

International Journal of Computational Vision and Robotics

ISSN online: 1752-914X - ISSN print: 1752-9131

<https://www.inderscience.com/ijcvr>

A multi-modal image encoding and self-attention-based transformer framework with sentiment analysis for financial time series prediction

Ravi Prakash Varshney, Dilip Kumar Sharma

DOI: [10.1504/IJCVR.2023.10057531](https://doi.org/10.1504/IJCVR.2023.10057531)

Article History:

Received:	28 January 2023
Last revised:	23 March 2023
Accepted:	03 May 2023
Published online:	02 December 2024

A multi-modal image encoding and self-attention-based transformer framework with sentiment analysis for financial time series prediction

Ravi Prakash Varshney*

Department of Computer Engineering and Applications,
GLA University,

Mathura, India

Email: ravi_varshney@outlook.com

*Corresponding author

Dilip Kumar Sharma

Department of Computer Engineering and Applications,
GLA University,

Mathura, India

Email: dilip.sharma@gla.ac.in

Abstract: In this paper, we propose a novel approach for financial time series forecasting using feature selection, image encoding, and a self-attention-based CNN transformer. We use Markov transition field and candlestick chart encoding to extract features from historical stock data. Additionally, we incorporate the sentiment analysis of the financial news data in our model to improve the forecast accuracy. The proposed approach is compared to traditional time series forecasting methods, and the results show that our method outperforms the traditional method in terms of forecasting accuracy. The proposed approach can be used to improve risk management and make more informed trading decisions. Our experiments demonstrate that the proposed framework achieved an improvement of approximately 17.8% in root mean squared error and ~38.7% in mean absolute error for securities lending dataset and ~71.5% improvement in root mean squared error and around ~83.2% improvement in mean absolute error for pricing dataset.

Keywords: candlestick image encoding; computer vision; convolutional neural network transformer; feature selection; Markov transition field image encoding; multivariate time series; MTS; pattern recognition; long-short term memory; LSTM; sentimental analysis.

Reference to this paper should be made as follows: Varshney, R.P. and Sharma, D.K. (2025) 'A multi-modal image encoding and self-attention-based transformer framework with sentiment analysis for financial time series prediction', *Int. J. Computational Vision and Robotics*, Vol. 15, No. 1, pp.31–58.

Biographical notes: Ravi Prakash Varshney received his BE (honours) in Computer Science and Engineering from the Rajiv Gandhi Technical University in 2010 and MTech (honours) in Information Management from the University of Mysore in 2014. He is currently a PhD candidate in the GLA University. He has around a decade of industry experience with different

information technology giants like Infosys Ltd., Deloitte LLP, IHS Markit and S&P Global. He has published various research papers in international journals/conferences of repute indexed in Scopus.

Dilip Kumar Sharma obtained his BE in CSE, MTech in CSE and PhD in Computer Engineering. He is a senior member of IEEE, IEEE-CS, IEEE-SIGHT, IEEE-IAS, ACM, USA, CSI, fellow of IETE and IE in India and member of IET in UK. He has published more than 200 research papers in international journals/conferences of repute indexed in SCI, Scopus and DBLP databases and participated in 50 international/national conferences. Presently he is working as a Dean (International Relations & Academic Collaborations) in Department of Computer Engineering and Applications, GLA University, Mathura, UP, India since March 2003.

1 Introduction

In the modern digital age, huge piles of information are being produced at a never seen scale. Such information is collected by a variety of sensors and other collection mechanisms from different sources regularly over a period. The data in context is usually of high dimension (Yang et al., 2015) and contains numerous correlated features. These datasets are analysed and used in a wide array of applications, like health monitoring, vehicle control devices, volatility analysis in financial markets, data centre monitoring, and avionics, etc. Availability of such indicators in advance can aid a lot in the applications. Forecasting in various domains has become a promising development orientation and aids a lot in better risk management and decision-making analysis. It has always been of interest to researchers to provide more accurate and flexible forecasting methods. An improved time series forecasting method is proposed in the paper (Farsani and Pazouki, 2020) that makes better forecasts for larger periods among its peer methods. As, neural networks are nonlinear and can be used for different types of time series even in absence of specific information about those series. Neural networks in various forecasting models have proved their effectiveness in time series prediction. Transformers using encoders-decoders have been used in a few models for time series forecasting. By decoding the output using the encoded input, a decoder network can produce the desired output. Encoders encode input data based on an encoder network.

Many real-life applications rely on the availability of time series predictions. In a few studies, transformers have been combined with convolutional neural network (CNN) in varying degrees to enhance the efficiency and locality of Transformer. In such combinations, CNN is not fully utilised as they are loosely coupled. The authors (Shen, and Wang, 2022) proposed three tightly coupled convolutional transformer (TCCT) architectures that transformed CNN architectures into transformers to address this issue. An output is generated by estimating the unobserved values using a multilayer network based on the observations in the time series (Brownlee, 2018). Another model based on CNN named Model WaveNet was proposed by the authors (Borovykh et al., 2017) in their work. When forecasting, the network can access a wide range of history by using stacks of dilated convolutions and uses a ReLU activation function. Using multiple convolutional filters to process separate time series in parallel, the proposed network

allows fast data processing and the possibility of exploiting correlation structures between multivariate time series.

One of the domains which have a key dependency on time series predictions is financial stock markets. In the last few decades, financial analysts and researchers have been trying hard to keep up with the problem of predicting stock market movements. When future indicator values can be predicted (Nagula and Alexakis, 2022), it becomes possible to make profitable trades by anticipating those values. Without such a system, it can be difficult to manage risk or make informed trading decisions. Technical analysis of patterns can be used to predict financial indicators and market trends. The repetition of certain patterns in a time series can be used to predict future trends or values based on those patterns and the specific indicator being analysed. A classification task for time series data is mapped to the stock market prediction problem in the paper (Picasso et al., 2019). It incorporates both technical analysis indicators and sentiment indicators from news articles. A robust predictive model based on 20 NASDAQ100 companies resulted in a model that can forecast the trend of the portfolio. It can be observed from the real data collected by social media that users on these channels frequently post short, informal messages. The existing works, especially those that use deep learning approaches, face many difficulties in handling this kind of data. In the paper (Vo et al., 2019), the authors suggested a solution to the issue. The work builds on the earlier research of the authors, in which they suggested fusing domain expertise with the common deep learning method of CNN. The combination is used to enrich training data with more information and create a loss function that makes more sense.

By converting pre-flare geostationary operational environmental satellite (GOES) data from 2010 to 2016, authors (Nagem et al., 2022) have created Gramian angular fields (GAF) images and Markov transition fields (MTFs) images. Next, a deep learning neural network generates predictions from both MTF and GAF images using the solar flare production platform. Using RBMs, human behaviour can be mapped to a high-dimensional space from various viewing angles. The paper (Yang et al., 2019) suggests a paradigm for sensor categorisation using data from multivariate time series sensors as inputs. The method concatenates two-dimensional coloured images of multivariate time series data into a single larger image for classification using a convolutional neural network (ConvNet). The major contributions of this paper are as follows:

- 1 A novel state-of-art framework in which financial data is fed into self-attention-based convolutional transformer and merged with technical indicators, derived from financial news for final security lending forecasting.
- 2 A Multi-Modal computer vision-based time-series data modelling has been proposed. The multi-modal includes either MTF or Candlestick charts image encoding to identify patterns in the data and improves the modelling and overall forecast accuracy. This modelling approach leverages the insights and patterns identified through visual analysis.
- 3 Proposed a new hybrid feature selection method that extracts the most relevant features using grey-wolf optimisation and a novel window length identification algorithm to optimise the window length selected for the time-series data modelling.
- 4 The research methodology, for evaluating the impact of technical indicators from a sentiment of financial news on the security lending data.

The paper is structured as follows: A methodical review of related work is described in Section 2. In Section 3, the proposed framework and various components of the framework are described in detail. Analysis of the datasets, performance metrics is presented, and an analysis of the results is presented in Section 4. A conclusion about the work is explained in Section 5.

2 Related work

Time series forecasting of financial metrics has gained a lot of attention recently. Various strategies have been explored in these studies, including conventional time series analysis techniques, cutting-edge machine learning and deep learning techniques. Such studies have considered the historical stock prices and the recent news stories. In these developed systems, machine learning models have been used for key forecast task. Deep neural networks (DNNs), particularly recurrent neural networks (RNN) and their advanced variants like long-short term memory (LSTM) and gated recurrent units (GRU), have demonstrated strong learning capabilities while modelling a large amount of chronological information in recent articles.

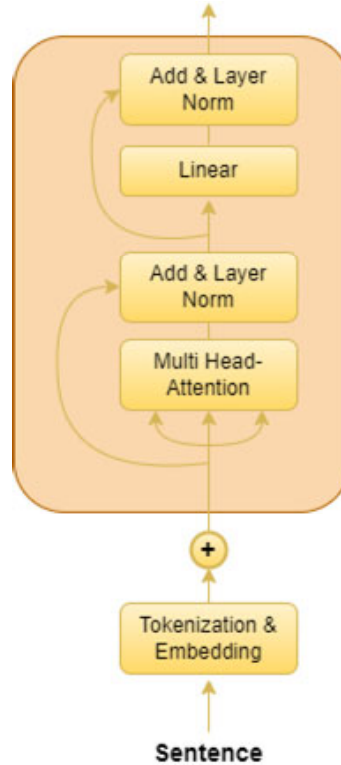
By combining time series augmentation techniques with the original series, authors (Abbasimehr et al., 2022) generated new time series that fit properly with deep learning models using enough samples generated from the augmented series. They used CNN, GRU, and LSTM to forecast COVID-19 time series using the proposed method. Time series forecasting in deep learning is examined in the paper (Agarwal et al., 2022) to give it a structure. Various deep learning-based time series forecasting techniques have been categorised and a report provided based on the underlying structures of these techniques, such as RNNs, CNNs, and transformers. The research (Sang et al., 2023) proposed a revolutionary skin texture preservation-based image enhancing algorithm. The mask was created using the Gaussian fitting, which allows for skin feathering and many instances of box blur. Based on application examples in expert and intelligent systems, the study (Speiser et al., 2019) is an important contribution since it examines variable selection techniques in the context of random forest classification. The paper (Li et al., 2022) presents an approach for improving short-term electric load forecasting by accounting for the dynamic temporal dependency pattern of actual load. The proposed method is based on an enhanced sequence-to-sequence gated recurrent unit network and a three-step adaptive framework. The first step involves estimating maximum duration of temporal dependence using autocorrelation coefficient and value is then used to set an initial upper limit for the window length when constructing time series samples.

2.1 *Transformers in time series prediction*

Recent years have seen that a significant number of studies have used transformers for time-series forecasting. However, original Transformers does not fare well in understanding time-series data's characteristics, so they tend to show promising results in various situations, but they suffer from some fundamental limitations, such as being unable to decompose or interpret the data and not being able to forecast for long periods of time effectively or efficiently. Using the principle of exponential smoothing, authors (Woo et al., 2022) proposed a state-of-art architecture (ETSformer) based on transformer to improvise the time-series forecasting. The authors were inspired from the original

exponential smoothing methods and proposed an enhanced exponential smoothing attention which improved the overall efficiency and accuracy. Figure 1 represents the architecture of a transformer.

Figure 1 Architecture of transformer (see online version for colours)



Spatial-temporal convolutional transformer network (STCTN) has been proposed (Huang et al., 2022) as an innovative model for multivariate time-series prediction. Modelling temporal and spatial dependencies is made possible by STCTN by combining two novel mechanisms for attention. Future price predictions are an area in the financial markets that are much sought after, but also complex and therefore challenging. Equity data of S&P 500 index was analysed as a part of the study (Jonsson and Deumic, 2022), which examined whether Transformers, a novel machine-learning model, could outclass traditional methods and the long-established ARIMA model. The temporal fusion transformer (TFT), a unique attention-based design that blends effective forecasting with comprehensible perceptions into temporal dynamics, is introduced in the study (Lim et al., 2021).

2.2 CNNs with encoder and decoder

A large scale DNN model usually comprises of extremely parameter efficient DNN. In the study (Wakayama and Saito, 2022) a CNN responsible for extracting the noteworthy features was replaced with a extremely parameter efficient DNN architecture in the field

of image identification (SAN). The crucial data that helps in prediction is the high-level characteristics. In our studies, we found that it is challenging to construct a sound event detection model that is more accurate than CNN-transformer in predicting the future. In the paper (Zhang et al., 2022), the authors present a unique deep RUL prediction approach called dual-aspect self-attention based on the transformer (DAST). DAST is an encoder-decoder structure that is only based on self-attention and does not include any RNN or CNN modules. DAST comprises of couple of encoders which work concurrently to identify features from various sensors.

2.3 Sentiment analysis financial data

When analysing and categorising sentiment texts, context is the key. To anticipate the orientation of attitudes, the research (Yusof et al., 2019) suggests a paradigm for sentiment analysis based on a context enrichment model. By taking context and semantic information into account, the system seeks to provide a methodology for sentiment classification. Pre-processing, a context enrichment model, classification, and performance evaluation are the four parts of this approach. The prediction processes (Li et al., 2020) have made use of both stock prices and news items. Despite this, a challenge remains in combining technical indicators from stock prices with sentiments from news articles to create a prediction model that can effectively learn sequential data within time series. In this study, authors have developed a stock metric prediction system and have suggested a method for representing metrics by technical indicators through methodological examination and news articles by sentiment vectors through sentiment assessment.

The rapid development around Natural Language Processing (NLP) has helped the financial industry to analyse a piles of text data. Financial sentiment analysis (FSA) (Xiang et al., 2022) has been the subject of some recent research, but there has not been a comprehensive review. A comprehensive survey of FSA was provided in the paper (Man et al., 2019). The review explained in detail about the sources of the data, conventional ML models and current deep learning advancements.

2.4 MTF for image encoding

MTF is a method for image encoding that is based on the Markov property, which states that the probability of a certain event occurring is dependent only on the state of the system immediately preceding it. In MTF, the image is divided into small overlapping blocks, and the probability of each pixel value in a block is determined based on the values of the pixels in the surrounding blocks. This information is then used to create a compact representation of the image, which can be used for efficient compression and transmission. In the study (Wang et al., 2022) presented a unique method for monitoring the aflatoxin B1 (AFB1) level in maize. A CNN methodology along with deep learning models based on near-infrared (NIR) and MTF was employed in the implementation. New architectures by combining 1D-CNN and 2D MTF-CNN was proposed in sync with the structure of NIR to form the model for the observing of aflatoxin B1 in maize. In their (de Paula et al., 2023) respective work, a strategy was proposed to encode EEG data to pictures as a data preparing step before classification with 2D-kerner-based CNNs. Various techniques such as Gramian angular difference, recurrence plots and MTD were used. Moreover, they also conducted a comparison analysis utilising different CNNs.

2.5 Candlestick charts for image encoding

Candlestick charts have been widely studied for their ability to forecast stock market trends. Various methods have been proposed to utilise candlestick-chart-formed data and pre-defined patterns for stock market prediction. One approach is the tree-based pattern-search method introduced by Karsten Martiny (2012), which aims to discover essential candlestick patterns to predict future price movements. Recent investigations have demonstrated the potential of analysing visual characteristics of candlestick charts to predict stock market trends. For example, studies (Hu et al., 2018) have used convolutional auto-encoders to summarise historical financial data as images, applied K-line theory (Fengqian and Chao, 2020) to characterise candlesticks, and proposed systems (Ananthi and Vijayakumar, 2021) that generate signals on the candlesticks to predict market price movement. Additionally, a novel variation of conventional candlesticks called RGBSticks (Eren, 2020) has been introduced as a means of predicting daily stock prices using an auto-encoder based on DNN.

In each period, candlestick charts show the high, low, opening, and closing prices. Human actions and emotions are patterned and repeatedly replicated, which leads to the formation of candlestick patterns. These patterns record data about the candles. The authors (Chen and Tsai, 2020) provide a two-step method to automatically recognise candlestick patterns to emphasise its capabilities.

3 Proposed work

In this section, the methodology of the proposed work for feature selection, sentiment analysis and the time series forecasting are explained. The strategies used in various modules are explained in this section.

3.1 Feature generation

In order to generate new features, we used the percentage change from each input feature. This was done by taking the difference between each value and the old value of a given input feature and dividing that difference by the old value. The resulting value was then multiplied by 100 to obtain the percentage change. The mathematical representation of percentage change is given in equation (1):

$$\text{percentage change} = \frac{\text{old value} - \text{new value}}{\text{old value}} * 100 \quad (1)$$

By applying this percentage change to the previous value of the target variable, we can generate new features that are useful for predicting future values of the target variable. These new features can provide additional information about how the target variable is changing over time, which can be helpful in making more accurate predictions.

3.2 Feature selection

A feature selection technique is applied to the feature set to determine the appropriate features for further analysis. This is achieved by utilising the grey wolf optimisation

technique to identify the subset of features that exhibit the highest mutual information with respect to the target variable. The features that possess the most optimal correlation with the target feature are then retained for subsequent steps, while all other features are eliminated from the feature set. Mutual information, a measure of association between two variables, quantifies the decline in vagueness of one variable provided knowledge of the other variable. Two random variables X and Y have the same mutual information if they are considered together. The equation (2) of mutual information is as follows:

$$I(X : Y) = H(X) - H(X | Y) \quad (2)$$

where

$I(X; Y)$ is the mutual information for X and Y

$H(X)$ is the entropy for X

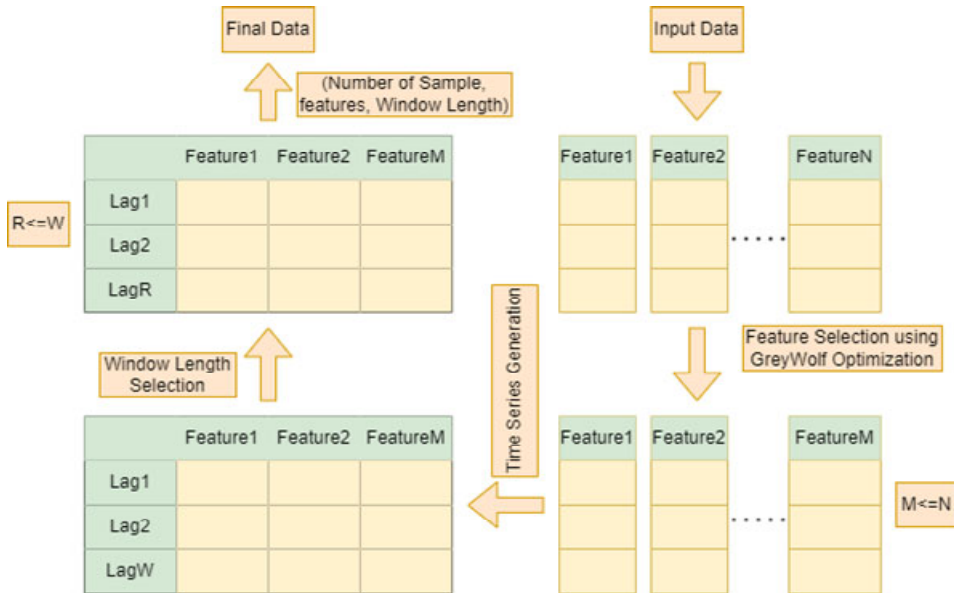
$H(X | Y)$ is the conditional entropy for X given Y .

The feature selection process in this approach employs a hybrid methodology, utilising both grey wolf optimisation and mutual information techniques to recognise the relevant features from the initial feature set. Once the relevant features have been identified, time series data is generated using these selected features, with the optimal window length being dependent on the specific characteristics of the data. During the generation of the time series data, it is important to consider the impact of different lag values of the window length on the target forecast variable. To determine all the relevant lag values, which reside within the range of the window length, the data is generated for each possible lag value and the mean mutual information of all the features present in the data is calculated. This allows for the identification of lag value features that have a significant impact on the target variable while filtering out those that do not.

It should be noted that lagging a time series refers to shifting the values forward one or more-time steps. This is a common technique in time series analysis to better understand the underlying patterns and trends in the data. In summary, the feature selection method employed in this approach is a hybrid one, where the initial feature set is optimised by maximising the mutual information between the candidate and target variables using grey wolf optimisation and mutual information techniques. Subsequently, during the generation of the time series data, different lag values of the window length are analysed, and features that do not have an impact on the target variable are dropped. This results in a more robust and accurate model for time series forecasting.

3.2.1 Grey wolf optimisation feature selection technique

In nature, grey wolves follow a hierarchy of leadership and have a hunting mechanism like that of grey wolves. They optimise by mimicking their system of leadership and hunting. An optimisation model for grey wolves focuses on the discovery of optimal sections of the complex search space over the interaction of individuals. Grey wolves have a four-layer hierarchical structure of α , β , δ , and ω . Grey wolves are dominated by the wolf α , which has the best adaptability among all. The main task of β and δ is to help wolf α in handling the wolf pack and making decisions in the hunting process.

Figure 2 Proposed multi-phase hybrid feature selection strategy (see online version for colours)

It can be described as follows: first, wolves α lead grey wolves to hunt, track, and move toward their prey; then β , and δ they command ordinary wolves to attack the prey under their command until it is captured by them. By mimicking grey wolf behaviours such as hunting, surrounding, and attacking, the GWO algorithm simulates predation. As a result, global optimisation can be achieved. Figure 2 represents that the input data with N features are taken then after applying feature selection using grey wolf optimisation the number of features is reduced to M ($M \leq N$). After that time series data is generated in which rows represent the lag values and columns are features. Each lag value i.e., row (total W rows) is represented by batch. Then we do batch filtering based on mutual information that provides the R lag values ($W \leq R$). This is the data that will be provided as input to the auto-encoder.

3.3 Proposed algorithm for window length selection

The proposed model for analysing securities lending data starts by calculating the autocorrelation of the utilisation feature using a specified lag value. This provides a measure of the similarity between different time steps in the time series and allows the model to identify temporal dependencies in the data. To determine the optimal window length for the time series data, a novel algorithm is proposed that considers both the autocorrelation of the target variable and the cosine similarity between the input features at different lag values. Cosine similarity is a measure of similarity between two vectors that are based on the cosine of the angle between the vectors. It is commonly used in NLP and information retrieval to measure the similarity between text documents.

By considering both the autocorrelation and similarity between features, the algorithm aims to identify the optimal window length that best captures the underlying patterns and dependencies in the data. Once the optimal window length is determined, the

time series data is generated and passed through a self-attention-based CNN transformer. The self-Attention mechanism aims to give more weightage to the important features for a specific time step and to reduce the weightage of less important features. This mechanism helps in modelling sequential data and improves the performance of the model. The final output is passed to the feed forward neural network (FNN) to obtain the final prediction of the target variable.

This model makes use of multiple advanced machine learning techniques such as autocorrelation, similarity measure, and self-attention-based transformer to capture temporal dependencies between variables in securities lending data to make accurate predictions. Furthermore, the use of advanced similarity measures such as cosine similarity in the window length selection algorithm provides a powerful method for identifying patterns and dependencies in the data and can improve the performance of the model. The following is the formula as per equation (3) to determine the cosine similarity between two vectors:

$$\text{CosineSimilarity} = \frac{x \cdot y}{\|x\| * \|y\|} \quad (3)$$

where

- $x \cdot y$ = product (dot) of the vectors x and y .
- $\|x\| * \|y\|$ = cross product of the two vectors x and y .

Algorithm 1 Window length selection

Initialisation

$$X = \left[[X_0^0, X_0^1, X_0^2, \dots, X_0^N], [X_1^0, X_1^1, X_1^2, \dots, X_1^N], \dots, [X_M^0, X_M^1, X_M^2, \dots, X_M^N] \right]$$

where

X = input

N = number of features

M = number of samples

t = lag value

y = target variable

$$\text{Correlation, } \alpha_y = \text{corr}(y_i, y_{i+t}) \quad (1)$$

$$\text{corr}(p, q) = \frac{\sum (p_i - \bar{p})(q_i - \bar{q})}{\sqrt{\sum (p_i - \bar{p})^2} \sqrt{\sum (q_i - \bar{q})^2}}$$

$$\text{Cosine Similarity, } \alpha_x = \text{cosine_sim}(X_i, X_{i+t}) \quad (2)$$

$$\text{cosine_sim}(p, q) = \frac{\sum_{i=1}^N p_i q_i}{\sqrt{\sum_{i=1}^N p_i^2} \sqrt{\sum_{i=1}^N q_i^2}}$$

From equations (1) and (2),

$$\alpha = \frac{\alpha_x + \alpha_y}{2}$$

$$score = \frac{1}{1 + e^{-\alpha}}$$

Table 1 represents the score for different window length (lag) values for the two datasets. As the lag value increases the score for that lag also increases and the optimal score is recorded. For the security lending dataset, the lag value 39 gives an optimal score and for stock price dataset the lag value 46 gives an optimal score.

Table 1 Window length with different values and their scores

<i>Security lending data</i>		<i>Stock price data</i>	
<i>Window length value</i>	<i>Window length score</i>	<i>Window length value</i>	<i>Window length score</i>
2	0.28324	13	0.29519
21	0.32152	19	0.30750
28	0.32901	34	0.33209
33	0.33701	43	0.34440
39	0.34249	46	0.34784

3.4 Transformer network

The Transformer model is a type of sequence-to-sequence (seq2seq) model that was initially developed for machine translation tasks. However, it has since been adapted to various NLP activities. The transformer is comprised of a pair of an encoder and a decoder module. The encoder module transforms the input text, which is pre-processed and tokenised, into a sequence of word embeddings, and then passed through a multi-head self-attention mechanism to compute a contextual representation of the input. Unlike RNN and LSTM models, the transformer can accept and analyse multiple embedding values at once, up to the length of a sentence, allowing for the evaluation of the context of the text in both forward and backward sequences, making it more computationally efficient. The output of the encoder module is then passed to the decoder module, which uses another multi-head self-attention mechanism to translate it into the desired output format (e.g., translated text).

The transformer model has some limitations that should be considered when evaluating its effectiveness and potential use cases. These include:

- High computational cost: The transformer model requires a significant amount of computation and memory, which can make it challenging to train and deploy on low-resource devices or systems with limited computational power.
- Difficulty handling rare or unseen words: The transformer model relies on a pre-trained vocabulary of words and may struggle to handle words that are not included in the vocabulary.
- Limitations in processing sequential dependencies: Although the transformer model is effective in modelling long-range dependencies, it may struggle with certain types of sequential dependencies that involve nonlinear patterns or multiple layers of abstraction.

Researchers have proposed various modifications and extensions to the transformer model to overcome these limitations. For example, pruning techniques can be used to reduce the model size and complexity, external knowledge sources can be incorporated to handle rare or unseen words, and more advanced architectures can be designed to capture complex sequential dependencies. Taking these limitations and potential solutions into account is important when assessing the suitability of the transformer model for a given task.

Figure 3 Self attention-based CNN transformer (see online version for colours)

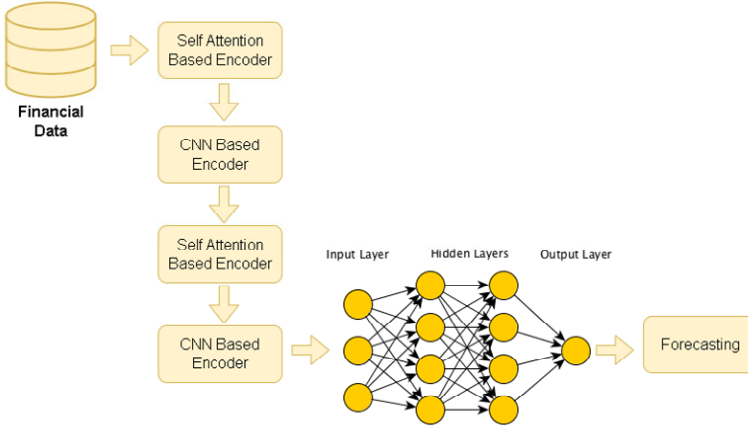


Figure 3 show the architecture of self-attention-based CNN Transformer that is used in the proposed model. Recent research in the field of transformers has focused on their strong potential for learning, as they have been shown to offer a paradigm with sound scalability. In comparison to CNN, transformers have been found to perform better in scenarios where pre-training data is available and have also been shown to fix the limitations of the transformer’s lack of inductive bias, while also improving the transfer effect in downstream tasks. Researchers (Dosovitskiy et al., 2020) in their work established the fact the use of transformers offers a paradigm with sound scalability. Transformers fares better in performance as compared to CNN in the scenarios’ where pre-training data is available. A recent model built by authors (He et al., 2022) is an example of this, as it reconstructs the masked pixels after randomly masking a portion of the input image. This model is known as a masked language model, which is trained to predict the missing tokens in a sentence and has been proven to be effective for pre-training in NLP tasks.

Financial data is transmitted to a CNN-based encoder layer after passing via a self-attention-based encoder. The output of this CNN encoder layer was once more sent to CNN-based encoder and self-attention-based encoder. After flattening the output, an artificial neural network (ANN) receives it. The final predicting value of the self-attention-based CNN transformer is the ANN’s output.

3.5 Self-attention based CNN transformers

The self-attention mechanism, compression of attention toward oneself, is a key component of the transformer model. It allows for parallel computation and eliminates

the need for searching deep networks for lengthy sentences, as compared to RNN and CNN architectures. The self-attention layer performs simultaneous examination of all words in the same sentence, making it a computationally efficient matrix calculation that can be performed in parallel across multiple computer units. The multi-head architecture allows the self-attention layer to extend its scope, or the distance between related words in a sentence, by analysing different subspaces of words simultaneously. Each head in the multi-head architecture can attend to different positions of the input and generate different representations, which are concatenated and passed through a fully connected layer to produce the final representation.

Figure 4 CNN and attention-based encoder (see online version for colours)

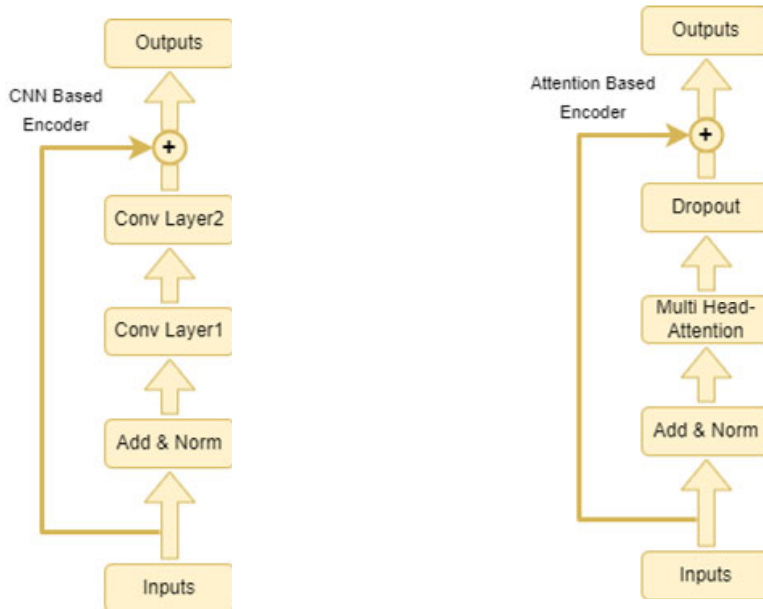


Figure 4 displays the structure of CNN based encoder and attention based encoder used in the proposed CNN transformer model. The proposed CNN Transformer model utilises an encoder, built upon a multi-head convolutional self-attention layer, to encode the sequential dependencies between the input data, and a decoder to decode the long-term dependencies between the target data and the input data. This model differs from traditional CNNs, in that it performs convolution in different subspaces of words simultaneously, allowing for capturing more semantic or syntactic features from n-grams and better handling of complex NLP tasks such as sentiment analysis. Furthermore, CNN Transformer uses various convolution window sizes in different convolutional heads, allowing for effective capture of n-grams with different lengths, while each head focuses on specific n-grams. Additionally, the model discards the pooling operation typically used in conventional CNNs, preserving the sequential information of texts and making it a sequence-to-sequence model. CNN Transformer utilises convolution that focuses more on the local context rather than the global context. This allows the model to capture local n-gram features more effectively, which is crucial for sentiment analysis.

Additionally, CNN Transformer has an abridged structure, making it more memory and computationally efficient, especially for long texts. The convolutional operation in this model allows for capturing local dependencies in the input text, which are important for tasks like sentiment analysis where the meaning of a word can change based on its context. The multi-head convolutional self-attention layer in the encoder allows the model to capture a diverse set of features and the decoder, with its multi-head self-attention layer, generates the output based on the encoded representation and the context of the input.

3.6 Markov transition field

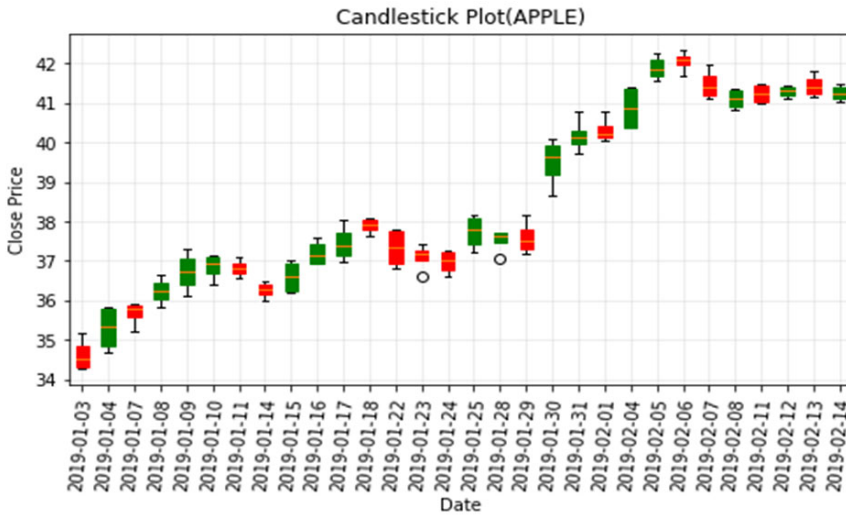
MTF is a type of probabilistic graphical model that is used for modelling temporal dependencies between variables. An MTF is a generalisation of a Markov random field (MRF) to handle time-varying variables. In an MTF, each variable can depend on the values of the other variables at the current time as well as at previous times, and the probability distribution of the variables is defined over a sequence of time steps. This allows an MTF to capture temporal dependencies between variables, which can be useful for modelling time-series data. The features generated are also encoded into an image using MTF and CNN is applied on the images encoded. Then the output of the self-attention-based transformer and CNN-based forecasting of the encoded image is blended with the financial news dataset. The equation (4) shows the mathematical representation of MTF:

$$M = \begin{bmatrix} m11 & m12 & m13 & m14 & m15 & \dots & m1l \\ m21 & m22 & m23 & m24 & m25 & \dots & m2l \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ ml1 & ml2 & ml3 & ml4 & ml5 & \dots & mll \end{bmatrix} \quad (4)$$

MTF is an architecture that utilises a sequence of operations for each user, with a spatial dimension of $l \times l$. In a similar fashion to a convolutional network used for image processing, a batch of n records is fed into the network, resulting in a matrix of dimensions $n \times l \times l$ being inputted into the network.

3.7 Candlestick charts for image encoding

Candlestick charts are a type of financial chart that is commonly used in technical analysis to represent the price movements of securities, derivatives, or currencies. These charts are popular in stock market, forex market, and other financial markets. Candlestick charts provide a detailed visual representation of the price action of an asset over a given period. This period can vary from a single minute to a month, depending on the context. The candlestick chart is constructed using four key data points: the open price, the high price, the low price, and the close price. These prices are plotted as a series of candlesticks, with each candlestick representing a specific period, such as one day or one week. Figure 5 represents the candlestick plot for Apple Company. The body of the candlestick represents the range between the open and close prices and is typically colour either black or white. The wicks or lines above and below the body represent the high and low prices for the period.

Figure 5 Candlestick chart for APPLE company (see online version for colours)

A bearish candlestick, characterised by a filled body, signifies that the security closed at a lower price than its opening price, thereby indicating a downward trend and high selling momentum. Conversely, a bullish candlestick, characterised by a hollow body, signifies that the security closed at a higher price than its opening price, indicating an upward trend and high buying momentum. In algorithmic trading, machine learning models are trained on historical data of candlestick charts. Once the model is trained, it can predict future price movements based on the patterns and trends it has learned from historical data. These predictions can be utilised to make trading decisions like buying or selling an asset.

3.8 Sentimental technical indicators

Sentiment technical indicators are tools that track and measure the emotions and attitudes of traders and investors towards a stock or market using sentiment analysis. They help to analyse market trends and make informed investment decisions. While they provide valuable insights into market sentiment, they should be used with other fundamental and technical analysis techniques for well-informed investment decisions.

Sentiment analysis is a method of understanding and extracting emotions, opinions, and attitudes from written or spoken language. It is used to identify and extract subjective information from a given text. It is used to classify the overall sentiment of a piece of text as positive, negative, or neutral. The process of sentiment analysis can be performed at different levels such as document, sentence, and aspect levels. The main idea behind it is to identify specific aspects or features of a product or service and determine the sentiment associated with them. Machine learning algorithms are used to analyse text data, identify patterns and features that indicate the sentiment, and then classify new text data based on its sentiment. To perform this task, a labelled dataset is needed which can be obtained from various sources such as social media platforms, news articles, or customer reviews. Sentiment analysis has a wide range of applications in various industries such as

marketing, customer service, and politics. Companies can use sentiment analysis to track their brand reputation on social media and to improve customers.

- *Subjectivity*: The degree to which a person is immersed or involved in an object at a personal level is described as subjectivity. In other words, they are personal ties or experiences with the object in a context that may or may not synchronise with another person's perspective. Subjectivity can be either positive or negative. As the user is sharing personal experience and feelings towards an object, the statement is bound to be subjective as this experience and feeling may differ from one another. Equation (5) shows the range of subjectivity.

$$\text{Subjectivity} = [0,1] \quad (5)$$

- *Polarity*: On the contrary, polarity quantifies how strong or impactful is the opinion. In simple words, it describes the strength of opinion. This could be positive or negative. Positive feelings or emotions such as affection, trust, or admiration have a certain orientation towards all other facets of an object's existence. This exactly holds true for suppressing or negative polarities as well. Range of polarity is shown in equation (6).

$$\text{Polarity} = [-1,1] \quad (6)$$

- *Probability*: Finance news data is semi-annotated manually and followed by applying TextBlob on this data. Subsequently, subjectivity and polarity are calculated for each sample. Then we convert text into embedded text by using universal sentence encoder (USE). The embedded text is passed into a simple ANN and the output we are getting from ANN for each sample is the probability and added to the finance news data. The range of probability is represented in equation (7).

$$\text{Probability} = [0,1] \quad (7)$$

Datasets that are manually annotated are used by most previous research and studies. According to this hypothesis; manual annotations are the most accurate since humans are the experts in them. Unfortunately, many important points have been overlooked during this process. These annotations are subjective since they are created by humans, so there is a degree of subjectivity involved, and annotations may vary from person to person. Secondly, the mental state of the expert while doing annotations also does play a key role. Annotations can also be skewed by human error and biasness. Considering the above-mentioned issues, this study adopts the use of TextBlob (Aljedaani et al., 2022).

3.9 Artificial neural network

Finance news data is semi-annotated manually and followed by training ANN model on this data. Semi-annotation combines manual and automated methods to label the data. In financial news data, a portion of the data is manually labelled while the rest is automatically labelled using machine learning models based on patterns learned from the labelled data. Let's discuss about ANN models.

ANNs are a type of machine learning algorithm inspired by the human brain's structure and function. It is made up of interconnected processing nodes or neurons that work together to make predictions and perform complex computations. An ANN

typically consists of an input layer, one or more hidden layers, and an output layer. Data is received by the input layer, which is then processed through the hidden layers and ultimately provides the final prediction in the output layer. During training, a large dataset with known inputs and outputs is provided to the ANN. The weights and biases of the neurons are then adjusted through a process called backpropagation. This involves calculating the difference between the predicted output and the actual output and using that error to modify the weights and biases of the neurons. Once an ANN has been trained, it can be used to make predictions on new, unseen data. For instance, an ANN can predict stock prices based on historical data or classify images based on their content. ANNs have been used in various applications, including speech recognition, NLP, image classification, and financial prediction. However, ANNs require substantial training data and can be computationally expensive.

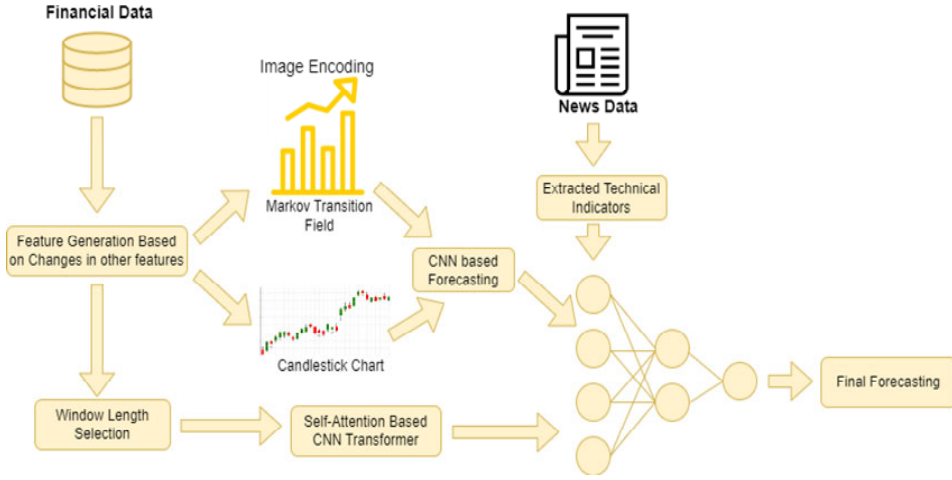
3.10 Universal sentence encoder

The USE is a deep learning model developed by Google for NLP tasks such as text classification, semantic similarity, and text clustering. The model uses a neural network architecture called the Transformer, which relies on self-attention to encode and decode sequences of input tokens. During the encoding process, the USE model tokenises the input sentence using the byte pair encoding (BPE) algorithm, which compresses the input text by replacing frequent pairs of adjacent characters with a single token. The pre-trained transformer layers are then applied to the tokenised input sequence, producing a fixed-length vector representation of the sentence.

This article (Fattoh et al., 2022.) introduces a sentiment analysis model that uses the USE on Twitter data. The model classifies the data into positive, neutral, and negative categories. The study highlights the potential of deep learning models for analysing social media sentiments to provide insights for decision-makers.

3.11 Proposed model

The proposed model for analysing securities lending data starts by analysing and engineering features from the dataset, such as stock loan availability, volume on loan, short interest, and key indicators that define the supply, demand, and borrowing costs of different securities across the globe. These indicators are used to depict the flow of liquidity and funds, which are considered key market indicators. The target variable in the study is the utilisation of security, which determines the popularity or demand of the security in the securities lending industry. The multi-phase feature selection is applied to retrieve the optimal features from the dataset. The feature subset is selected by using the feature selection technique, allowing the relevant features to be analysed further. Selecting the features involves using the grey-wolf optimisation technique to find the feature subset that has the maximum mutual information about the target variable. This is followed by selecting the features whose dimensions are closest to those of the target feature and dropping all the other features. To identify temporal dependencies in the data, the model calculates the autocorrelation of the target variable using a lag value.

Figure 6 Architecture of the proposed framework for time series forecasting (see online version for colours)

A novel algorithm is proposed for the selection of window length, which considers the autocorrelation of the target variable as well as the similarity between the input features at different lag values, to find the optimal value for the window length. Once the optimal window length is determined, the time series data is generated and passed through a self-attention-based CNN transformer. If dataset have the high, low, opening, and closing prices then the features generated are encoded into image form using Candlestick charts.

Algorithm 2 Proposed framework

Initialisation

$$L = \{L_1, L_2, L_3, \dots, L_m\}$$

where L is the set of m lag values selected by the random search for finding the optimal value of lag.

The total number of data samples, is represented by the variable n .

$$\text{lag_score} \leftarrow []$$

for $l \leftarrow 1$ to m :

$$\alpha_x, \alpha_y \leftarrow [], []$$

for $i \leftarrow 1$ to n :

$$\alpha_x^i \leftarrow \text{cosine_sim}(X_i, X_{i+l})$$

$$\alpha_y^i \leftarrow \text{corr}(y_i, y_{i+l})$$

$$\alpha \leftarrow \text{mean}\left(\frac{\sum_{i=1}^n \alpha_x^i}{n}, \frac{\sum_{i=1}^n \alpha_y^i}{n}\right)$$

$$\text{lag_score}_l \leftarrow \frac{1}{1 + e^{-\alpha}}$$

#Selecting lag with optimal score

$$\text{optimal_index} \leftarrow 0$$

for $l \leftarrow 1$ to m :

 if $lag_score_l > lag_score_{optimal_index}$:

$optimal_index \leftarrow l$

Optimal lag: l

Additionally, the features generated are encoded into an image representation using a MTF model, and CNNs are applied to this encoded image representation. The output of the self-attention-based transformer and CNN-based forecasting of the encoded image are then blended with a financial news dataset. Finally, a feed FNN is applied to decide the final prediction of the target variable. The proposed model makes use of multiple advanced machine learning techniques such as MTF and CNN, for capturing temporal dependencies between variables and makes accurate predictions. The blending with financial news data provides additional context and information that can be beneficial in the forecasting target variable. Furthermore, the use of a self-attention-based transformer in the model improves its ability to capture the dependencies between different elements of the input sequence, which is crucial in time series analysis.

4 Results

4.1 Performance measures

The model's errors are measured to assess how well the forecasts performed. At the same scale as the data, the errors are measured. Root mean square error (RMSE) and mean absolute error (MAE) are the two distinct error measurements we used in this study.

- RMSE: RMSE aids in determining the average error magnitude. The square root of the difference is used to get the average of the squared discrepancies between the forecasted and observed values. It is illustrated by the equation (8) below.

$$Root\ Mean\ square\ error(RMSE) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \bar{y}_t)^2} \quad (8)$$

- MAE: MAE aids in determining the average error magnitude, regardless of its direction. It is computed as an average over samples using the absolute disparities between the predicted and observed values. It is illustrated by the equation (9) below.

$$Mean\ absolut\ error(MAE) = \frac{1}{T} \sum_{t=1}^T |y_t - \bar{y}_t| \quad (9)$$

4.2 Datasets

4.2.1 Security lending data

The information present in the dataset can be used to create asset allocation models and risk management procedures with historical coverage from 2002. This dataset can also be used for screening potential investment candidates, assessing market sentiment, determining the cost of borrowing security, calculating the likelihood of short squeezes,

and measuring lender liquidity of both equities and fixed income instruments. This dataset contains:

- Data on short interest rates in fixed income and stock markets are available globally.
- Prime brokers, hedge funds, and loan agencies provide daily and intraday updates with the opportunity to compare the information to the stock exchange. Aggregated short-interest data is delivered every two months with a 10-day delay.
- Timely data on offshore transactions and information from unregistered businesses.

In security finance data we have a total of six attributes. Data have 403,309 training sample sizes and 134,437 number of testing sample sizes. So, we have a total of 537,746 samples in the dataset.

4.2.2 Stock price data

Stock price data is a crucial aspect of financial research and analysis. One common source of stock price data is through Yahoo Finance, which provides historical and real-time financial data for various publicly traded companies. This data can include information such as the opening and closing prices, high and low prices, trading volume, and other financial metrics. By using stock price data collected through Yahoo Finance, researchers can analyse market trends, predict future stock prices, and make informed investment decisions. Additionally, this data can be used to train machine learning models for various financial applications such as algorithmic trading and portfolio management. In stock price dataset collected through yahoo finance we have 70,596 total samples and six attributes.

Dataset link: <https://finance.yahoo.com/>.

4.2.3 Finance news data

This dataset, which has been web scraped for individual stocks, is created by scraping the entire Benzinga news database. It is stock market news data and contains information for over 6,000 different stocks. The data has been collected from the year 2019 to 2021, providing a comprehensive overview of the stock market trends and developments over a period of two years. The dataset is a valuable resource for anyone interested in understanding the stock market, as it includes a wide range of information on different stocks and companies, including financial news, market analysis, and expert commentary. The dataset can be used for a variety of research and analysis purposes, such as stock market prediction, trend analysis, and portfolio optimisation. Additionally, the dataset can be used to train machine learning models for stock market prediction and trading.

Dataset link: <https://www.benzinga.com/>.

4.2.4 Results

This section discusses and explains the results of the study that assesses the model proposed. During the evaluation and benchmarking of the proposed model, we evaluate the performance of the proposed model and the comparison of it is done against the performance and results of the other models discussed in Table 2. To compare the

proposed model with the benchmark sets, we have used the finance news dataset. The presented results on the dataset are on the testing dataset.

Table 2 Comparing model results for securities lending dataset

<i>Model</i>	<i>RMSE</i>		<i>MAE</i>	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
LSTM	6.112	6.083	2.764	2.767
Transformer (lag = 2)	5.942	5.933	2.609	2.635
Transformer (lag = 39)	5.818	5.836	2.251	2.227
<i>Proposed final framework</i>	<i>4.774</i>	<i>4.995</i>	<i>1.615</i>	<i>1.694</i>

Table 2 represents the results and analysis of the proposed model in comparison to different models for security lending dataset. From the evaluation metrics, it can be easily deduced that the proposed framework is performing better than all other models i.e., LSTM, transformer (lag = 2) and transformer (lag = 39). This proves that the proposed strategy surely improves the forecasting. The proposed framework is better than all other models with regard to error metrics MAE and RMSE. Comparing the proposed framework where self-attention-based CNN transformer is used is giving better results. The proposed framework with CNN transformer attains the best overall RMSE and MAE across all the other models used. For test data proposed framework brings around $\sim 17.8\%$ improvement in RMSE and around $\sim 38.7\%$ improvement in MAE in comparison to LSTM model, around $\sim 15.8\%$ improvement in RMSE and around $\sim 35.7\%$ improvement in MAE when compared to transformer (lag = 2), and around $\sim 14.4\%$ improvement in RMSE and around $\sim 23.9\%$ improvement in MAE when compared to transformer (lag = 39).

Table 3 Comparing model results for stock price dataset

<i>Model</i>	<i>RMSE</i>		<i>MAE</i>	
	<i>Train</i>	<i>Test</i>	<i>Train</i>	<i>Test</i>
LSTM	55.567	40.898	9.188	8.797
Transformer (lag = 13)	23.323	17.002	4.433	4.110
Transformer (lag = 46)	21.421	16.086	3.484	3.150
<i>Proposed final framework</i>	<i>8.178</i>	<i>11.647</i>	<i>1.362</i>	<i>1.474</i>

Table 3 represents the results and analysis of the proposed model in comparison to different models for security price dataset. From the evaluation metrics, it can be easily deduced that the proposed framework is performing better than all other models i.e., LSTM, transformer (lag = 13) and transformer (lag = 46). The proposed framework with CNN transformer attains the best overall RMSE and MAE across all the other models used. For test data proposed framework brings around $\sim 71.5\%$ improvement in RMSE and around $\sim 83.2\%$ improvement in MAE in comparison to LSTM model, around $\sim 31.4\%$ improvement in RMSE and around $\sim 64.1\%$ improvement in MAE when compared to transformer (lag = 13), and $\sim 27.5\%$ improvement in RMSE and around $\sim 53.2\%$ improvement in MAE when compared to transformer (lag = 46).

Figure 7(a) shows the graphical relationship and variance between the predicted value and the original observed values for transformer model with lag =2 for security lending dataset.

Figure 7(b) shows the graphical relationship and variance between the predicted value and the original observed values for transformer model with lag =39 for security lending dataset.

Figure 7(c) shows the graphical relationship and variance between the predicted value and the original observed values for proposed model for security lending dataset.

Figure 7(d) clearly establish fact that proposed framework is better in performance and precision as compared to LSTM, transformer with lag = 2 and lag = 39 models for security lending dataset.

Figure 8(a) shows the graphical relationship and variance between the predicted value and the original observed values for transformer model with lag = 13 for stock price dataset.

Figure 8(b) shows the graphical relationship and variance between the predicted value and the original observed values for transformer model with lag = 46 for stock price dataset.

Figure 7 (a) Training of transformer model with lag = 2 (b) Training of transformer model with lag = 39 (c) Training of proposed framework for security lending dataset (d) Plot comparing all models for security lending data (see online version for colours)

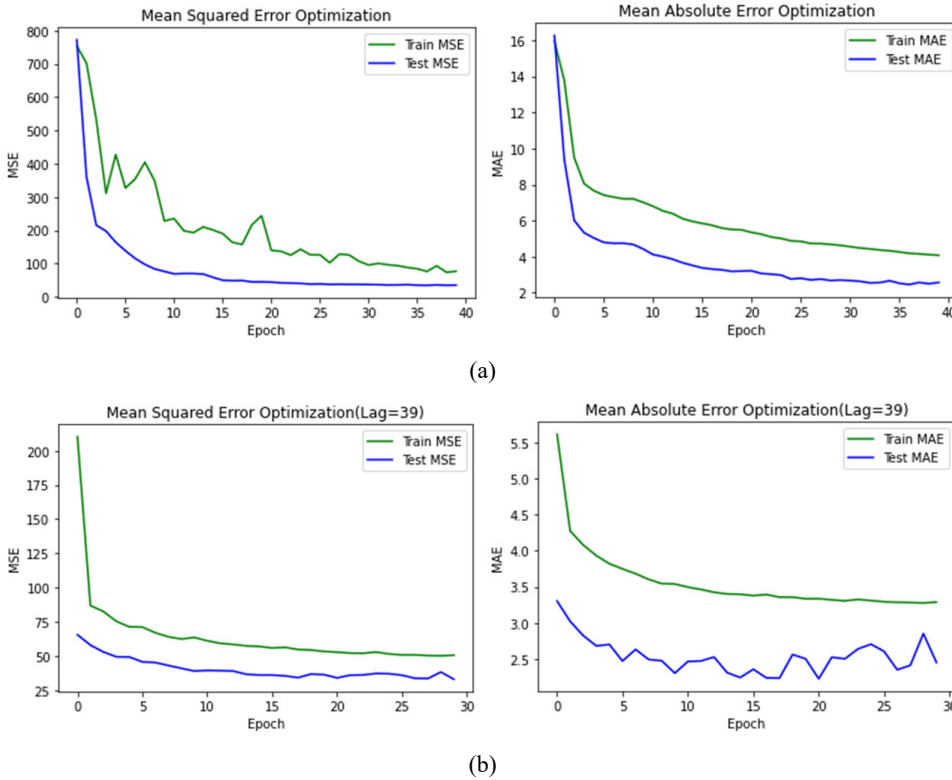


Figure 7 (a) Training of transformer model with lag = 2 (b) Training of transformer model with lag = 39 (c) Training of proposed framework for security lending dataset (d) Plot comparing all models for security lending data (continued) (see online version for colours)

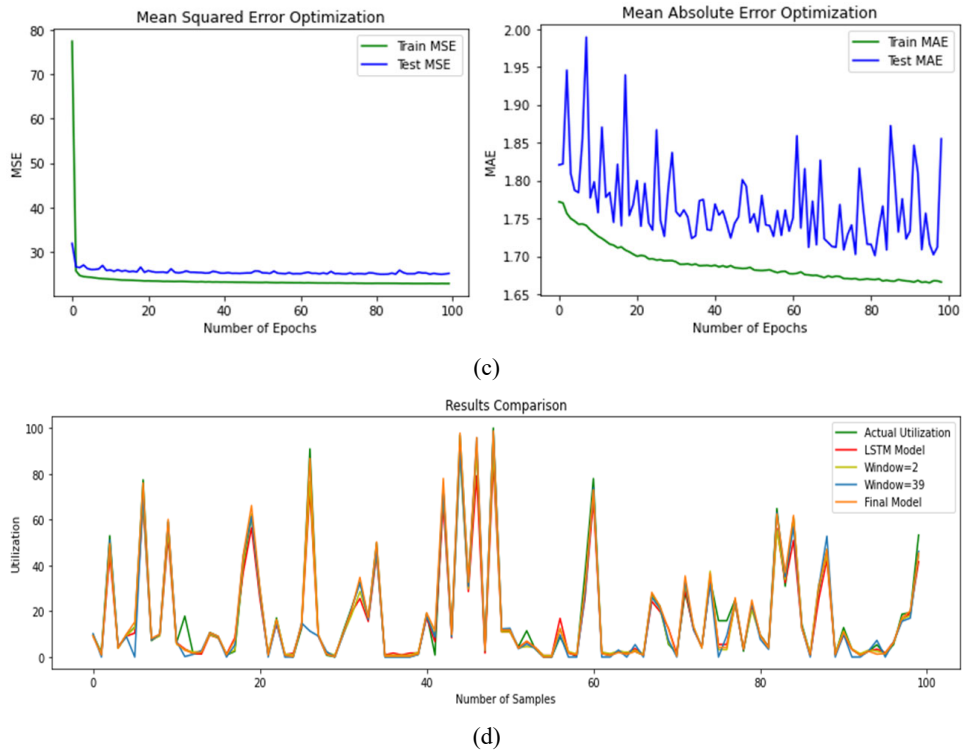


Figure 8(c) shows the graphical relationship and variance between the predicted value and the original observed values for proposed model for stock price dataset.

Figure 8(d) results for proposed framework is better in performance as compared to LSTM model, transformer with lag = 13 and lag = 46 models for stock price dataset.

Figure 8 (a) Training of transformer model with lag = 13 (b) Training of transformer model with lag = 46 (c) Training of proposed framework for stock price dataset (d) Plot comparing all models (see online version for colours)

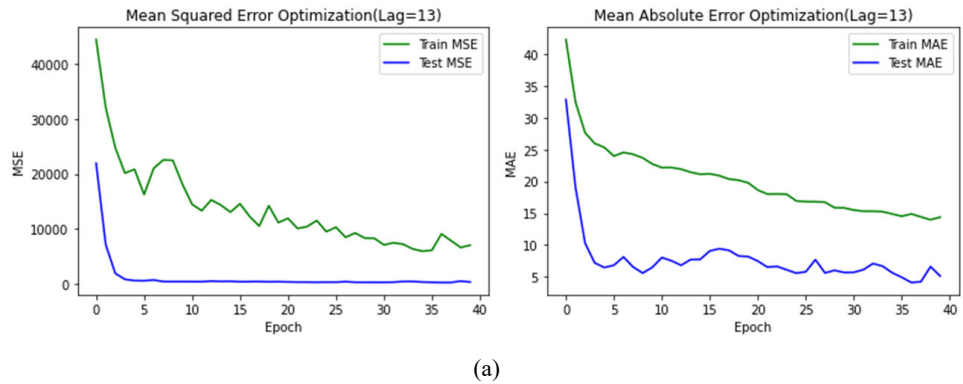
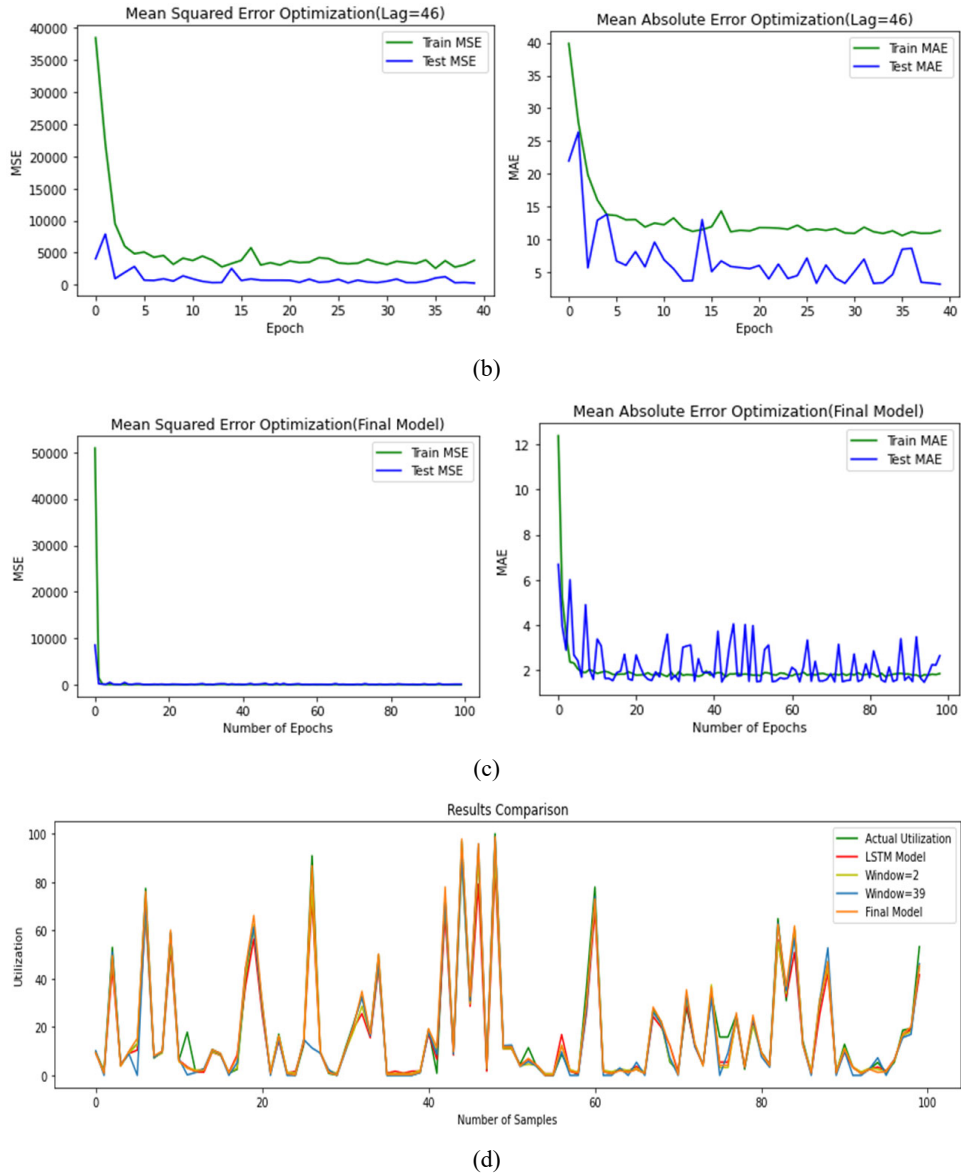


Figure 8 (a) Training of transformer model with lag = 13 (b) Training of transformer model with lag = 46 (c) Training of proposed framework for stock price dataset (d) Plot comparing all models (continued) (see online version for colours)



5 Conclusions

This research paper presents a novel approach for the time series prediction of financial data using the ConvTransformer model. To solve the Curse of Dimensionality problem in multivariate time series data, a hybrid feature selection technique is proposed. This technique extracts relevant features using grey-wolf optimisation and then finds the

relevant lag values in window lengths based on the mutual information technique to solve the Curse of Dimensionality issue. The proposed method combines CNNs with self-attention mechanisms, image encoding with MTF, and dynamic window length selection to capture complex patterns in the data. The model has an encoder that encodes the sequential dependencies in the input data, and a decoder that decodes the long-term dependencies between the target data and the input data. The performance of the model was evaluated on securities lending historical data and technical indicator data, as well as incorporating sentiment analysis of financial news for financial forecasting. The proposed model with transformer improves the RMSE by $\sim 17.8\%$ and MAE by $\sim 38.7\%$ in comparison to standard LSTM implementation. The novel window length optimisation algorithm brings an improvement of $\sim 15.8\%$ in RMSE and $\sim 35.7\%$ in MAE. The final model integrated with the news provides improvement of $\sim 14.4\%$ in MAE and $\sim 23.9\%$ in RMSE over the last improvisation for security lending dataset. Owing to the complex nature of the proposed model, it takes computation time, storage, memory, and resources in the training and testing phases. There are potential avenues for future work on this research paper. Other financial indicators can be incorporated into the model to improve the accuracy of predictions. The proposed model can be combined with other models such as reinforcement learning to improve the performance and generalisation of the model. The proposed model can be applied to other financial markets such as commodity and foreign exchange markets to evaluate its generalisability. Overall, the proposed model provides a promising approach for financial time series prediction, and there are many opportunities for further research to improve its performance and applicability in the financial industry.

6 Limitations and future work

One of the limitations is the high computational requirements of the ConvTransformer model, which may limit its scalability and applicability in real-time prediction scenarios. Another limitation is the assumption of feature independence, which may not hold for all types of financial data. Moreover, the proposed feature selection technique may not be suitable or effective for all financial data or prediction tasks. To address these limitations and improve the proposed approach, there are several opportunities for future research. For instance, exploring alternative techniques for reducing the computational requirements of the ConvTransformer model, such as pruning or compression methods, could be beneficial. Incorporating real-time data and news sources could also enhance the model's accuracy and generalisability. Additionally, integrating other types of financial data, such as macroeconomic indicators or social media sentiment, could improve the model's ability to capture complex relationships in the data. Finally, investigating the use of reinforcement learning or other model combination techniques may further enhance the performance and robustness of the proposed approach.

References

- Abbasimehr, H., Paki, R. and Bahrini, A. (2022) 'A novel approach based on combining deep learning models with statistical methods for COVID-19 time series forecasting', *Neural Computing and Applications*, Vol. 34, No. 4, pp.3135–3149.
- Agarwal, K., Dheekollu, L., Dhama, G., Arora, A., Asthana, S. and Bhowmik, T. (2022) 'Deep learning-based time series forecasting', in *Deep Learning Applications*, Vol. 3, pp.151–169, Springer, Singapore.
- Aljedaani, W., Rustam, F., Mkaouer, M.W., Ghallab, A., Rupapara, V., Washington, P.B., Lee, E. and Ashraf, I. (2022) 'Sentiment analysis on Twitter data integrating TextBlob and deep learning models: the case of US airline industry', *Knowledge-Based Systems*, 14 November, Vol. 255, p.109780.
- Ananthi, M. and Vijayakumar, K. (2021) 'Stock market analysis using candlestick regression and market trend prediction (CKRM)', *Journal of Ambient Intelligence and Humanized Computing*, Vol. 12, No. 5, pp.4819–4826.
- Borovykh, A., Bohte, S. and Oosterlee, C.W. (2017) *Conditional Time Series Forecasting with Convolutional Neural Networks*, arXiv preprint arXiv:1703.04691.
- Brownlee, J. (2018) *Deep Learning for Time Series Forecasting: Predict the Future with MLPs, CNNs, and LSTMs in Python*, Machine Learning Mastery.
- Chen, J.H. and Tsai, Y.C. (2020) 'Encoding candlesticks as images for pattern classification using convolutional neural networks', *Financial Innovation*, Vol. 6, No. 1, pp.1–19.
- de Paula, P.O., da Silva Costa, T.B., de FaissolAttux, R.R. and Fantinato, D.G. (2023) 'Classification of image encoded SSVEP-based EEG signals using convolutional neural networks', *Expert Systems with Applications*, 15 March, Vol. 214, p.119096.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. and Uszkoreit, J. (2020) 'An image is worth 16x16 words: Transformers for image recognition at scale', arXiv preprint arXiv:2010.11929.
- Eren, U.N.L.U. (2020) 'RGBSticks: a new deep learning based framework for stock market analysis and prediction', *Journal of Soft Computing and Artificial Intelligence*, Vol. 1, No. 2, pp.78–85.
- Farsani, M.R. and Pazouki, E. (2020) 'A transformer self-attention model for time series forecasting', *Journal of Electrical and Computer Engineering Innovations (JECEI)*, Vol. 9, No. 1, pp.1–10.
- Fattoh, I.E., Kamal Alsheref, F., Ead, W.M. and Youssef, A.M. (2022) 'Semantic sentiment classification for covid-19 tweets using universal sentence encoder', *Computational Intelligence and Neuroscience*, 5 October, Vol. 2022.
- Fengqian, D. and Chao, L. (2020) 'An adaptive financial trading system using deep reinforcement learning with candlestick decomposing features', *IEEE Access*, 23 March, Vol. 8, pp.63666–63678.
- He, K., Chen, X., Xie, S., Li, Y., Dollár, P. and Girshick, R. (2022) 'Masked autoencoders are scalable vision learners', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp.16000–16009.
- Hu, G., Hu, Y., Yang, K., Yu, Z., Sung, F., Zhang, Z., Xie, F., Liu, J., Robertson, N., Hospedales, T. and Miemie, Q. (2018) 'Deep stock representation learning: from candlestick charts to investment decisions', in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April, IEEE, pp.2706–2710.
- Huang, L., Mao, F., Zhang, K. and Li, Z. (2022) 'Spatial-temporal convolutional transformer network for multivariate time series forecasting', *Sensors*, Vol. 22, No. 3, p.841.

- Jonsson, E. and Deumic, A. (2022) *Machine Learning For Financial Time Series Prediction: A Comparative Study On Stock Market Data*, Polytechnic University of Catalonia, Barcelona.
- Li, D., Sun, G., Miao, S., Gu, Y., Zhang, Y. and He, S. (2022) ‘A short-term electric load forecast method based on improved sequence-to-sequence GRU with adaptive temporal dependence’, *International Journal of Electrical Power & Energy Systems*, May, Vol. 137, p.107627.
- Li, X., Wu, P. and Wang, W. (2020) ‘Incorporating stock prices and news sentiments for stock market prediction: a case of Hong Kong’, *Information Processing & Management*, Vol. 57, No. 5, p.102212.
- Lim, B., Arık, S.Ö., Loeff, N. and Pfister, T. (2021) ‘Temporal fusion transformers for interpretable multi-horizon time series forecasting’, *International Journal of Forecasting*, Vol. 37, No. 4, pp.1748–1764.
- Man, X., Luo, T. and Lin, J. (2019) ‘Financial sentiment analysis (FSA): a survey’, in *2019 IEEE International Conference on Industrial Cyber Physical Systems (ICPS)*, May, IEEE, pp.617–622.
- Martiny, K. (2012) ‘Unsupervised discovery of significant candlestick patterns for forecasting security price movements’, in *KDIR*, October, pp.145–150.
- Nagem, T.A.H., Alshareef, S., Mohamed, A., Gihedan, A. and Albargathe, S. (2022) ‘Comparison between Gramian angular fields (GAF) and Markov transition field (MTF) images data by using them as input to the deep learning neural network solar flare production platform’, in *2022 International Conference on Engineering & MIS (ICEMIS)*, July, IEEE, pp.1–5.
- Nagula, P.K. and Alexakis, C. (2022) ‘A Novel Machine Learning Approach for Predicting the NIFTY50 Index in India’, *International Advances in Economic Research*, 6 December, Vol. 28, pp.155–170.
- Picasso, A., Merello, S., Ma, Y., Oneto, L. and Cambria, E. (2019) ‘Technical analysis and sentiment embeddings for market trend prediction’, *Expert Systems with Applications*, 30 November, Vol. 135, pp.60–70.
- Sang, H., Chen, B., Chen, S. and Yan, L. (2023) ‘Image enhancement based on skin-colour segmentation and smoothness’, *International Journal of Computational Vision and Robotics*, Vol. 13, No. 1, pp.1–20.
- Shen, L. and Wang, Y. (2022) ‘TCCT: tightly-coupled convolutional transformer on time series forecasting’, *Neurocomputing*, 1 April, Vol. 480, pp.131–145.
- Speiser, J.L., Miller, M.E., Tooze, J. and Ip, E. (2019) ‘A comparison of random forest variable selection methods for classification prediction modeling’, *Expert Systems with Applications*, 15 November, Vol. 134, pp.93–101.
- Vo, K., Nguyen, T., Pham, D., Nguyen, M., Truong, M., Mai, T. and Quan, T.T. (2019) ‘Combination of domain knowledge and deep learning for sentiment analysis of short and informal messages on social media’, *International Journal of Computational Vision and Robotics*, Vol. 9, No. 5, pp.458–485.
- Wakayama, K. and Saito, S. (2022) ‘CNN-transformer with self-attention network for sound event detection’, in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May, pp.806–810, IEEE..
- Wang, B., Deng, J. and Jiang, H. (2022) ‘Markov transition field combined with convolutional neural network improved the predictive performance of near-infrared spectroscopy models for determination of Aflatoxin B1 in Maize’, *Foods*, Vol. 11, No. 15, p.2210.
- Woo, G., Liu, C., Sahoo, D., Kumar, A. and Hoi, S. (2022) *ETSTformer: Exponential Smoothing Transformers for Time-series Forecasting*, arXiv preprint arXiv:2202.01381.
- Xiang, C., Zhang, J., Li, F., Fei, H. and Ji, D. (2022) ‘A semantic and syntactic enhanced neural model for financial sentiment analysis’, *Information Processing & Management*, Vol. 59, No. 4, p.102943.
- Yang, C.L., Chen, Z.X. and Yang, C.Y. (2019) ‘Sensor classification using convolutional neural network by encoding multivariate time series as two-dimensional colored images’, *Sensors*, Vol. 20, No. 1, p.168.

- Yang, X.S., Lee, S., Lee, S. and Theera-Umpon, N. (2015) 'Information analysis of high-dimensional data and applications', *Mathematical Problems in Engineering*, 16 November, Vol. 2015, 2 pp.
- Yusof, N.N., Mohamed, A. and Abdul-Rahman, S. (2019) 'Context enrichment model based framework for sentiment analysis', *International Conference on Soft Computing in Data Science*, August, Springer, Singapore, pp.325–335.
- Zhang, Z., Song, W. and Li, Q. (2022) 'Dual-aspect self-attention based on transformer for remaining useful life prediction', *IEEE Transactions on Instrumentation and Measurement*, 17 March, Vol. 71, pp.1–11.