

A multi-channel cross-residual deep learning framework for news-oriented stock movement prediction

Duo Xu, Zeshui Xu, Shuixia Chen & Hamido Fujita

To cite this article: Duo Xu, Zeshui Xu, Shuixia Chen & Hamido Fujita (2023) A multi-channel cross-residual deep learning framework for news-oriented stock movement prediction, Economic Research-Ekonomiska Istraživanja, 36:2, 2106271, DOI: [10.1080/1331677X.2022.2106271](https://doi.org/10.1080/1331677X.2022.2106271)

To link to this article: <https://doi.org/10.1080/1331677X.2022.2106271>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 10 Aug 2022.



Submit your article to this journal [↗](#)



Article views: 620



View related articles [↗](#)



View Crossmark data [↗](#)

A multi-channel cross-residual deep learning framework for news-oriented stock movement prediction

Duo Xu^a, Zeshui Xu^b, Shuixia Chen^b and Hamido Fujita^c

^aBusiness School, Stevens Institute of Technology, Hoboken, NJ, USA; ^bBusiness School, Sichuan University, Chengdu, China; ^cRegional Research Center, Iwate Prefectural University, Iwate, Japan

ABSTRACT

Stock market movement prediction remains challenging due to random walk characteristics. Yet through a potent blend of input parameters, a prediction model can learn sequential features more intelligently. In this paper, a multi-channel news-oriented prediction system is developed to capture intricate moving patterns of the stock market index. Specifically, the system adopts the temporal causal convolution to process historical index values due to its capability in learning long-term dependencies. Concurrently, it employs the Transformer Encoder for qualitative information extraction from financial news headlines and corresponding preview texts. A notable configuration to our multi-channel system is an integration of cross-residual learning between different channels, thereby allowing an earlier and closer information fusion. The proposed architecture is validated to be more efficient in trend forecasting compared to independent learning, by which channels are trained separately. Furthermore, we also demonstrate the effectiveness of involving news content previews, improving the prediction accuracy by as much as 3.39%.

ARTICLE HISTORY

Received 24 November 2021
Accepted 21 July 2022

KEYWORDS

Stock market prediction; deep learning; temporal causal convolution; cross-residual learning; information fusion

JEL CODES

A1; C4; C5; G1

1. Introduction

A time series exhibits a wide variety of features whose historical properties and future evolution are worth investigating (Mills, 2019). Time series prediction is perceived to be an essential technique in many real-world applications. Among them, financial markets have been the focus of much research to date given its importance in facilitating economic development (Hsu et al., 2016). Individuals and organisations can be well-supported to make rational investment decisions and manage financial and operational risks efficiently. Yet, the prediction task remains challenging due to the dynamic, nonlinear, nonstationary, nonparametric properties of the financial market (Abu-Mostafa & Atiya, 1996). To grasp such volatile nature, researchers in many

CONTACT Zeshui Xu  xuzeshui@scu.edu.cn

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

disciplines focus on developing intelligent techniques for effective financial market forecasting, in particular the stock market.

Early approaches to stock market behaviour prediction were primarily based on statistical techniques pre-assuming linearity, stationarity, and normality (Shah et al., 2019). These conventional methods, such as autoregressive integrated moving average (ARIMA), work for a particular time series but fail to capture the nonlinear characteristics in general (Zhang, 2003). The emergence of artificial intelligence helps overcome such limitation and deep neural network models are being explored by recent research, in various financial and economic applications, aiming to reach the full potential (Li, Xu, et al., 2020; Liu et al., 2021; Long et al., 2019; Zheng et al., 2021).

Although such deep learning (DL) approaches perform well according to the prediction accuracy, they, to some extent, ignore the fact that the market also reacts arbitrarily to unexpected external events. Stock market forecasting solely based on quantitative financial indicators can be vulnerable to economic and social changes. There has been an increasing trend towards integrating heterogeneous sources of information to improve prediction efficiency. Such data sources, e.g., financial news and social media feeds, can complement valuable information that is not reflected in the intrinsic historical market values.

1.1. Literature review

The most extensively used DL model architectures dealing with financial time series are recurrent neural network (RNN) and convolutional neural network (CNN). RNN, mainly its variants, has been dominantly adopted in stock market prediction due to its powerful capability in sequence processing (Ghosh et al., 2022; Kim & Kang, 2019). It forms an information loop to explore the temporal dynamic patterns, allowing learning from the previously received input features. Emerged from the visual cortex structure, CNN has exhibited its strength in feature extraction. Compared to RNN, CNN benefits from a lower number of parameters by filtering with the kernel window function. One way to apply CNNs to financial forecasting is to convert the selected predictors into a time-series graph (Sim et al., 2019). Besides, the prediction task can also be implemented with simple and compact 1D CNNs, realising a lower implementation cost than feeding two-dimensional image data (Cavalli & Amoretti, 2021).

Yet, Dingli and Fournier (2017) indicated that state-of-the-art classifiers (e.g., LR, SVM) yield slightly better outputs than vanilla 1D CNNs, which thus demand further configurations to surpass the techniques above. For example, since the output at a specific time is conditioned only on the past input samples in financial prediction tasks, a temporal convolution network (TCN) can be a suitable candidate. TCNs introduce dilated causal convolutions (DCC) and residual connections to a simple 1D CNN to maintain such causality. Based on an empirical evaluation (Bai et al., 2018), a TCN is substantially superior to LSTMs and GRUs in capturing long-term dependencies. It also presents longer effective memory and higher execution speed (Borovykh et al., 2019; Wiese et al., 2020).

More than half of the recent existing DL implementations in the financial domain have focussed on stock price or stock market index (Sezer et al., 2020). In addition to

Table 1. Recent research on stock price or index forecasting using financial news.

Literature	Dataset	Period	Variables	Lag	Horizon	Method
Ding et al. (2015)	S&P500 index, 15 individual stocks from S&P500, news from Reuters and Bloomberg	2006–2013	price & index data, technical indicators, news titles	–	–	CNN
Vargas et al. (2017)	S&P500, news from Reuters and Bloomberg	2006–2013	index data, technical indicators, news titles	1d	1d	RCNN
Chen et al. (2018)	HS300, social media news from Sina Weibo	2015–2017	index data, news contents	1d	1d	RNN boost with LDA
Hu et al. (2018)	2527 individual stocks, news from Sina Finance, and Eastmoney	2014–2017	price data, news titles & contents	–	–	Bi-GRU
Daiya et al. (2020)	S&P500 index, 15 individual stocks from S&P500, news from Reuters and Bloomberg	2006–2013	price & index data, technical indicators, news titles	40d	1d	CNN
Li, Wu, et al. (2020)	HKEx (HSI), FINET news	2003–2008	price data, news contents	–	–	LSTM
Ren et al. (2020)	18 individual stocks, news from Sina Finance, and Oriental Wealth Network	2013–2017	price data, news titles	–	–	Bi-LSTM
Li and Pan (2022)	S&P500, news from CNBC, Reuters, WSJ, and Fortune	2017–2018	index data, news titles	10d	1d	LSTM, GRU
Xu et al. (2022)	240 Chinese stocks, news from Eastmoney, and Tecent Finance	2016–2019	price data, movement data, news titles	–	5d	Bi-GRU

Source: Authors.

its historical patterns, the stock market is also highly influenced by certain public events, which can be extracted from online news articles. Table 1 lists the implementation details of recent stock price (index) forecasting works using financial news data. In Table 1, *Period* refers to the training and testing period, and *Variables* lists different input features that are fed into the model. *Lag* denotes the time length of the input vector, and *Horizon* denotes the time length of the output vector to be predicted.

As illustrated in Table 1, news-oriented stock market forecasting has been a popular research area. RNN and its variants, including LSTM and GRU, are still dominant for time series data processing. Hybrid architectures that maximise the strengths of different models are also attractive. More studies prefer a 2- to 5-year research period due to the high computational costs for collecting online news articles over an extended period. In this study, the input data length is around five years, sufficient for producing satisfactory results. Regarding the qualitative datasets, news from Reuters and Bloomberg, released by Ding et al. (2015), is a popular news dataset but no longer publicly available for research purposes. As a result, the news data collection relies on requiring from legit agencies or web scraping. In addition, most of the literature utilised news headlines solely as the textual data source. It is believed that highly concentrated titles can convey the most indicative information, and incorporating news content risks adding irrelevancies. Although price prediction is essentially

a regression problem, a correct directional movement is perceived as a more crucial research area. In this sense, trend prediction becomes a classification problem, taking upward and downward price movements into account.

1.2. Contribution

When designing DL prediction models, three factors play an essential role in the model accuracy, namely network design, input feature, and quantity of training data (Walczak, 2001). Most of the innovations for deep prediction models are embodied in the network design. Considering CNNs as a natural starting point for sequence modelling, this paper develops a novel DL architecture, namely a multi-channel cross-residual TCN (mc-CRTCEN), to forecast the stock market behaviour. The proposed prediction system incorporates stock market data and financial news to handle unexpected abrupt changes or unexplainable extreme fluctuations. The representation of both news headlines and abstracts is obtained via the Transformer Encoder (TE). A few studies (Liu et al., 2019; Tang et al., 2018) have shown the superiority of TE to CNN- and RNN- based models for deep semantic features extraction.

Given multimodal information, a vital issue is concerning how they are processed and fused. Existing knowledge- or event-driven stock market prediction has already utilised the multi-channel system, with one channel responsible for processing stock price data and others dealing with external stock-related tasks (Chung & Shin, 2020; Nti et al., 2021; Zhang et al., 2018). Generally, these related channels are trained simultaneously but separately and then concatenated at the end for an information fusion. This paper presents a different fusion method by jointly training these channels through cross-connections, by which a stronger coupling of multimodal information is realised at an earlier stage. The cross-residual learning (CRL) technique, pioneered by Jou and Chang (2016), has been integrated in visual recognition and pattern classification tasks (Lyu et al., 2020). According to these existing applications, CRL is regarded as an in-network regularisation, possessing greater generalisation ability.

Since an accurate prediction of the exact stock index value can be highly challenging, a classification problem is formulated instead, aiming at predicting the specific movement direction for the next trading day according to relevant information extracted from the financial market news. Our contributions can be summarised as:

1. A hybrid configuration to the baseline model is presented using TCN and TE. It effectively enhances the performance of vanilla 1D CNNs in stock market prediction.
2. Both sentence- and paragraph-level features are extracted via TE to capture indicative information from publicly available news data. The influence of adding content previews or description to the news input channel is evaluated in the comparison analysis.
3. A CRL procedure is implemented through cross-connections between multiple channels. The primary motivation is to closely couple multimodal information at an earlier stage to reveal cross-task dependencies. A modification to the original CRL is proposed for stock market forecasting.

The remainder of this study is constructed as follows. Section 2 introduces the relevant preliminaries. The proposed prediction strategy based on mc-CRTCNN is elaborated in Section 3. Further, Section 4 describes the data collected for the prediction model and presents the numerical experimental results and their implications. Finally, Section 5 concludes this study and prompts several future research directions.

2. Preliminaries

2.1. Temporal convolution network

Using multiple DCC layers, TCN differentiates itself from standard 1D CNNs when handling long sequences. Causal convolution was initially proposed as the main ingredient of WaveNet (Oord et al., 2016), attempting to avoid the involvement of future information. In other words, it helps inspect that the channel only considers historical index values, as required in financial time series prediction. For a convolution operator to be causal, it must involve causality described as:

$$p(X) = \prod_{t=1}^T p(x_t | x_1, x_2, \dots, x_{t-1}) \quad (1)$$

where $p(X)$ represents the maximum likelihood given the observed past index values and T denotes the total length of the observation period.

Financial market forecasting normally desires a full coverage of historical patterns to improve model robustness and accuracy. In this case, it necessitates a deeper network or a larger filter (also kernel) size, neither of which is feasible for prediction with very long sequences. The use of dilated convolution enables TCN to efficiently address this problem without adding computational complexity. The dilated rate d is introduced to represent the spacing between the skipped input values in each convolutional layer. Since d is doubled every layer up, TCN eventually procures an exponential growth in the receptive field without losing resolution or coverage (Yu & Koltun, 2015). Formally, DCC on element s from a sequential input series $x \in \mathbb{R}^n$ with a filter $f : \{0, \dots, k-1\}$ is defined as:

$$(x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (2)$$

where $*_d$ denotes the dilated convolutional operand and $s - d \cdot i$ indicates the past direction. In addition, a TCN yields an output tensor of the same dimension as the input tensor, thus zero-padding is applied on every subsequent layer to secure the equal length of the previous layer.

2.2. Transformer encoder

The textual features are captured through the TE, entirely based on self-attention without recurrence or convolution. The TE stacks a certain number of multi-head attention (MHA), where the self-attention is implemented, and pointwise fully connected

feed-forward layers (Vaswani et al., 2017). Each input embedding $X \in \mathbb{R}^d$ has three different vectors, namely query Q , key K and value V , which are linearly projected h times with learnable parameter matrices $W_q \in \mathbb{R}^{d \times d_k}$, $W_k \in \mathbb{R}^{d \times d_k}$, and $W_v \in \mathbb{R}^{d \times d_v}$. The scaled dot-product attention function performed in the MHA can be described as:

$$Q^{(h)}(x_i) = W_{h,q}^T x_i, K^{(h)}(x_i) = W_{h,k}^T x_i, V^{(h)}(x_i) = W_{h,v}^T x_i, \quad (3)$$

$$\alpha_{i,j}^{(h)} = \text{softmax}_j \left(\frac{\langle Q^{(h)}(x_i), K^{(h)}(x_j) \rangle}{\sqrt{d_k}} \right) \quad (4)$$

$$\text{Attn}(Q^{(h)}, K^{(h)}, V^{(h)}) = \alpha_{i,j}^{(h)} V^{(h)} \quad (5)$$

where $Q^{(h)}(x_i)$, $K^{(h)}(x_i)$, $V^{(h)}(x_i)$ denote the vector representation of the i^{th} input token in the head h . $\langle Q^{(h)}(x_i), K^{(h)}(x_j) \rangle$ defines the compatibility of x_i with x_j in the head h and is normalised to compute the attention weight. Equation (5) defines the self-attention, a weighted sum of values. Therefore, MHA can be obtained stacking h parallel layers, or attention heads, given by

$$\text{MultiHead}(X) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_o \quad (6)$$

where $\text{head}_i = \text{Attn}(Q^{(i)}, K^{(i)}, V^{(i)})$ and $W_o \in \mathbb{R}^{hd_v \times d}$.

Such architecture weights each word in the news pieces according to its attention weight through training the query and key matrices of the attention layers to determine the usefulness of different words, namely the attention weight. The attention mechanism was initially integrated into sequence-to-sequence models (i.e., RNNs) for a selective concentration on relevant information. Compared to the standard attention mechanism, self-attention creates a shorter path between any combination of either distant input or output positions. Thus, it is more efficient to compute, parallelise, and capture long-range dependencies, all of which are crucial for contextual embedding (Mishev et al., 2020).

2.3. Cross residual learning

The proposed architecture integrates heterogeneous input sources by adding a CRL module between different data types. Initially, a residual learning framework is introduced to deep neural networks to address the degradation problem and ease the training process (He et al., 2016), given by

$$y = F(x, W_i) + W_S x \quad (7)$$

where F indicates the residual mapping fit by stacked convolutional layers, and W_i is the weight associated with the i^{th} layer. Since $F + x$ is implemented through a shortcut connection and element-wise addition, F and x must have the same dimension. Otherwise, a linear projection W_S is required on the shortcut connection to match the dimension.

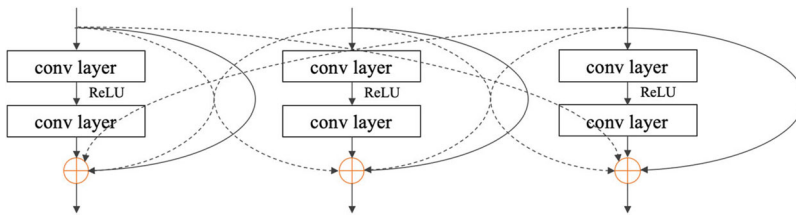


Figure 1. A sample cross-residual module with three tasks.
Source: Authors.

Jou and Chang (2016) put forward an extension that enables intuitive learning across domains in the field of visual recognition, aiming at building a more versatile multimodal network that leverages cross-task dependencies. Given a target task t and $N - 1$ other related tasks, a cross-residual module is defined as:

$$y^{(t)} = F\left(x, \left\{W_i^{(t)}\right\}\right) + \sum_{j=1}^{N-1} W_S^{(j)} x \quad (8)$$

where the superscript (\cdot) denotes the task index and $j \neq t$. It comprises the residual learning for t and the additive contribution of j to t . Figure 1 illustrates a CRL module between multiple related tasks, where solid shortcuts represent identity mapping and dashed shortcuts represent cross-residual connections. It serves as a form of in-network regularisation by biasing at a layer level and enables greater network generalisation.

3. The proposed prediction strategy

This research develops a novel news-driven prediction system for the stock market trend. Its main novelty lies in how these different input data are fused during the training procedure to output better model performance. In general, the current state of the stock market can be influenced not only by its past performance but also by the latest publicly-released news. That is, dealing with the stock index values and the news pieces are considered as related tasks. The proposed model adopts different channels to handle feature extraction for these related tasks, respectively. An overview of the multi-channel setting is illustrated in Figure 2.

Different types of representation are processed simultaneously and eventually concatenated to a fully connected layer followed by Softmax activation. Notably, a cross-residual module is highlighted with dashed lines. It helps to select textual information that is highly relevant to the stock movement and to investigate the interrelatedness between multiple sources of input data by leveraging cross-task dependencies. In this setting, an earlier integration is realised between channels, attempting to avoid information loss, thus to improve the prediction performance for the stock market behaviour.

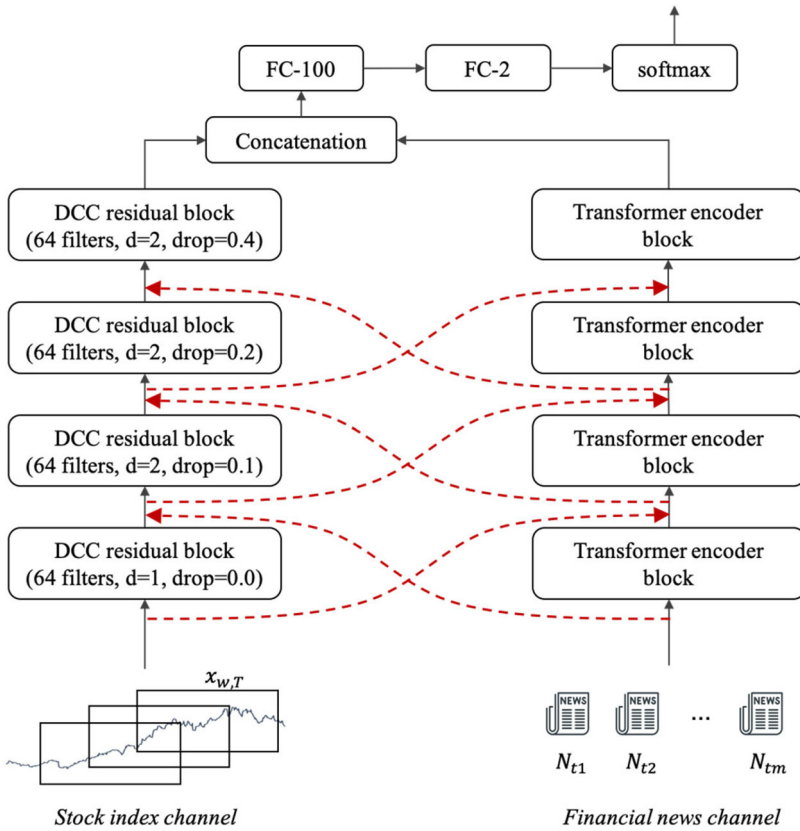


Figure 2. An overview of the proposed multi-channel cross-residual model architecture. Source: Authors.

3.1. Stock index channel

When processing sequential data, a 1D feature map is produced per kernel by sliding several kernels across the series. That is, within a fixed window of the length w , the partial sequential pattern is detected. The observed stock market value $x_{w,T} = (x_{t-w+1}, x_{t-w+2}, \dots, x_t)$ is encoded to an index vector $V_{x_{w,T}}$ to represent the features captured in the latest time window.

The stock index channel employs four DCC blocks (as in Figure 3). Each block consists of two stacked DCC layers, followed by the weighted normalisation to counteract gradient explosions, SELU to introduce the non-linearity, and dropout with skip connections to prevent the DCC from overfitting. Although ReLU is predominantly adopted in CNNs to realise lower error rates, Daiya et al. (2020) suggested that SELU performed better when the output needed to preserve negative values (i.e., negative stock movement). The drop rates remain the same for DCC layers in the same residual block but differ among blocks to ensure the latent representations learned are not unduly influenced by past patterns far back in the time series. Specifically, a dropout rate of 0.0, 0.1, 0.2, and 0.4 is assigned.

In the DCC residual block, an 1×1 convolution is optional and added only when the residual input and output are of discrepant dimensions. Besides dilated

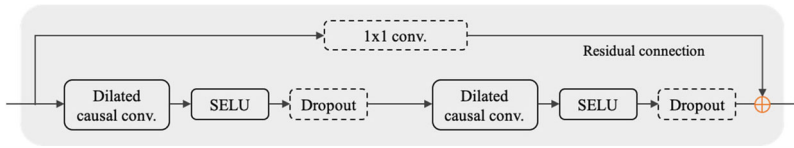


Figure 3. A DCC residual block.

Source: Authors.

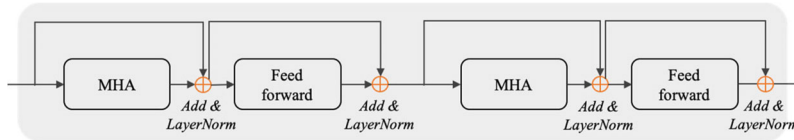


Figure 4. A TE block.

Source: Authors.

convolutions, residual connections can also effectively capture long-range temporal patterns (Bai et al., 2018). The output of an identity mapping, the mapping performed by skipping a certain number of layers, is added to the output of stacked DCC layers to handle the degradation problem exposed with a deeper network structure.

3.2. Financial news channel

On the textual input side, every single piece of news is encoded to derive the news representation. Several existing studies (Ding et al., 2015; Radinsky et al., 2012) indicated news headlines are more informative than news contents, since covering the entire content may introduce irrelevancies that degrades the prediction accuracy. Suppose that there are in total m pieces of news data published on the target date t , the channel obtains the news vector $E(h_{t,1}, \dots, h_{t,m}) = V_{N_t}$, where $h_{t,j}$ ($j \in [1, m]$). The news channel stacks four standard encoder blocks (as in Figure 4), using three self-attention heads for MHA and ReLU activation for position-wise fully connected feed-forward layers with no dropout. A Mask function is adopted in the attention layer to maintain the causality.

Although the full length of news content can influence the prediction results reversely, a more condensed abstract may not. In addition to the news headlines, the corresponding preview texts are encoded to examine its effect on stock forecasting. Given a paragraph set $\text{para} = \{s_1, \dots, s_{N_s}\}$, where N_s denotes the number of sentences in the preview content, the channel obtains another news vector $E(\text{para}_{t,1}, \dots, \text{para}_{t,m}) = V_{N_t}^*$ whose length is fixed at 80. To avoid significant information loss during compressing too large the input size, V_{N_t} is not spliced with $V_{N_t}^*$ when these vectors are fed into the Transformer. Therefore, two independent TE channels are adopted for news headlines and preview texts, respectively. The whole process is realised through PyTorch 1.10.0. For the model initialisation, the channel utilises the word embedding pre-trained on Google news using Word2Vec, and fine-tunes the word vector along with the model.

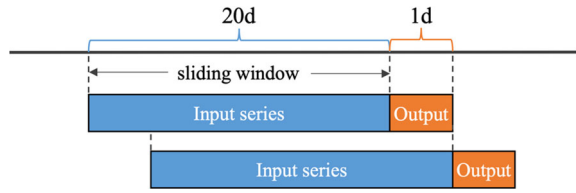


Figure 5. A schematic diagram of the sliding window.

Source: Authors.

3.3. Channel concatenation

The proposed model leverages an attention mechanism to investigate the interrelatedness between multiple channels by selecting textual information that is highly relevant to stock movements. A cross-residual module is highlighted in this architecture that replaces the self-residual unit in a conventional network. Compared to the setting proposed by Jou and Chang (2016), some modifications are made when the module is applied to the stock market, given by

$$V_{x_w, T}^{(l+1)} = F(V_{x_w, T}^{(l)}) + W_S^1 V_{N_t}^{(l)} \quad (9)$$

$$V_{N_t}^{(l+1)} = F(V_{N_t}^{(l)}) + W_S^2 V_{x_w, T}^{(l)} \quad (10)$$

where the superscript l denotes the block (or layer) index and F denotes the residual learning function. W_S^j refers to the linear projection when the news representation has a different dimension from the index representation. The training, validation, and testing split is 70%, 15%, and 15%, respectively. The training process comprises both independent- and cross-learning, which primarily deals with parameter fine-tuning, i.e., we train one channel while maintaining the parameters fixed of the other channel. As a result, there are relatively fewer training parameters, thereby improving the learning efficiency and generalisation. The model is trained for 100 epochs with a batch size of 32, and the initial learning rate decays linearly by 0.1 every 10,000 iterations.

The model employs an observation window w of 20 days, meaning the index data from past 20 trading days are required, to capture a relatively large receptive field. In addition, the financial news published only on the last day of w is collected since historical events normally influence the stock market less than the latest news. The target horizon is one next trading day. By sliding w (as in Figure 5), the sample predicted each time is fed back into the proposed network to predict the next index sample.

4. Experiment, results and discussion

4.1. Data description

The stock market index is selected as the input variable (typically computed as the weighted arithmetic mean of the prices of different selected stocks) since it is less volatile than a single stock and more indicative of a general state of the national economy. S&P 500 (Standard & Poor's 500) is one of the most commonly used index data when predicting the stock market index. It covers 500 listed companies in the

U.S. that are highly representative of the rise and fall of the national economy. The closing price of S&P 500 are collected from Yahoo Finance during November 2015 and October 2019. There are in total $N_t = 1003$ index instances collected. Considering two output categories only (i.e., either positive or negative), the input data distribution is relatively balanced (54.14% vs. 45.86%). There is only one instance that experienced no change in value, which is excluded for simplicity.

Meanwhile, financial news that are publicly released only on trading days are also collected over this period. Both news headlines and preview texts are scraped from Reuters and CNBC (Consumer News and Business Channel), two of the most prominent multimedia agencies that provide an expansive range of real-time business and financial news. A total of 68,253 pieces of news is collected from target websites; each piece is marked with the last updated date, headlines, and description. The punctuation and stop-words are removed. To limit the length of the input texts, zero padding is applied on the right side of the headline and preview input tensor that fails to reach the fixed size of 20 and 80. This setting ensures the implementation of cross-connections through linear projection.

4.2. Performance metrics

In the presented binary classification problem, the confusion matrix is utilised to assess the model quality. Confusion matrix, also error matrix, is another typical criterion for statistical classification in supervised learning, not limited to binary categories. The most basic terms in a confusion table are true positives (*TP*), true negatives (*TN*), false positives (*FP*), and false negatives (*FN*). *TP* indicates both predicted and actual values are positive, and *TN* indicates both are negative. *FP* represents positive prediction but negative observation, and it also refers to as *Type I* error related to overestimation. Conversely, *FN* refers to as *Type II* error related to underestimation.

The confusion matrix derives four evaluation indicators, namely accuracy, precision, recall, and *F1*-score. *Accuracy* is one of the most intuitive measures that explains the consistency between the predicted output and the actual level. Nevertheless, *Accuracy* may produce misleading results if the dataset is asymmetric (i.e., *FP* significantly differs from *FN*). Besides accuracy, *F1*-score is also adopted. It represents the weighted average of *Precision* and *Recall*, taking both *FP* and *FN* into account. When dealing with imbalanced datasets, *F1*-score is usually more beneficial than *Accuracy* by penalising extreme values.

$$\text{Acc} = (TP + TN)/(TP + FP + FN + TN) \quad (11)$$

$$\text{Precision} = TP/(TP + FP) \quad (12)$$

$$\text{Recall} = TP/(TP + FN) \quad (13)$$

$$\text{F1-score} = (2 \times \text{Precision} \times \text{Recall})/(\text{Precision} + \text{Recall}) \quad (14)$$

4.3. Model performance

Performance of the proposed mc-CRTCEN is evaluated in three progressive aspects: (i) an assessment of the basic TCN architecture, (ii) the effect of different residual learning patterns, i.e., CRL and independent learning (IL), and (iii) the effect of involving news previews in the news channel. The numerical results are primarily shown in the form of comparative analysis. In addition, other than a binary classification problem, different output settings are applied to examine the effectiveness and robustness of the proposed mc-CRTCEN.

4.3.1. TCN vs. baseline models

The stock index channel that deals with market behaviour itself is the base of the proposed framework. Considering 1D CNN as a natural starting point, the framework extracts such intrinsic historical features through the generic TCN architecture. In order to demonstrate its strength in financial time series forecasting, a comparison analysis is implemented between TCN and some conventional prediction models, including ARIMA, LSTM, and CNN. The stochastic gradient descent (SGD) is employed to optimise the training process, with a kernel size of 2 and residual blocks of 10. All the settings for the input data keep the same.

These baseline models are evaluated only with the stock index data, and the results are shown in Table 2. The generic ARIMA model based on a linear assumption shows its weakness in processing dynamic stock market data. Deep neural network models achieve better performance. In terms of accuracy, TCN greatly outperforms other baseline models on the stock trend prediction over S&P500 index, showing an obvious advantage as a basic component of the proposed framework. Compared to vanilla CNN, TCN obtains superior outcomes as it benefits from stacked DCCs. Therefore, DCC neural network can be an ideal model structure for stock market forecasting.

4.3.2. CRL vs. IL

In addition, since this paper primarily emphasises a different residual pattern among various processing channels, another comparison analysis is conducted between normal and cross residual learning. The index prediction results are contrasted with and without cross-connections when using news headlines only. It helps to assess the effect of adding a CRL module between channels. The model performance is then evaluated with an inclusion of previews texts (or descriptions) of news contents. Notably, when feeding both headlines and descriptions to the proposed model, three channels are established, one TCN channel dealing with the stock market index and two TE channels extracting news representation.

Table 2. Stock trend prediction with different baseline models.

	Accuracy (%)	Precision (%)	Recall (%)	F1-score
ARIMA	53.80	39.17	42.92	0.4096
LSTM	59.25	56.66	56.12	0.5639
CNN	56.87	53.46	55.68	0.5455
TCN	62.72	54.58	57.10	0.5581

Note: The highest values of Accuracy, Recall, and F1-score achieved by the baseline models are shown in bold.
Source: Authors' calculations.

Table 3. Stock trend prediction results with different residual learning pattern using different set of news inputs.

	Accuracy (%)	Precision (%)	Recall (%)	F1-score
IL (headlines only)	64.62	59.51	68.70	0.6381
CRL (headlines only)	65.25	59.16	69.59	0.6396
IL (headlines & previews)	67.68	61.41	72.02	0.6630
CRL (headlines & previews)	68.64	61.60	73.15	0.6688

Note: The highest values of Accuracy, Recall, and F1-score achieved using different learning patterns and news input sets are shown in bold.

Source: Authors' calculations.

Table 3 summarises the prediction results. In Table 3, IL refers to the concatenation procedure without CRL, in which stock index and news headlines are trained independently and eventually combined to a fully connected layer. Based on the confusion matrix-relevant measurements, we draw the following findings:

1. Overall, the proposed prediction model that applies CRL using both news headlines and preview texts as the external inputs delivers the most satisfying results.
2. Using the same set of input variables, the model with CRL procedures outperforms the model with IL in Accuracy and F1-score. The measurement increases by 0.63% and 0.0015, respectively, when employing headlines solely.
3. The increments in (2) are not significant mainly because the proposed model is evaluated benchmarking a comparative mature DCC-TE setting. In other words, this paper primarily assesses the effect of cross-connections, which indeed improves the prediction accuracy. These increases are still appreciable for stock market prediction due to the highly uncertain properties of the financial market.
4. An enhancement of model performance presents in every evaluation criterion when feeding headlines and descriptions into TEs, compared to feeding the news titles only. An average increase is observed of 3.39% and 0.0294, respectively, in Accuracy and F1-score. Therefore, if available, the news preview text or description is a suitable input candidate when employing financial news for stock market prediction.

4.3.3. Illustration of news previews

A few sample news instances are tabulated below to exemplify (4) above, demonstrating how preview texts provide [supplementary information](#) relevant to the stock market. In Table 4, bold texts indicate key information of the above news headlines. Bold underlined texts represent information that is parallel to the headlines. Bold italic texts represent information that is additional to the headlines. All key words are selected from a subject perspective. Among the five selected pieces of news, preview texts of news (a) and (b) provide repetitive messages. This type of contents risks lower prediction accuracy as it may bring in more irrelevancies. On the other hand, descriptions of news (c) to (e) demonstrate additional information that is not revealed from the headlines, which further facilitate the prediction tasks. Based on this study, despite a potential risk of input redundancies, the involvement of the concentrated descriptive contents still efficiently improves the model performance.

Table 4. Sample news instances in the dataset collected from Reuters.com.

Headlines	Preview texts
(a) Social media stocks tumble as Wall Street fears regulation	Shares of Facebook, Twitter and Snapchat-owner Snap fell further on Tuesday as Wall Street fretted over potential regulatory scrutiny that could hobble the business of the social networks.
(b) Dropbox raises IPO price range by \$2 on strong demand	Cloud storage firm Dropbox Inc [DBX.O] raised the price range for its initial public offering by \$2 on Wednesday as investors bid strongly for the first big tech IPO this year ahead of final pricing expected on Thursday.
(c) BMW says electric car mass production not viable until 2020	BMW will not mass produce electric cars until 2020 because its current technology is not profitable enough to scale up for volume production, the chief executive said on Thursday.
(d) Exclusive: On Amazon, a quarter of merchants' sales are cross-border	Amazon.com Inc is making a push for merchants on its website to sell goods into other countries, setting the stage for greater competition with rival marketplaces run by eBay Inc and potentially Alibaba Group Holding Ltd.
(e) Toys 'R' Us says 'making every effort' to pay vendors	Toys 'R' Us said at a bankruptcy court hearing on Tuesday that it was working hard to maximise payments to suppliers and lenders , as it starts to shutter 735 big-box toy stores across the United States.

Source: Authors.

4.3.4. Classification output setting

The original mc-CRTC architecture formulates a binary classification problem for stock trend prediction, where the output is either positive or negative. In reality, investors may be more interested in a particular interval where the next day index value may fall into. Therefore, different output settings can be applied to the classification task.

When the stock market index movement is discretised into three classes using $\pm 1\%$ as the thresholds, the distribution of input data is relatively balanced. The output labelled as 'Negative', 'Neutral', and 'Positive' accounts for 28.5%, 32.5%, and 39%. Another attempt is to formulate a 5-class problem using $\pm 0.3\%$ and $\pm 1\%$ as four thresholds. These five categories are labelled as 'Negative', 'Neutral-Negative', 'Neutral', 'Neutral-Positive', and 'Positive', with a distribution of 12.1%, 16.4%, 32.5%, 24.7%, and 14.3%.

The model performances using different classification modes are assessed based on prediction accuracy, as illustrated in Figure 6. The prediction accuracy decreases with narrowed output intervals. Still, CRL with both headlines and preview texts achieves a superior result over IL. Even in a 5-class trend prediction task, it is able to correctly forecast the stock index value more than half of the observation period.

5. Conclusion and future directions

Financial market forecasting has been one of the most discussed DL applications. The market values are intrinsically unpredictable; nevertheless, various types of information that are publicly accessible can help discover potential contributors to the market movements. This paper proposes a novel DL architecture, namely mc-CRTC, that leverages online financial news articles to forecast the stock market movement for the next trading day. It is a multi-channel news-driven system established primarily based

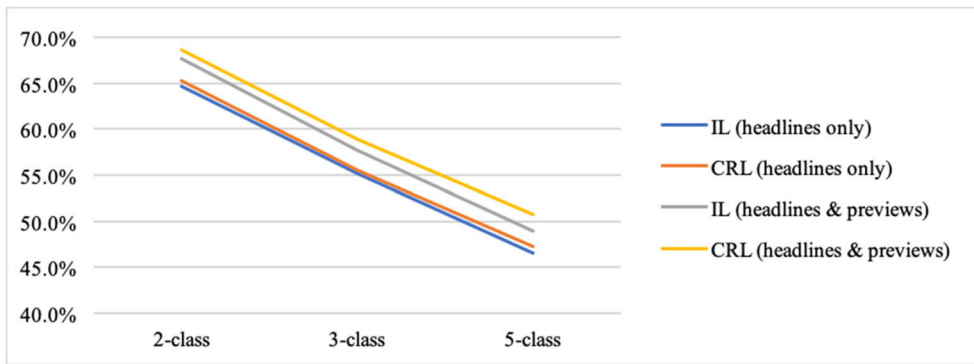


Figure 6. Prediction accuracy over different output classification settings.

Source: Authors.

on TCN (or DCC) and TE. DCC is applied to capture the long-range temporal patterns of stock market index fluctuations within a fixed time window. The index value (e.g., S&P500) is less volatile than single stocks; thus, it is used as the input for the stock data channel. In the other channel, the news data are sent to stacked encoder blocks using the self-attention mechanism for feature extraction. We adopt two separate encoder channels when processing both news headlines and preview texts simultaneously.

The highlight of this paper lies in the method of concatenating these multi-channels. Specifically, cross-connections are introduced between different channels to replace self-residual learning procedures. We empirically demonstrate the feasibility and effectiveness of multi-modal CRL with a blending of input sources. Interpreted by a confusion matrix, our proposed architecture yields better predictive results. Furthermore, the experiment implemented also confirms our assumption that the involvement of news content descriptions, if available, indeed upgrades the model accuracy.

We believe there is still a lot of room for improvement to our current work regarding input data sources and the model design. Below we put forward main future directions.

- More efforts will be devoted to raw data collection, primarily to expand the news dataset. We plan to discover more online news sources, not only from the financial domain, that provide an abstract of contents. Using an increased training dataset size, we expect the proposed architecture to yield more robust outputs. Different input sources such as technical indicators can also be considered to construct other channels.
- The balance between model complexity and predictive performance needs to be evaluated. Neither too sophisticated model design nor excessive input sources can guarantee an improvement. Moderate complexity is acceptable; however, we recommend further discussions on if the increment of accuracy is worth the higher computational cost.
- We propose to change the sliding window size dynamically. Too many zero paddings are not desired when linear projecting the news input tensors. In this study,

we employ the news data collected on the last day of the observation window. We can also reconsider, for example, using news data for the last three days to capture the lagging effect.

- Some modifications to the model design may be achievable. For example, some CRL units can be replaced by cross-attentions, i.e., feeding the output of DCC blocks into TE as the ‘Key’ for MHA.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Abu-Mostafa, Y. S., & Atiya, A. F. (1996). Introduction to financial forecasting. *Applied Intelligence*, 6(3), 205–213. <https://doi.org/10.1007/BF00126626>
- Bai, S., Kolter, J. Z., & Koltun, V. (2018). An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1803.01271>
- Borovykh, A., Bohte, S., & Oosterlee, C. W. (2019). Dilated convolutional neural networks for time series forecasting. *Journal of Computational Finance*, 22(4), 73–101. <https://doi.org/10.21314/JCF.2019.358>
- Cavalli, S., & Amoretti, M. (2021). CNN-based multivariate data analysis for bitcoin trend prediction. *Applied Soft Computing*, 101, 107065. <https://doi.org/10.1016/j.asoc.2020.107065>
- Chen, W., Yeo, C. K., Lau, C. T., & Lee, B. S. (2018). Leveraging social media news to predict stock index movement using RNN-boost. *Data & Knowledge Engineering*, 118, 14–24. <https://doi.org/10.1016/j.datak.2018.08.003>
- Chung, H., & Shin, K. S. (2020). Genetic algorithm-optimized multi-channel convolutional neural network for stock market prediction. *Neural Computing and Applications*, 32(12), 7897–7914. <https://doi.org/10.1007/s00521-019-04236-3>
- Daiya, D., Wu, M. S., & Lin, C. (2020). Stock movement prediction that integrates heterogeneous data sources using dilated causal convolution networks with attention. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 8359–8363). New York, NY: IEEE.
- Ding, X., Zhang, Y., Liu, T., & Duan, J. (2015). Deep learning for event-driven stock prediction. In Q. Yang, & M. Wooldridge (eds.), *IJCAI 2015. Proceedings of the 24th international joint conference on artificial intelligence* (pp. 2327–2333). Freiburg, Germany: IJCAI.
- Dingli, A., & Fournier, K. S. (2017). Financial time series forecasting – A deep learning approach. *International Journal of Machine Learning and Computing*, 7(5), 118–122. <https://doi.org/10.18178/ijmlc.2017.7.5.632>
- Ghosh, P., Neufeld, A., & Sahoo, J. K. (2022). Forecasting directional movements of stock prices for intraday trading using LSTM and random forests. *Finance Research Letters*, 46, 102280
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778). New York, NY: IEEE.
- Hsu, M. W., Lessmann, S., Sung, M. C., Ma, T., & Johnson, J. E. (2016). Bridging the divide in financial market forecasting: Machine learners vs. financial economists. *Expert Systems with Applications*, 61, 215–234. <https://doi.org/10.1016/j.eswa.2016.05.033>
- Hu, Z., Liu, W., Bian, J., Liu, X., & Liu, T. Y. (2018). Listening to chaotic whispers: A deep learning framework for news-oriented stock trend prediction. In *WSDM 2018. Proceedings*

- of the 11th ACM international conference on web search and data mining (pp. 261–269). New York, NY: ACM.
- Jou, B., & Chang, S. F. (2016). Deep cross residual learning for multitask visual recognition. In *MM 2016. Proceedings of the 24th ACM International Conference on Multimedia* (pp. 998–1007). New York, NY: ACM.
- Kim, S., & Kang, M. (2019). Financial series prediction using Attention LSTM. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1902.10877>
- Li, X., Wu, P., & Wang, W. (2020). Incorporating stock prices and news sentiments for stock market prediction: A case of Hong Kong. *Information Processing & Management*, 57(5), 102212. <https://doi.org/10.1016/j.ipm.2020.102212>
- Li, Y., & Pan, Y. (2022). A novel ensemble deep learning model for stock prediction based on stock prices and news. *International Journal of Data Science and Analytics*, 13(2), 139–149.
- Li, Y., Xu, Z., Wang, X., & Wang, X. (2020). A bibliometric analysis on deep learning during 2007 to 2019. *International Journal of Machine Learning and Cybernetics*, 11(12), 2807–2826. <https://doi.org/10.1007/s13042-020-01152-0>
- Liu, J., Lin, H., Liu, X., Xu, B., Ren, Y., Diao, Y., & Yang, L. (2019). Transformer-based capsule network for stock movement prediction. In *Proceedings of the First Workshop on Financial Technology and Natural Language Processing* (pp. 66–73). Macao, China.
- Liu, M., Li, G., Li, J., Zhu, X., & Yao, Y. (2021). Forecasting the price of Bitcoin using deep learning. *Finance Research Letters*, 40, 101755. <https://doi.org/10.1016/j.frl.2020.101755>
- Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, 164, 163–173. <https://doi.org/10.1016/j.knsys.2018.10.034>
- Lyu, J., Bi, X., & Ling, S. H. (2020). Multi-level cross residual network for lung nodule classification. *Sensors*, 20(10), 2837. <https://doi.org/10.3390/s20102837>
- Mills, T. C. (2019). *Applied time series analysis: A practical guide to modeling and forecasting*. Academic Press.
- Mishev, K., Gjorgjevikj, A., Vodenska, I., Chitkushev, L. T., & Trajanov, D. (2020). Evaluation of sentiment analysis in finance: From lexicons to transformers. *IEEE Access*, 8, 131662–131682. <https://doi.org/10.1109/ACCESS.2020.3009626>
- Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2021). A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. *Journal of Big Data*, 8(1), 1–28. <https://doi.org/10.1186/s40537-020-00400-y>
- Oord, A. V. d., Dieleman, S., Zen, H., Simonyan, K., Vinyals, O., Graves, A., Kalchbrenner, N., Senior, A., & Kavukcuoglu, K. (2016). Wavenet: A generative model for raw audio. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1609.03499>
- Radinsky, K., Davidovich, S., & Markovitch, S. (2012, April). Learning causality for news events prediction. In *WWW 2012. Proceedings of the 21st International Conference on World Wide Web* (pp. 909–918). New York, NY: ACM.
- Ren, Y., Liao, F., & Gong, Y. (2020). Impact of news on the trend of stock price change: An analysis based on the deep bidirectional LSTM model. *Procedia Computer Science*, 174, 128–140. <https://doi.org/10.1016/j.procs.2020.06.068>
- Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- Shah, D., Isah, H., & Zulkernine, F. (2019). Stock market analysis: A review and taxonomy of prediction techniques. *International Journal of Financial Studies*, 7(2), 26. <https://doi.org/10.3390/ijfs7020026>
- Sim, H. S., Kim, H. I., & Ahn, J. J. (2019). Is deep learning for image recognition applicable to stock market prediction? *Complexity*, 2019, 1–10. <https://doi.org/10.1155/2019/4324878>
- Tang, G., Müller, M., Rios, A., & Sennrich, R. (2018). Why self-attention? A targeted evaluation of neural machine translation architectures. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1808.08946>

- Vargas, M. R., De Lima, B. S., & Evsukoff, A. G. (2017, June). Deep learning for stock market prediction from financial news articles. In *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)* (pp. 60–65). IEEE. <https://doi.org/10.1109/CIVEMSA.2017.7995302>
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *NIPS 2017. Advances in Neural Information Processing Systems* (pp. 5998–6008). Red Hook, NY: Curran Associates Inc.
- Walczak, S. (2001). An empirical analysis of data requirements for financial forecasting with neural networks. *Journal of Management Information Systems*, 17(4), 203–222.
- Wiese, M., Knobloch, R., Korn, R., & Kretschmer, P. (2020). Quant Gans: Deep generation of financial time series. *Quantitative Finance*, 20(9), 1419–1440. <https://doi.org/10.1080/14697688.2020.1730426>
- Xu, H., Chai, L., Luo, Z., & Li, S. (2022). Stock movement prediction via gated recurrent unit network based on reinforcement learning with incorporated attention mechanisms. *Neurocomputing*, 467, 214–228. <https://doi.org/10.1016/j.neucom.2021.09.072>
- Yu, F., & Koltun, V. (2015). Multi-scale context aggregation by dilated convolutions. *arXiv Preprint*. <https://doi.org/10.48550/arXiv.1511.07122>
- Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, 50, 159–175. [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
- Zhang, X., Qu, S., Huang, J., Fang, B., & Yu, P. (2018). Stock market prediction via multi-source multiple instance learning. *IEEE Access*, 6, 50720–50728. <https://doi.org/10.1109/ACCESS.2018.2869735>
- Zheng, Y. H., Xu, Z. S., & Wang, X. X. (2021). The fusion of deep learning and fuzzy systems: A state-of-the-art survey. *IEEE Transactions on Fuzzy Systems*, 1–1. <https://doi.org/10.1109/TFUZZ.2021.3062899>