

$$\text{RMSE} = \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{N}$$

(19)
where $y_i$ and $y_i$ are prediction and actual temperature distribution in cooling zone, respectively, while $N$ is number of the dataset. In addition, to evaluate whether it is competitive on the dynamic control for temperature distribution prediction in control cooling process, computation time of the proposed method is compared with traditional model.

This project is using Python and the Keras framework. The training and testing were performed on a system equipped with an Intel(R) Core(TM) i5-6300HQ CPU, 16 GB of RAM.

The evaluation criteria result of prediction accuracy is as depicted in Table 2. As can be seen in Table 2, models based on LSTM and GRU slightly outperformed than ones with standard RNN. The models based on bidirectional unit show similar performance regularity, because bidirectional unit performs in sequential dependencies of forwarding and backward direction [39]; it is observed that the bidirectional unit of each model generally enhances the prediction accuracy. In addition, it is shown that the length of input sequence (timesteps) affects prediction accuracy of each recurrent unit. Due to higher velocity prediction accuracy in more timesteps, the models with 48 timesteps have higher prediction performance in both unidirectional and bidirectional models, even if long sequence prevents from capturing a piece of extended temporal [40]. As aforementioned, more timesteps mean more times of velocity modification calculation, and higher accuracy velocity prediction will be realized, which also makes the proposed model yield higher performance. In conclusion, the bi-LSTM model yielded the better value for each evaluation criterion than did the other units, and the model based on bi-LSTM and 48 timesteps showed the best prediction performance with the lowest RMSE 8.03 and MAE 5.7 and highest $R^2$ 0.976.

In order to further evaluate the performance of the proposed model for temperature distribution prediction, several segments with different rolling velocity of 11.8 mm 380CL strip were selected as examples; temperature distribution prediction is described in Fig. 13. Compared with the baseline model, the proposed model with 48 timesteps yielded desirable prediction as shown in Fig. 13, which proved the ability of temperature distribution prediction. Moreover, the final CT prediction performance of models based on bi-LSTM is shown in Fig. 14. It can be clearly seen that

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i|$$

(20)

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

(21)

where $\hat{y}_i$ and $y_i$ are prediction and actual temperature distribution in cooling zone, respectively, while $N$ is number of the dataset. In addition, to evaluate whether it is competitive on the dynamic control for temperature distribution prediction in control cooling process, computation time of the proposed method is compared with traditional model.

This project is using Python and the Keras framework. The training and testing were performed on a system equipped with an Intel(R) Core(TM) i5-6300HQ CPU, 16 GB of RAM.

The evaluation criteria result of prediction accuracy is as depicted in Table 2. As can be seen in Table 2, models based on LSTM and GRU slightly outperformed than ones with standard RNN. The models based on bidirectional unit show similar performance regularity, because bidirectional unit performs in sequential dependencies of forwarding and backward direction [39]; it is observed that the bidirectional unit of each model generally enhances the prediction accuracy. In addition, it is shown that the length of input sequence (timesteps) affects prediction accuracy of each recurrent unit. Due to higher velocity prediction accuracy in more timesteps, the models with 48 timesteps have higher prediction performance in both unidirectional and bidirectional models, even if long sequence prevents from capturing a piece of extended temporal [40]. As aforementioned, more timesteps mean more times of velocity modification calculation, and higher accuracy velocity prediction will be realized, which also makes the proposed model yield higher performance. In conclusion, the bi-LSTM model yielded the better value for each evaluation criterion than did the other units, and the model based on bi-LSTM and 48 timesteps showed the best prediction performance with the lowest RMSE 8.03 and MAE 5.7 and highest $R^2$ 0.976.

In order to further evaluate the performance of the proposed model for temperature distribution prediction, several segments with different rolling velocity of 11.8 mm 380CL strip were selected as examples; temperature distribution prediction is described in Fig. 13. Compared with the baseline model, the proposed model with 48 timesteps yielded desirable prediction as shown in Fig. 13, which proved the ability of temperature distribution prediction. Moreover, the final CT prediction performance of models based on bi-LSTM is shown in Fig. 14. It can be clearly seen that

![Fig. 13 Prediction and measured temperature distribution of 11.8 mm 380CL under rolling stage](image-url)
temperature prediction error decreased with increasing input length. And prediction accuracy within $\pm 20 \, ^\circ C$ was 94.9%, 95.6%, and 97.6% for 12, 24, and 48 timesteps, respectively, equaling or higher than that of the baseline model, whose best accuracy within $\pm 20 \, ^\circ C$ is 95% in application. Furthermore, the model based on bi-LSTM and 48 timesteps could realize 95% prediction accuracy of $\pm 15 \, ^\circ C$, meaning more times of modifying velocity are an advantage for temperature prediction, and control.

### 5.2 Computation time

We aim to develop not only a high-precision prediction but also a real-time model which could just in time deal with variable-velocity rolling. Thus, the model should deal with variable-velocity dynamic process and satisfy the requirement of online temperature monitoring and modification. Herein, the proposed deep neural network-based model and traditional mechanism model were built to compare to computation time. Table 3 shows the comparison result for each model with different timesteps. It is shown that computational cost of the proposed model is considerably lower than the baseline mechanism model. The baseline mechanism model requires excessive computational resources by 1.2, 1.7, and 2.5 s for 12, 24, and 48 timesteps to accomplish re-calculation of temperature distribution for all segments in the cooling process, respectively. The proposed model based on bi-LSTM shows relatively similar consumption time of less than 0.001 s.

To summarize, the proposed model outperformed the mechanism model concerning computation time, which makes it possible to establish a real-time dynamic control in an actual control cooling process.

### 6 Conclusions and future work

In this paper, a temperature distribution prediction based on recurrent neural network for variable-velocity hot rolling strips was proposed. The isolated forest algorithm was used to eliminate outliers. And different recurrent cell and length of input sequences were tested to determine the optimal model. The key results were as follows:

A traditional mechanism model of temperature distribution prediction based on finite-difference method was introduced as baseline model. Moreover, to match precise velocity value, an online cycle velocity calculation strategy was proposed. The result showed that the proposed velocity calculation strategy could realize high-precision velocity prediction.

| Table 3 | Comparison of the computation time between the mechanism model and proposed model |  |
| --- | --- | --- | --- |
| Model | 12 timesteps s | 24 timesteps s | 48 timesteps s |
| Proposed model | 0.0009 | 0.0009 | 0.0008 |
| Mechanism model | 1.2 | 1.7 | 2.5 |
The Pauta criterion and isolate forest algorithm were used to eliminate outliers, successively. Loss function was modified to make it more suitable for training the proposed model, due to that the measured temperature should occupy more important influence.

The prediction performance with standard RNN, LSTM, GRU, and bidirectional operation recurrent cell and length of input sequence of 12, 24, and 48 were compared with evaluation criteria of MSE, MAE, and $R^2$. Among them, the model built with bi-LSTM and 48 timesteps yielded best performance in terms of prediction accuracy and has the lowest RMSE 8.03 and MAE 5.7 and the highest $R^2$ 0.976.

Compared with the baseline model, the proposed model cost less times (less than 0.001 s for one cycle temperature distribution prediction of all segment in cooling zone), which increases the feasibility of the model used online for industrial application. It is feasible to unitize RNN model to realize temperature distribution prediction for hot strip control cooling process on run-out table. The follow-up work is going to improve prediction accuracy by obtaining more input features and testing more effective model structures, such as convolutional neural network+recurrent neural network, which is beneficial for recognizing more information from sequence data. In addition, the input feature of water distribution should be considered for influence effect on cooling, and influence effect should be recognized by the proposed model to improve the prediction performance.

Author contribution Dong Chen: methodology, data curation, visualization, investigation, writing — original draft. Rui Zhang data curation, software, visualization. Zhenlei Li: conceptualization, methodology, supervision, validation, writing — review and editing. Yunjie Li: visualization. Guo Yuan: conceptualization, supervision, validation, writing — review and editing.

Funding This work was supported by the National Natural Science Foundation of China (51804074), Project funded by China Postdoctoral Science Foundation (2020M680964), Fundamental Research Funds for the Central Universities (N2107004), and Northeastern University Postdoctoral Foundation (2020323).

Data Availability All the data that supports the findings of study is available from the corresponding author upon reasonable request.

Declarations

Ethics approval We promise that our research in this work is in compliance with ethical standards all the while.

Consent to participate All the authors consent to participate in this work.

Consent to publication The authors agree to publication in the Journal indicated below and also to publication of article in English by Springer in Springer’s corresponding English-language journal.

Competing interests The authors declare no competing interests.

References

1. Herman JC (2001) Impact of new rolling and cooling technologies on thermomechanically processed steels. Ironmak Steelmak 28:159–163. https://doi.org/10.1179/030192301677894
2. Tsuyama S (2015) Thick plate technology for the last 100 years: a world leader in Thermo Mechanical Control Process. ISIJ Int 55:67–78. https://doi.org/10.2355/isijinternational.55.67
3. Uranga P, Rodriguez-Ibabe JM (2020) Thermomechanical processing of steels Metals 10(5):641. https://doi.org/10.3390/met10050641
4. Li C, Yuan G, Fet Ji, Kang J, Misra D, Wang GD (2015) Mechanism of microstructural control and mechanical properties in hot rolled plain C-Mn steel during controlled cooling. ISIJ Int 55:1721–1729. https://doi.org/10.2355/isijinternational.ISIJINT-2015-040
5. Shi J, Yuan G, Jiang LY, Li ZL, Zhao K, Wang GD (2015) Heat transfer symmetry of the strip surface due to agroup oblique slot jet impingement after hot rolling. Steel Res Int 86:1548–1557. https://doi.org/10.1002/srin.201400529
6. Timm W, Weinzierl K, Leipertz A, Zieger H, Zouhar G (2002) Modelling of heat transfer in hot strip mill run out table cooling. Steel Res 73:97–104. https://doi.org/10.1002/srin.200200180
7. Li HJ, Li ZL, Yuan G, Wang ZD, Wang GD (2013) Development of new generation cooling control system after rolling in hot rolled strip based on UFC. J Iron Steel Res Int 20:29–34. https://doi.org/10.1016/S1006-706X(13)60122-3
8. Gong DY, Xu JZ, Cai XH, Zheng H, Liu XH, Wang GD (2007) Model building of finishing temperature control for accelerative hot strip rolling. Key Eng Mater 340–341:633–638. https://doi.org/10.4028/www.scientific.net/kem.340-341.633
9. Xie HB, Jiang ZY, Liu XH, Wang GD, Zhou TG, Tieu AK (2007) On-line optimization of cooling temperature control on run-out table for hot strip mills. Key Eng Mater 340–341:701–706. https://doi.org/10.4028/www.scientific.net/kem.340-341.701
10. Latzel S (2001) Advanced automation concept of run out table strip cooling for hot strip and plate mills. IEEE Trans Ind Appl 37:1088–1097. https://doi.org/10.1109/28.936401
11. Zhang DZ, Ye HL, Xiang XF (2010) Coiling temperature optimal setting control model based on genetic algorithms and application in hot strip rolling mill. 2010 International Conference on Electrical and Control Engineering. IEEE. https://doi.org/10.1109/ICECE.2010.151
12. Zheng Y, Li N, Li S (2013) Hot-rolled strip laminar cooling process plant-wide temperature monitoring and control. Control Eng Pract 21:23–30. https://doi.org/10.1016/j.conengprac.2012.09.004
13. Xie HB, Liu XH, Wang GD, Zhang ZP (2006) Optimization and model of laminar cooling control system for hot strip mills. J Iron Steel Res Int 13:18–22. https://doi.org/10.1016/S1006-706X(06)60019-8
14. Li S, Li X, Yang Y (2016) Intelligent model building and GPC-PID based temperature curve control strategy for metallurgical industry. Math Probl Eng. https://doi.org/10.1155/2016/745805
15. Li S, Zhou C, Deng ZH, Li X (2015) Application of genetic algorithm and least squares support vector machines in laminar cooling process. J Comput Theor Nanosci 12:1467–1471. https://doi.org/10.1166/jctn.2015.3915
16. Li S, Li X, Deng ZH (2014) A new kind of model of laminar cooling: by ls-svm and genetic algorithm. 9th International Conference on Bio-Inspired Computing-Theories and Applications (BIC-TA). Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-662-45049-9_41
17. Li S, Li X, Deng ZH (2015) A T-S fuzzy model-based intelligent temperature prediction model of laminar cooling system. 2015 Chinese automation congress (CAC) https://doi.org/10.1109/CAC.2015.7382685
18. Zheng Y, Li S, Wang XB (2011) Horizon-varying model predictive control for accelerated and controlled cooling process. IEEE
Trans Ind Electron 58:329–336. https://doi.org/10.1109/TIE.2010.2045321

19. Zheng Y, Li S, Wang X (2009) Distributed model predictive control for plant-wide hot-rolled strip laminar cooling process. J Process Control 19:1427–1437. https://doi.org/10.1016/j.jprocont.2009.04.012

20. Li HJ, Li LG, Li YL et al (2015) Online monitor and control of cooling temperature on run-out table of hot strip mill. Steel Res Int 86:1225–1233. https://doi.org/10.1002/str.201400269

21. Chen XL, Wang GD, Tian Y, Wang BX, Guo Y, Wang ZD (2014) An on-line finite element temperature field model for plate ultrafast cooling process. J Iron Steel Res Int 21:481–487. https://doi.org/10.1016/S1006-706X(14)60075-3

22. Li ZL, Li HJ, Yuan G, Wang GD, Wang XQ (2015) Research and application of ultra-fast cooling system and velocity controlled strategy for hot rolled strip. Steel Res Int 86:478–488. https://doi.org/10.1002/str.201400777

23. Zheng Y, Li S, Wang X (2010) An approach to model building for accelerated cooling process using instance-based learning. Expert Syst Appl 37:5364–5371. https://doi.org/10.1016/j.eswa.2010.01.020

24. Zhang T, Xie Q, Wang B, Wu ZQ, Wang ZD, Wang GD (2016) A novel variable scale grid model for temperature self-adaptive control: an application on plate cooling process after rolling. Steel Res Int 87:1213–1219. https://doi.org/10.1002/str.201500344

25. Schlang M, Broese E, Feldkeller B, Granckow O; Jansen M, Pappe T; Schaffner C, Sorgel G (1997) Neural networks for process control in steel manufacturing. 1997 IEEE International conference on Acoustics, Speech, and Signal Processing(ICASSP). IEEE. https://doi.org/10.1109/ICASSP.1997.599582

26. Liu EY, Peng W, Cao N, Yu SR, Xu J, Peng LG, Zhang DH (2014) Prediction of cooling temperature of hot rolled strip based on BP neural network. Appl Mech Mater 633–634:679–683. https://doi.org/10.4028/www.scientific.net/AMM.633-634.679

27. Deng J, Sun J, Peng W, Hu YH, Zhang DH (2019) Application of neural networks for predicting hot-rolled strip crown. Appl Soft Comput J 72:119–131. https://doi.org/10.1016/j.asoc.2019.02.030

28. Colla V, Vannucci M, Dimatteo A (2010) Diagnosis of the instability of the cooling behaviour of flat steel products through parametric characterisation, neural networks and statistics. ISA Trans 49:235–243. https://doi.org/10.1016/j.isatra.2010.01.001

29. Xing G, Ding J, Chai TY, Afshar P, Wang H (2012) Hybrid intelligent parameter estimation based on grey case-based reasoning for laminar cooling process. Eng Appl Artif Intell 25:418–429. https://doi.org/10.1016/j.engappai.2011.10.007

30. Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323:533–536. https://doi.org/10.1038/323533a0

31. Connor JT, Martin RD, Atlas LE (1994) Recurrent neural networks and robust time series prediction. IEEE Trans Neural Networks 5:240–254. https://doi.org/10.1109/72.279188

32. Gregor K, Danihelka I, Graves A, Rezende DJ, Wierstra D (2015) DRAW: a recurrent neural network for image generation. 2015 Proceedings of The 32nd International Conference on Machine Learning(ICML) 37: 1462–1471. http://proceedings.neurips. org/v3/2015/day-1/382

33. Graves A, Mohamed AR, Hinton G (2013) Speech recognition with deep recurrent neural networks, 2013 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP). IEEE. https://doi.org/10.1109/ICASSP.2013.6638947

34. Sak H, Senior A, Beaufays F (2014) Long short-term memory recurrent neural network architectures for large scale acoustic modeling. 2014 15th Annual Conference Of The International Speech Communication Association (INTERSPEECH). ISCA. 2014:338–342. https://doi.org/10.21437/Interspeech.2014-80

35. Graves A, Schmidhuber J (2005) Framewise phoneme classification with bidirectional LSTM and other neural network architectures. Neural Netw 18:602–610. https://doi.org/10.1016/j.neunet.2005.06.042

36. JC Heck FM Salem (2017) Simplified minimal gated unit variations for recurrent neural networks. 2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS) IEEE https://doi.org/10.1109/MWSCAS.2017.8053242

37. Dey R, Salient FM (2017). Gate-variants of Gated Recurrent Unit (GRU) neural networks. 2017 IEEE 60th 60th international Midwest Symposium on Circuits and Systems (MWSCAS). IEEE. https://doi.org/10.1109/MWSCAS.2017.8053243

38. Li L, Wen Z, Wang Z (2016) Outlier detection and correction during the process of groundwater lever monitoring base on pauta criterion with self-learning and smooth processing. Commun Comput Inf Sci 643:497–503. https://doi.org/10.1007/978-10-2663-8_51

39. Thireou T, Reczko M (2007) Bidirectional long short-term memory networks for predicting the subcellular localization of eukaryotic proteins. IEEE-ACM Transactions on computational biology and bioinformatics 4:441–446. https://doi.org/10.1109/TCBB.2007.1015

40. Lee SY, Tama BA, Choi C et al (2020) Spatial and sequential deep learning approach for predicting temperature distribution in a steel-making continuous casting process. IEEE Access 8:21953–21965. https://doi.org/10.1109/ACCESS.2020.2969498

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.