IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

Received 21 December 2023, accepted 5 January 2024, date of publication 15 January 2024, date of current version 2 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3354702

RESEARCH ARTICLE

Cross-Domain Disentanglement: A Novel Approach to Financial Market Prediction

HUMAM M. ABDULSAHIB AND FOAD GHADERI
Human-Computer Interaction Laboratory, Faculty of Electrical and Computer Engineering, Tarbiat Modares University, Tehran 1411713116, Iran
Corresponding author: Foad Ghaderi (fghaderi@modares.ac.ir)

ABSTRACT Profit maximization and risk mitigation require good financial market predictions. Financial markets have a correlated nature, which means that there are some shared patterns between them; therefore, learning about one market might help understand the behavior of others. End-to-end training techniques have proven successful in financial markets, but they have flaws, such as picking up noise and failing to account for the complicated relationships across markets. We present a promising model for predicting financial markets using the correlation between the two markets, which draws inspiration from the recent progress in disentanglement learning. This model learns to disentangle representations of features shared between markets from specific representations, and removes features that cause interference. We utilized a dilated convolutional neural network as an encoder to extract features while using self-attention and cross-attention to capture specifics and shared patterns. Our model uses Dynamic Time Warping (DTW) to minimize the similarity between specific and shared patterns. It also combines DTW's alignment-based similarity with the Mean Square Error (MSE) to determine the optimal balance between alignment and prediction accuracy. We conducted our experiments using datasets that included the closing prices of Apple, Samsung, Bitcoin, Ethereum, Meta platforms, and the X platform. Spearman's rank correlation coefficient was used to evaluate the disentanglement by describing the relationship between the extracted representations. The findings confirm that our model surpasses state-of-the-art approaches in prediction error, financial risk assessment, correlation evolution, and prediction net curves, thereby giving market participants more trust in their decisions.

INDEX TERMS Attention-mechanisms, cross domain, disentangled representation, financial market, time series prediction.

I. INTRODUCTION

Financial markets have a dynamic and complicated nature. Thanks to the ability of deep learning (DL) methods to process large datasets and perform in-depth pattern analyses, DL has attracted a great deal of interest and been put to widespread use in the financial markets [1], [2]. The prediction of financial time series, a key application of DL, involves the use of past data to make informed future decisions. There is a range of viable deep learning approaches, including models (e.g., recurrent [3], convolutional [4], attention [5], and hybrid networks), feature engineering (e.g., technical indicators [6], economic indicators, and sentiment data [7]), regularization techniques (e.g., batch normalization and weight regularization), and transfer learning [8]. The nature of the time series (including complexity and non-linearity [9]) and the scarcity of data [10], among other reasons, may make unsupervised time series prediction a challenge.

There have been many endeavors to identify potential resolutions, but most have focused on only one market [11] or used data from one market to predict results in another works [12]. However, multiple-market prediction may benefit from the cointegration of data from numerous markets to learn from different data distributions and discern patterns and underlying dynamics, which will improve the training dataset and make the model more stable [13]. In prediction techniques that use end-to-end learning, algorithms use raw input data, learn features, train the model, and then make

The associate editor coordinating the review of this manuscript and approving it for publication was Sajid Ali.

predictions based on the trained model. This technique has limitations in prioritizing meaningful representations [14] that contribute to behavior and patterns, which is the main objective of disentanglement learning. Harvesting the most important factors may simplify subsequent modelling by reducing the dimensions. Separating the various factors helps construct optimized portfolios [15] that balance risk and returns more effectively. These diverse merits motivated us to investigate this approach for financial market prediction.

In multiple-market prediction, it is beneficial to have shared representations that capture similarities between markets, along with domain-specific features. To successfully achieve this task, a model must learn the representations separated into three parts: (1) a shared pattern that captures the similarities between the two data markets, and (2) two domain-specific representations exclusive to each market. This objective requires the consideration of two key points. First, these representations must be disentangled, meaning that they should be dissimilar from one another. Second, they should be informative in a manner that accurately captures each element of the pattern in the appropriate part of the representation.

In this paper, we propose Cross-Domain Disentanglement: A Novel Approach to Financial Market Prediction, a prediction model that uses gains from two markets to disentangle financial market movements into shared and specific patterns. In essence, our work focuses on two main questions: (1) the capability of cross-domain disentanglement to improve prediction, which is our main objective, and (2) how we can attain robust disentanglement. Our framework uses a dilated convolutional neural network (CNN) module, since its proficiency enables the network to have a larger receptive field [16] without increasing the number of parameters [17] in order to capture the temporal feature relationships between historical data. Subsequently, self-attention and cross-attention are also used to capture the specific representation and shared representation, that is; (1) multi-head self-attention, taking advantage of its strength in capturing long-term dependencies and focusing on different aspects of temporal patterns [18], is used for specific representation, and (2) multi-head cross-attention, based on cosine similarity, is applied to represent the shared representation.

The Dynamic Time Warping (DTW) technique, which is a non-parametric method [19] based on dynamic programming techniques [20], allows for the calculation of the distance between two time series and the ability to handle complex time series [21]. We adopted this technique to filter out similar patterns between the specific and shared representations. To achieve equilibrium between the aligned values and the accuracy of the prediction at each aligned time point, we used the mean square error along with the DTW loss function. Contrastive learning, inspired by [16], was used to train the model to discern between pairs of data points that are similar and dissimilar. After concatenating the specific and shared representations, a linear regression model only needs to learn the representations once for each dataset. We used six prominent financial market classes, including stocks—representing assets tied to companies [22] like Apple, Samsung, Meta platform, and X platform—and cryptocurrencies, which are digital currencies not associated with any specific company and operate based on blockchain technology [23], such as Bitcoin or Ethereum. Each pair of inputs was the same size, meaning an equal number of data points. In all the experiments, the accuracy metrics (MSE and MAE), correlation evaluation, risk measurement, and prediction net curves showed that our results were consistent with the state-of-the-art. In short, our contributions include the following:

1) We present an innovative methodology for predicting financial market movements using shared and specific disentanglement between two markets that does not require adversarial training or gradient reversal layers.
2) The black-box nature of deep learning is effectively addressed in our approach, as we provide interpretability and elucidate the inner workings of the algorithm. Separation of the underlying components makes explainable artificial intelligence (XAI) possible.
3) We conducted extensive analyses of our approach in correlated settings on all datasets to ensure that the findings we obtained are consistently state-of-the-art.

The paper's structure can be summarised as follows. The next section will outline some of the key literature that relates to our work. We discuss the theoretical foundations of our model in Section III, while in Section IV, we describe the overall structure of our model. Experiment details and settings are provided in Section V, while Section VI closes with our findings.

II. RELATED WORKS

In recent years, academics and market participants [24], including investors, banks, dealers, financial institutions, and traders, have demonstrated a notable increase in interest in financial market predictions. As a result of this increase in attention, deep learning strategies have been used to improve the challenging task of making predictions in financial markets. Previous works that used DL approaches for predicting financial markets can be classified into two main categories:

1) Based on domain: Deep learning models use historical data in order to predict future values. The previous works in this category may be sub-classified based on the source of the data:
    a) Single domain: These models use a single source of data to make future predictions. They realize good results, but they possess notable limitations, such as ignoring the correlated movement

of markets. Single domains includes univariate features [25], multivariate features [26], and text data [27].

   b) Cross domains: Predictions in these models are based on relationships, correlations, or interactions between different financial markets and sources. Cross-domain disentanglement is a significant topic, and several methods have been proposed for dealing with this task at hand, including transfer learning [28], [29], cointegration [30], and sentiments such as [31] and [32].

2) Style of learning: There are different approaches for predicting time series of prices, and in this study, we categorise them into the following based on their training style:

   a) End-to-end methods: These models are trained to perform an entire task, from raw input data to producing the desired output, and often pick up unwanted or noisy data. There are many algorithms that have proven to be outstanding for this objective, such as recurrence [33], [34], convolutional [35], attention mechanism [5] and hybrid models [11].

   b) Disentanglement methods: These methods aim to separate the underlying factors or features in data so that each factor corresponds to a distinct and interpretable dimension. These methods are particularly valuable when dealing with complex data that may have multiple intertwined factors. Learning disentangled representations has yielded impressive results in a set of studies through some common approaches, including Independent Component Analysis (ICA) [36], Principal Component Analysis (PCA) [37], adversarial training [38], information theory [39], similarity attention [40], and causally disentangled generation [41].

Although these previous efforts have generally led to robust results, certain drawbacks have been pointed out. These include issues like the collection of noisy data or inaccurate correlations, as well as limitations relating to model capacity. Some endeavors to address this point by disentangling representation have achieved good findings, such as GLAD [42]. Nevertheless, they ignore the correlation effects between markets and the absence of correlation analysis. Financial market data usually contains temporal correlations, and while end-to-end learning could be used to model these correlations, it does not provide interpretable predictions, as does disentanglement learning. This work explicitly delves into these limitations and strengthens the positive points for better prediction results as well as interpretability, which is crucial for many downstream tasks. This effort is based on the promising progress of cross-domain disentanglement methods and other similar ideas.

III. CROSS-DOMAIN DISENTANGLEMENT AND ITS THEORETICAL INTERPRETATION

Cross-domain disentanglement addresses the process of isolating and extracting the informative representations that constitute the demeanour of a domain, i.e., unraveling the shared patterns that reflect the similarity between domains in addition to the exclusive patterns of each domain. Our work begins with the following cornerstone theories:

1) Better generalization: Generalization is a vital component for a model to be useful, in that it must be able to generalize its performance far beyond its training data and onto new, unexplored data. The cross-domain approach endows the model with the ability to learn well-to-do temporal sequences, which leads to better generalization [43].
2) Multi-domain prediction may reveal insights when data from one domain inform or improve the knowledge of another; this is especially true for financial market data, where many factors have a role, including complex correlations with other markets [44].
3) Market integration [45] describes the extent to which various financial markets, including stock and cryptocurrency markets, are interdependent or interact with one another. It entails looking at how changes in one market might affect other related markets via trends, events, and price changes. Studying market integration through the identification of similarity patterns is essential to comprehending intermarket dynamics, since fluctuations in one market might potentially initiate an effect in other markets [46].
4) Model performance may be enhanced by increasing the amount of the dataset, which can be done via cross-domain, particularly when data are scarce [47].

In addition to the model performance presented above, there are advantages related to market participants, which include the following:

1) Multiple sources offer traders a more comprehensive understanding of complex systems, which leads to deeper insights and a holistic perspective with the ability to rebalance portfolios.
2) Financial markets are dynamic environments [48], which is a challenge since prices can fluctuate rapidly in response to many factors, including their correlations. Cross-domain disentanglement furnishes market attitudes by virtue of shared representations between markets, along with their specifics.

IV. PROPOSED METHOD

In the following section, we will start with the formulation of the problem. We then proceed to describe the proposed method in further depth.

A. PROBLEM FORMULATION

The time series of financial market prices draws attention to the correlated sequence of data points that reflects the dynamics of financial markets. Similar behavior, known as shared representation, and specific movements, known as specific patterns, give rise to these dynamics. In other words, if we assume that X and Y are time series data from two financial markets, they can be represented as follows:

$$X = \{x_{i,1:T_o}\}_{i=1}^{n} \tag{1}$$
$$Y = \{y_{j,1:T_o}\}_{j=1}^{m} \tag{2}$$

where $x_{i,1:T_o} = \{x_{i,1}, x_{i,2}, \ldots, x_{i,t_o}, \ldots, x_{i,T_i}\}$, $y_{j,1:T_o} = \{y_{j,1}, y_{j,2}, \ldots, y_{j,t_o}, \ldots, y_{j,T_i}\}$, T_0 is the number of data points observed, and x_{i,t_o} and y_{j,t_o} represent the observations of financial markets i and j at time steps t_o. There are some common factors between the representations of X and Y, such as:

1) Specific representations, i.e., X-specific (x^{sp}) and Y-specific (y^{sp}), represent the specific representations of the financial markets.
2) X-Yshared (xy^{sh}) captures similar factors between the two financial markets.

The goal should take into account the following requirements for cross-domain disentanglement:

1) Decomposition: The first step is to decompose the data into domain-specific ((x^{sp}) and (y^{sp})) and shared features ((xy^{sh})), where (x^{sp}), (y^{sp}), and (xy^{sh}) are the X-specific, Y-specific, and X-Yshared, respectively. This can be achieved by multi-head self-attention to capture the specific feature of each domain, whereas multi-head cross domain-based cosine can be similarly used to ascertain the share representation.
2) Disentanglement: Once the decomposition is carried out, the next step is to disentangle the domain-specific features from the shared features. This can be achieved by enforcing each representation to be dissimilar to the others through, e.g. DTW alignment to minimize the similarities. To balance the alignment quality and prediction accuracy, the training objective will combine the alignment loss function with the accuracy metric, which in our model will be the Mean Square Error.

Our aspiration is to assist market participants in making more informed decisions by analyzing market behavior and refining financial market predictions.

B. METHODOLOGY

In this paper, we propose an innovative approach for analyzing and predicting the behavior of financial market movements by separating these movements into shared and specific representations. It is called Cross-Domain Disentanglement: A Novel Approach to Financial Market Prediction, and it takes into account what we discussed above. As an alternative to end-to-end learning from observed data or learning from a single source of data, this work aims to capture and learn usable features from observed data from two sources. We plan to learn representations so that, for each time step, we have disentangled representations for shared and specific parts, i.e., $X = [(x^{sp}); (xy^{sh})]$ and $Y = [(y^{sp}); (xy^{sh})]$, where X and Y are two financial markets, $(x^{sp}) \in \mathbb{R}^{h \times d^{sp}}$, $(y^{sp}) \in \mathbb{R}^{h \times d^{sp}}$, and $(xy^{sh}) \in \mathbb{R}^{h \times d^{sh}}$. such that $d = d_{specific} + d_{share}$. Figure1 provides an overview of the proposed approach.

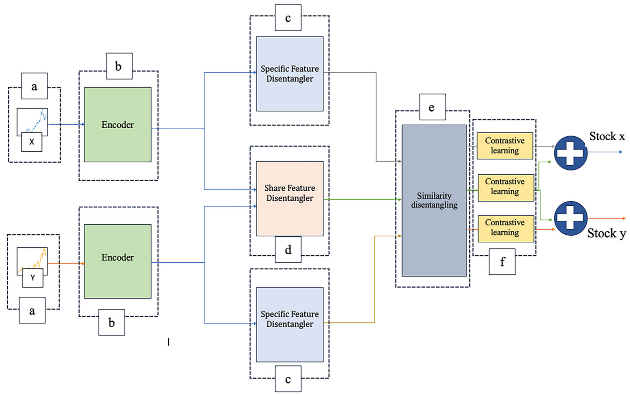

FIGURE 1. Overall Framework. These components make up the model as a whole: (a) Data preprocessing; (b) a dilated CNN to capture the temporal relationship between historical data; (c) multi-head self-attention to capture the specific representation of financial markets; (d) multi-head cross-attention based on cosine similarity to extract the similar patterns between financial markets; (e) the alignment module to minimize similarity between the three representations. This step will combine with MSE to balance alignment and prediction accuracy; (f) contrastive learning as representation discrimination.

1) ENCODER: MAPS INPUT REPRESENTATIONS TO LATENT SPACE

First, we encode raws X and Y into the hidden state representations $f_x : \mathbb{R}^{h\_x \times m\_x} \rightarrow \mathbb{R}^{h\_x \times d\_x}$ and $f_y : \mathbb{R}^{h\_y \times m\_y} \rightarrow \mathbb{R}^{h\_y \times d\_y}$. Our encoder, which draws inspiration from [16], has three parts: an input projection layer, a timestamp masking module, and a dilated CNN module with ten residual blocks. The fully connected layer will map the observations $x_{i:t}$ and $y_{i:t}$at timestamp t to latent vectors $z_{xi:t}$ and $z_{yi:t}$. Subsequently, latent vectors are masked by timestamps to find tokens for data when the raw is challenging owing to the unbounded values. We used a dilated convolutional neural network (CNN) module with ten residual blocks to pull out complex hierarchical features at each timestamp because it can pull out features at different scales, which means that they are stable when the sizes of the inputs change. Two $1D$ convolutional layers with a dilation parameter (2^l for the $l - th$ block) are the structure of the residual block.

2) SPECIFIC FEATURE REPRESENTATIONS: EXTRACTION OF A FINANCIAL MARKET'S UNDERLYING SPECIFIC PATTERNS

This module is designed to capture specific temporal patterns where we take advantage of multi-head self-attention for this purpose. The primary objective of self-attention is to model complex dependencies within sequences, and this is what we need at this point, particularly when capturing diverse and multi-faceted relationships by using multi-head self-attention. The result is obtained by computing a weighted sum of the values for each token, where the weights indicate the similarity between the relevant query token and key tokens. The attention is expressed as:

$$Attention(Q,\ K,\ V)\ = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3}$$

where Q, K, and V are tokens representing the query, key, and value, respectively, and the scaling factor d_k is the dimension of the query and key. Next, we use a ReLU activation in the middle of two fully connected layers to project the attention token outputs.

3) SHARE FEATURE REPRESENTATIONS: FIGURING OUT THE FINANCIAL MARKETS' SIMILARITY PATTERNS

Exploiting the similarity patterns between financial markets is the mission of this module, and in pursuit of this aim, we present a cross-attention mechanism. The cross-attention mechanism is used in time series analysis to capture the interactions between different variables and improve learning performance. This involves utilizing attention modules to discover discriminative cues underlying temporal data and balance feature weights within or across modalities. Cross-attention is similar to self-attention in terms of its structure; however, it processes information differently [40]. One of the paired inputs will be utilized for the query in cross-attention, whereas the other input is used for the key and value (i.e., $Q = input\_x, K = input\_y$). In cross-attention, the result is the weighted sum of the value (which is also the key) input, where the weights reflect the similarity between the query and key inputs. To ascertain the similarity patterns, a common method is cosine similarity, which maximizes the similarity score between financial markets. Finding the cosine of the angle between two vectors allows us to obtain a sense of how close they are, and this is what cosine similarity does. It treats each data point as separate and does not look at the order in which they appear in a time series, which is very important in our case. The timestamps used in the encoder help us overcome this limitation by providing the chronological context. The similarity score was calculated using the following formula:

$$CosSim(x, y)\ =\ \frac{(x\ .y)}{(||x||\ *\ ||y||)} \tag{4}$$

where $(x \cdot y)$ represents the dot product of vectors x and y, and $||.||$ represents the Euclidean norm of a vector.

4) SIMILARITY DISENTANGLING: FILTERING OUT SIMILAR INFORMATION

As discussed above, we should filter out the similar patterns between specific and shared representations for disentanglement. To train our model, we adopt DTW as an alignment-similarity measure with a fixed window to maximize the Euclidean distance between the aligned series (x^{sp}), (y^{sp}), and (xy^{sh}). DTW quantifies the dissimilarity between the three representations, and we employed MSE to weigh how closely the predicted values at each time point matched the actual values. In light of this, to balance alignment quality and prediction accuracy, the training objective is as follows:

$$\begin{aligned}\mathcal{L} = {} & Sim(x^{sp}, y^{sp}) + Sim(x^{sp}, xy^{sh}) + Sim(y^{sp}, xy^{sh}) \\ & + MSE(x\_predicted, x\_input) \\ & + MSE(y\_predicted, y\_input)\end{aligned} \tag{5}$$

where $Sim(.)$ is the DTW function and $MSE(.)$ is the Mean Square Error.

Minimizing DTW and combining it with the MSE loss function will encourage the model to align the time series accurately (according to DTW), while also making accurate predictions (according to MSE) once they are aligned.

5) CONTRASTIVE LEARNING

The key idea behind contrastive learning, as a self-supervised learning technique, is to learn meaningful representations by training a model to distinguish between positive pairs (similar data) and negative pairs (dissimilar data). We apply the MoCo [49], which makes use of two encoders, "query" and "key", with the same architecture, which consists of two fully connected layers in each encoder, and in the middle is a ReLU activation. The goal of these encoders is to find the key representation, the momentum encoder, that is most similar to the query representation, which aims to generate discriminative representations. Positive pairs, which represent samples of data augmentations and consists of scaling, shifting, and jittering techniques as three typical augmentation methods, and negative pairs, which consider the remaining samples in the mini-batch as negative samples, are used to create a contrastive loss function. In our work, we used the InfoNCE loss function, as follows:

$$CL = \sum_{i=1}^{N} -\log \frac{\exp(q_i \cdot \frac{k_i}{\tau})}{\exp(q_i \cdot \frac{k_i}{\tau}) + \sum_{j=1}^{K} \exp(q_i \cdot \frac{k_j}{\tau})} \tag{6}$$

where CL is contrastive learning, τ is a hyper-parameter for the temperature, q is an encoded query, and k is a set of encoded samples. This part $(\exp(q_i \cdot \frac{k_i}{\tau}))$calculates the exponentiation of the dot product between (q) and (k), divided by a temperature parameter. The negative logarithm, denoted as $-\log$, is utilized to transform probabilities into a format suitable for minimizing during training. In simpler terms, this function encourages the model to bring similar representations closer together while pushing dissimilar ones farther apart.

After, specific and shared representations will concatenate and pass through the ridge regression model for prediction.

V. EXPERIMENTS AND DISCUSSION

In the sections that follow, we discuss the results of our in-depth empirical study of the model and how it compares to other methods of predicting financial markets.

A. DATASETS

We performed extensive experiments on six different financial markets to demonstrate the predictive ability of our model. Each pair of financial markets that provided input to our model had the same period, as shown in Table 1.

TABLE 1. Datasets and their data points.

	Financial market	From	To	Data points
First inputs	Apple	2005-01-03	2022-12-30	4289
	Samsung	2005-01-03	2022-12-30	4289
Second inputs	Bitcoin	2017-12-01	2023-08-01	2069
	Ethereum	2017-12-01	2023-08-01	2069
Third inputs	Meta Platforms	2013-11-07	2022-10-27	2258
	X platform	2013-11-07	2022-10-27	2258

Features setting: We set the "Close" feature as the target value for our prediction, while the input was "Close". **Data processing**: The raw data for each feature is a one-dimensional time series; to achieve good data quality, we scale the features to unit variance and zero mean to decrease volatility. **Data setup**: We use the time feature with a fixed-size window to ensure that the values are taken at equal intervals. Accurate predictions for the next day can be obtained using the Ridge regression model.

B. EXPERIMENTAL DETAILS

The basic information about the components and setups is summarized in the following sections:

Metrics: Mean squared error (MSE) and mean absolute error (MAE) on each prediction window were the evaluative metrics in this work, through a 60/20/20 train/valid/test splitting of the dataset, as shown in the following equations.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2 \tag{7}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{8}$$

where N is the sample size, y_i is the observed value, and $\hat{y}_i$ is the predicted value.

Risk measurement: In addition to accuracy, it is imperative to evaluate pertinent facets of the trading process, such as the measurement of risks related to financial market returns, which may be calculated using both real value y and predicted value $\hat{y}$. We inspired a trading strategy from [50]. In this strategy, if the predicted value $\hat{y}_{t+1}$ surpasses the actual value y_t, we initiate a long position. Conversely, decrease the position index by one. Otherwise, we refrain from taking any position. The expression for the return R at time $t+1$ can be written:

$$R_{t+1} = \ln\frac{y_{t+1}}{y_t} \times sign(\hat{y}_{t+1} - y_t) \tag{9}$$

where $sign(.)$ is the sign function. To assess our model, we are going to employ two of the risks' metrics, which are volatility and max drawdown:

1) Volatility: This is a key metric used to quantify the extent of fluctuations in financial market returns throughout a certain time period [51]. It may be formulated as follows:

$$Volatility = \sigma(R_i) \tag{10}$$

where R_i is the return of the financial market at time i.

2) Max Drawdown: Participants use the financial indicator "Max Drawdown" to calculate their maximum loss for the entire trading period. The equation could be represented as:

$$MaxDrawdown = \max_{i<j}\frac{NV_j - NV_i}{NV_i} \tag{11}$$

where $NV(.)$is the total return.

Correlations:

To evaluate the disentanglement, we will use Spearman's rank correlation coefficient, which is used to assess the association between two time series without making assumptions about linearity [52]. The resulting value of ρ will range from -1 to 1 based on the strength relation between two sequences, i.e., result $= 1$ means the relation is strong, while close zero refers to a weak relation; otherwise, close to -1 indicates a perfect negative relation. The formula for calculating Spearman's rank correlation coefficient is as follows:

$$\rho = 1 - \frac{6\sum d_i^2}{N(N^2 - 1)} \tag{12}$$

where ρ is Spearman's rank correlation coefficient,$\sum d_i^2$is the sum of the squared differences between the ranks of the corresponding data points in the time series, and N is the number of observed data points.

Experimental setup: The experimental conditions are listed in Table 2.

TABLE 2. The experimental environments' setups.

	Configuration
Processor	2.9 GHz Quad-Core Intel Core i7
Operating system	macOS Ventura Version 13.2.1
Python version	Python 3.9.13 64-bit
Pytorch version	Pytorch 1.8.0

Hyper-parameter tuning: For our model, the backbone encoder used is the TS2VC [16]. The Adam optimizer was employed with an initial learning rate of $1e^{-4}$, and set the batch size to 32, and temperature to 0.07.

C. BASELINES

We compared our work with state-of-the-art approaches so as to demonstrate its proficiency. All their findings are reported based on our replication with dataset modifications for day, week, and year. The details of this are outlined below:

1) End-to-end learning methods: Two of the models that are based on end-to-end training, TS2VS, and Informer will be discussed as following:

a) Informer [5]: The proposed model is an enhanced version of the Transformer model [53] that addresses several limitations, such as the quadratic time complexity. Its primary objective is to provide efficient long-sequence time series prediction. This paper was used to predict ET,[1] ECL,[2] and Weather dataset,[3] and we used the open-source implementation[4] as is.

b) TS2Vec [16]: In this paper, the authors introduce a model that can learn contextual representations for arbitrary sub-series at various semantic levels. We used the open source implementation[5] as is.

2) Disentanglement methods: These models rely on extracting the key features from the raw data in order to learn a meaningful representation while filtering out irrelevant factors.

a) GLAD: This effort utilized the Informer, as a backbone, used time-frequency domains to capture global-local patterns, and employed contrastive learning to improve the results.

b) CoST [54] This model[6] was used to predict ETT, electricity, and weather by using TS2Vec [16] as the backbone, capturing seasonal-trend by time-frequency analysis, and learning them by contrastive loss.

D. INTERPRETABILITY AND EXPLAINABILITY OF OUR MODEL

The concepts of trustworthiness and transparency, which are crucial to making rational decisions, are intricately associated with the interpretability and explainability of machine learning. Despite their centrality in the financial market model, these concepts have gotten less academic attention than other disciplines, such as computer vision. Our approach is designed to attain both of these objectives, and we will explain in more detail how our goal is to go about doing so below.

1) INTERPRETABILITY: THE OVERALL UNDERSTANDABILITY OF THE MODEL

An interpretable model is related to its ability to offer a clear understanding of its prediction process, usually by exposing the relationships between input data and output predictions. Interpretability is concerned with making the model's inner workings more transparent and understandable [55]. First, we need to ascertain which representations are the most important, and then we can train the model to learn them [56]. Our key contribution is the realization that the factors underpinning the two markets' data, as a correlation point in a time series, may be represented by a shared pattern between them with separate representations for each. In other words, in prior works, like GLAD, the authors focused on a single market, ignoring the effects of others. In this work, we focused on the complex relationship between markets and how they affect each other. Since the value range for time series is perhaps boundless and a special token for raw data cannot be found, one of the merits of our work is that we performed disentanglement and augmentation in the latent space.

2) EXPLAINABILITY: WHY DID THE MODEL MAKE THIS PREDICTION?

The lack of transparency in conventional DL models is a major drawback, and as a result of this behavior, like the black box, even the people who designed it cannot explain the reasoning behind any given outcome. The amount of clarity with which the inner workings of a model can be understood is commonly referred to as its explainability. For this purpose, decomposability models, among which the disentanglement approach is one, are employed. Our model zooms in on specific predictions and offers deep explanations, demonstrating the manner in which the model came to its findings for a given set of data as well as providing extensive reasons for its conclusions. Our model includes the necessity for stakeholders, such as market participants, to have trust in, accept, and comprehend the model's predictions and the reasoning behind them.

TABLE 3. Results of closing price prediction on six datasets. The first-best results will be in bold, while the second-best will be contained in brackets.

	Financial market		Our Model	GLAD	CoST [55]	TS2Vec [16]	Informer [5]
First Inputs	Apple	MSE	0.00761	**0.00756**	[0.00758]	0.00820	0.00853
		MAE	0.0076	[0.00748]	**0.00746**	0.0083	0.0086
	Samsung	MSE	[0.00789]	**0.00786**	0.00793	0.00859	0.00849
		MAE	0.00953	**0.00941**	[0.00946]	0.0168	0.0158
Second Inputs	Meta platforms	MSE	**0.0061**	[0.00682]	0.00689	0.00832	0.00841
		MAE	**0.00889**	[0.00892]	0.00894	0.00953	0.00955
	X platform	MSE	[0.0092]	**0.00881**	0.00947	0.0138	0.0142
		MAE	**0.0109**	0.0113	[0.0111]	0.0231	0.0219
Third Inputs	Bitcoin	MSE	**0.0079**	[0.00831]	0.00842	0.00951	0.00957
		MAE	**0.00924**	0.00937	[0.00931]	0.00972	0.00966
	Ethereum	MSE	**0.00851**	[0.00855]	0.00857	0.00883	0.00879
		MAE	**0.00881**	0.00884	[0.00882]	0.00955	0.00946

E. RESULTS AND ANALYSIS

The experimental findings obtained from the six datasets are listed in Table 3, which provides a summary of our model results and the top-performing baselines. The best results are in bold, whereas the second-best results are enclosed within brackets. This table demonstrates that cross-domain disentanglement could be a promising solution to the challenge

[1] https://github.com/zhouhaoyi/ETDataset.git
[2] https://archive.ics.uci.edu/ml/ datasets/ElectricityLoadDiagrams 20112014.
[3] https://www.ncei.noaa.gov/ data/local-climatologicaldata/
[4] https://github.com/zhouhaoyi/Informer2020.git
[5] https://github.com/yuezhihan/ts2vec
[6] https://github.com/salesforce/CoST.git

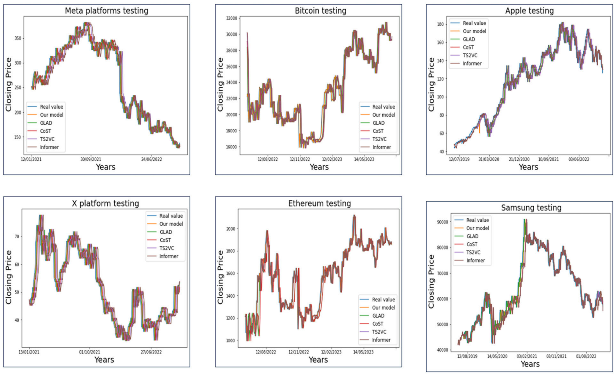

FIGURE 2. The predicted curves of data.

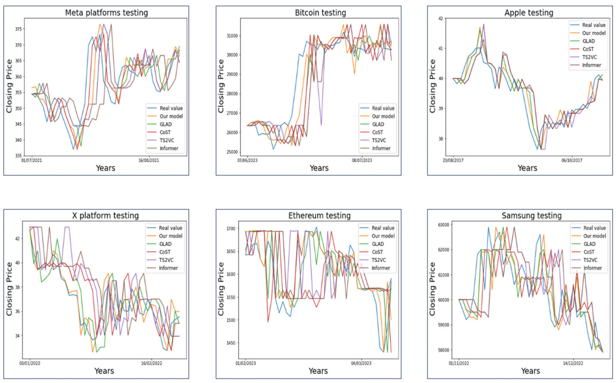

FIGURE 3. The predicted curves of a sample of data.

of data scarcity, particularly in cryptocurrencies, leveraging the concept of market integration. Figure 2 presents the fitted curves in the testing sets generated by our model and other models for six main financial market indices, while Figure 3 presents a sample of the fitted curves (40 days) in testing. Correlations among the three representations are very weak, as seen in Table 4, demonstrating that our disentanglement strategy has been successful. We used max drawdown and volatility, two standard measures of risk in the financial markets, to assess the model's efficacy and accuracy. The results for volatility and max drawdown are shown in Table 5, where they show that our model performs competitively.

TABLE 4. Results of correlations between financial markets and their representations.

		Input_x and Input_y	Input_x and x_specific	Input_y and y_specific	x_specific and share	y_specific and share	x_specific and y_specific
First Inputs	Apple Samsung	0.631	0.201	0.214	0.0182	0.0139	0.0081
Seconnd Inputs	Meta platforms X platform	0.556	0.181	0.113	0.0188	0.0132	0.0041
Third Inputs	Bitcoin Ethereum	0.621	0.156	0.228	0.0191	0.0120	0.0052

TABLE 5. The return(%), maximum drawdown(%) and volatility(%) for four datasets. The first-best results will be in bold, while the second-best will be contained in brackets.

Financial market	Models	Return(%)	Volatility(%)	MaxDrwon(%)
Apple	Our model	[79.21]	[1.292]	**-17.13**
	GLAD	**82.99**	**1.288**	[-17.51]
	CoST [54]	78.32	1.295	[-17.51]
	TS2VC [16]	67.99	1.297	-19.48
	Informer [5]	68.43	1.297	-19.34
Samsung	Our model	[78.32]	**1.42**	[-17.42]
	GLAD	**78.72**	1.433	**-17.39**
	CoST [54]	78.19	[1.432]	**-17.39**
	TS2VC [16]	60.99	1.441	-18.44
	Informer [5]	64.95	1.438	-17.55
Meta platforms	Our model	**51.34**	**1.258**	**-24.10**
	GLAD	44.20	[1.262]	-25.34
	CoST [54]	40.92	[1.262]	[-24.31]
	TS2VC [16]	35.99	1.268	-32.44
	Informer [5]	[47.64]	1.277	-31.11
X platforms	Our model	**43.02**	**1.38**	**-33.09**
	GLAD	[29.59]	[1.39]	[-34.12]
	CoST [54]	29.44	[1.39]	-35.30
	TS2VC [16]	6.98	1.41	-36.66
	Informer [5]	9.06	1.40	-36.65
Bitcoin	Our model	**56.38**	**1.222**	-29.13
	GLAD	44.92	[1.231]	**-29.44**
	CoST [54]	[44.95]	1.241	[-29.46]
	TS2VC [16]	17.43	1.246	-30.99
	Informer [5]	25.27	1.245	-30.87
Ethereum	Our model	**51.88**	**1.317**	**-31.29**
	GLAD	[50.49]	1.322	[-33.01]
	CoST [54]	50.36	[1.321]	-33.25
	TS2VC [16]	4.08	1.327	-34.99
	Informer [5]	16.41	1.326	-34.86

Ablation Study To evaluate the efficacy of each module in our work, we designed our model according to the following scenarios:

1) Backbone: Dilated CNN vs. multi layer perceptron MLP: Our work leverages a dilated CNN as an encoder to capture complex temporal patterns, while MLP was used in other works. For evolution, we replaced dilated CNNs with two-layer MLPs. The results showed that dilated CNN outperformed the MLP, as shown in Table 6. This result is related to the capability of dilations in the convolution layers to control the receptive field, thereby enabling the model to capture local and global patterns efficiently.
2) Share Feature Representations: Cosine Similarity vs Euclidean Distance Cosine similarity and Euclidean distance are the two most familiar metrics used as similarity metrics in a time series. We used cosine similarity in the share module to ascertain the similarity, while Euclidean distance was used in WTD to filter out similar information. For appraisal reasons, we swapped Cosine with Euclidean distance, i.e., we used Euclidean distance instead of Cosine and vice versa. Table 7 shows that the cosine method is better when we want to capture the similarity, and this may be because it is based on the angle between time series vectors

TABLE 6. Comparison of results obtained from dilated CNN and MLP encoders.

			Dilated CNN	MLP
First Inputs	Apple	MSE	0.00761	0.088
		MAE	0.0076	0.0923
	Samsung	MSE	0.00789	0.0901
		MAE	0.00953	0.0954
Second Inputs	Meta platforms	MSE	0.0061	0.0712
		MAE	0.00889	0.0781
	X platform	MSE	0.0092	0.0986
		MAE	0.0109	0.0992
Third Inputs	Bitcoin	MSE	0.0079	0.0493
		MAE	0.00924	0.0534
	Ethereum	MSE	0.00851	0.0602
		MAE	0.00881	0.0689

TABLE 7. Comparison of results obtained from swapping cosine similarity with Euclidean distance.

			Our model	After swap
First Inputs	Apple	MSE	0.00761	0.00952
		MAE	0.0076	0.0083
	Samsung	MSE	0.00789	0.00943
		MAE	0.00953	0.00996
Second Inputs	Meta platforms	MSE	0.0061	0.0089
		MAE	0.00889	0.00928
	X platform	MSE	0.0092	0.0162
		MAE	0.0109	0.0247
Third Inputs	Bitcoin	MSE	0.0079	0.00920
		MAE	0.00924	0.0159
	Ethereum	MSE	0.00851	0.00939
		MAE	0.00881	0.0093

TABLE 8. The relationship between markets correlation and prediction accuracy.

		Input_x and Input_y	MSE	MAE
Inputs	Apple	0.631	0.00761	0.0076
	Samsung		0.00789	0.00953
Inputs	Meta platforms	0.556	0.0061	0.00889
	X platform		0.0092	0.0109
Inputs	Bitcoin	0.621	0.0079	0.00924
	Ethereum		0.00851	0.00881
Inputs	Apple	0.210	0.099	0.0110
	Ethereum		0.0102	0.0115
Inputs	Apple	0.231	0.081	0.0123
	Meta platforms		0.089	0.0110

and takes into account the frequency of the data point, whereas the Euclidean distance measures the straight-line distance between two points, making it superior for grouping similar or dissimilar data points.

3) Relationship between the strength of correlation and prediction accuracy: We also compared the markets' correlation and the prediction's accuracy, as shown in Table 8. The results show that, for markets with a high correlation, the prediction accuracy will be better.

4) Contrastive learning: The objective of contrastive learning is to learn useful representations by contrasting similar and dissimilar patterns. To evaluate the role of this module in our work, we compared the model before and after using it, and the results are shown in Table 9. The results indicate that contrastive learning plays a crucial role in acquiring meaningful representations without explicit supervision. This becomes particularly significant in predicting financial markets, especially when leveraging cross-domains, where obtaining labeled data for domains can be challenging.

TABLE 9. The effect of contrastive learning on prediction accuracy.

			With contrastive	Without contrastive
Inputs	Apple	MSE	0.00761	0.00913
		MAE	0.0076	0.0120
	Samsung	MSE	0.00789	0.00991
		MAE	0.00953	0.0196
Inputs	Meta platforms	MSE	0.0061	0.0152
		MAE	0.00889	0.00925
	X platform	MSE	0.0092	0.0154
		MAE	0.0109	0.0192
Inputs	Bitcoin	MSE	0.0079	0.00937
		MAE	0.00924	0.0158
	Ethereum	MSE	0.00851	0.00972
		MAE	0.00881	0.00967

VI. CONCLUSION AND FUTURE WORK

Our study shows that our model is better at predicting financial markets than the end-to-end methods. This is true for both accuracy metrics (2.53% improvement in MSE and 1.21% improvement in MAE) and net value analysis, as well as for measuring financial risk. The results also shed light on how the use of a multi-market model improves the prediction accuracy, particularly in cases where data points are scarce. We conducted experimental tests to validate our model, and the results demonstrated that our technique has a competitive advantage over the existing state-of-the-art methodologies. For financial market prediction, we introduced our model, a framework that deals with two markets and disentangles their representations into shared and specific patterns. The empirical results demonstrated that contrastive learning may improve both learning and the prediction model. Our results make it easier for real-world users to understand what is going on by showing where the variance and influencing factors come from. In our future research, we will investigate (1) whether the shared representation has any correlation or causality with the share representation of the other two different markets; and (2) whether our model has the capability to extend to other time series datasets.

REFERENCES

[1] M. N. Ashtiani and B. Raahemi, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review," *Expert Syst. Appl.*, vol. 217, May 2023, Art. no. 119509.

[2] Y. Zhao and G. Yang, "Deep learning-based integrated framework for stock price movement prediction," *Appl. Soft Comput.*, vol. 133, Jan. 2023, Art. no. 109921.
[3] X. Zhang, C. Zhong, J. Zhang, T. Wang, and W. W. Y. Ng, "Robust recurrent neural networks for time series forecasting," *Neurocomputing*, vol. 526, pp. 143–157, Mar. 2023.
[4] Z. Hajirahimi and M. Khashei, "Hybridization of hybrid structures for time series forecasting: A review," *Artif. Intell. Rev.*, vol. 56, no. 2, pp. 1201–1261, Feb. 2023.
[5] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond efficient transformer for long sequence time-series forecasting," 2020, *arXiv:2012.07436*.
[6] S. Verma, S. P. Sahu, and T. P. Sahu, "Stock market forecasting with different input indicators using machine learning and deep learning techniques: A review," *Eng. Lett.*, vol. 31, no. 1, pp. 1–12, 2023.
[7] K. Nyakurukwa and Y. Seetharam, "The evolution of studies on social media sentiment in the stock market: Insights from bibliometric analysis," *Sci. Afr.*, vol. 20, Jul. 2023, Art. no. e01596.
[8] A. N. Sayed, Y. Himeur, and F. Bensaali, "From time-series to 2D images for building occupancy prediction using deep transfer learning," *Eng. Appl. Artif. Intell.*, vol. 119, Mar. 2023, Art. no. 105786.
[9] M. R. A. Khalil and A. A. Bakar, "A comparative study of deep learning algorithms in univariate and multivariate forecasting of the Malaysian stock market," *Sains Malaysiana*, vol. 52, no. 3, pp. 993–1009, Mar. 2023.
[10] Q. Wen, L. Sun, F. Yang, X. Song, J. Gao, X. Wang, and H. Xu, "Time series data augmentation for deep learning: A survey," 2020, *arXiv:2002.12478*.
[11] H. V. Dudukcu, M. Taskiran, Z. G. C. Taskiran, and T. Yildirim, "Temporal convolutional networks with RNN approach for chaotic time series prediction," *Appl. Soft Comput.*, vol. 133, Jan. 2023, Art. no. 109945.
[12] Y. Cai, H.-W. Chang, and T. Chang, "Evaluating time-varying Granger causality between U.S.-China political relation changes and China stock market," *Finance Res. Lett.*, vol. 55, Jul. 2023, Art. no. 103918.
[13] G. Wilson, J. R. Doppa, and D. J. Cook, "CALDA: Improving multi-source time series domain adaptation with contrastive adversarial learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14208–14221, Dec. 2023.
[14] H. Tang, L. Wu, W. Liu, and J. Bian, "ADD: Augmented disentanglement distillation framework for improving stock trend forecasting," 2020, *arXiv:2012.06289*.
[15] N. Musmeci, T. Aste, and T. Di Matteo, "Relation between financial market structure and the real economy: Comparison between clustering methods," 2014, *arXiv:1406.0496*.
[16] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "TS2Vec: Towards universal representation of time series," 2021, *arXiv:2106.10466*.
[17] Y. Li, K. Li, C. Chen, X. Zhou, Z. Zeng, and K. Li, "Modeling temporal patterns with dilated convolutions for time-series forecasting," *ACM Trans. Knowl. Discovery Data*, vol. 16, no. 1, pp. 1–22, Feb. 2022.
[18] L. Chen, W. Chen, B. Wu, Y. Zhang, B. Wen, and C. Yang, "Learning from multiple time series: A deep disentangled approach to diversified time series forecasting," 2021, *arXiv:2111.04942*.
[19] R. J. Kate, "Using dynamic time warping distances as features for improved time series classification," *Data Mining Knowl. Discovery*, vol. 30, no. 2, pp. 283–312, Mar. 2016.
[20] L. Muda, M. Begam, and I. Elamvazuthi, "Voice recognition algorithms using mel frequency cepstral coefficient (MFCC) and dynamic time warping (DTW) techniques," 2010, *arXiv:1003.4083*.
[21] C.-J. Hsu, K.-S. Huang, C.-B. Yang, and Y.-P. Guo, "Flexible dynamic time warping for time series classification," *Proc. Comput. Sci.*, vol. 51, pp. 2838–2842, Jan. 2015.
[22] A. N. Fernandes, "An evolutionary computing approach to financial portfolio management based on growth stocks & sector/industry distribution," Instituto Superior Técnico, Lisbon, Portugal, 2016.
[23] B. Sriman and S. G. Kumar, "Decentralized finance (DeFi): The future of finance and Defi application for Ethereum blockchain based finance market," in *Proc. Int. Conf. Adv. Comput., Commun. Appl. Informat. (ACCAI)*, Jan. 2022, pp. 1–9.
[24] K. Czech, M. Wielechowski, P. Kotyza, I. Benešová, and A. Laputková, "Shaking stability: COVID-19 impact on the visegrad group Countries' financial markets," *Sustainability*, vol. 12, no. 15, p. 6282, Aug. 2020.
[25] J. Eapen, D. Bein, and A. Verma, "Novel deep learning model with CNN and bi-directional LSTM for improved stock market index prediction," in *Proc. IEEE 9th Annu. Comput. Commun. Workshop Conf. (CCWC)*, 2019, pp. 264–270.
[26] L. Ni, Y. Li, X. Wang, J. Zhang, J. Yu, and C. Qi, "Forecasting of forex time series data based on deep learning," *Proc. Comput. Sci.*, vol. 147, pp. 647–652, Jan. 2019.
[27] M. Yekrangi and N. Abdolvand, "Financial markets sentiment analysis: Developing a specialized lexicon," *J. Intell. Inf. Syst.*, vol. 57, no. 1, pp. 127–146, Aug. 2021.
[28] S. Chen, H. Yao, F. Qiao, Y. Ma, Y. Wu, and J. Lu, "Vehicles driving behavior recognition based on transfer learning," *Expert Syst. Appl.*, vol. 213, Mar. 2023, Art. no. 119254.
[29] T. Yang, X. Yu, N. Ma, Y. Zhao, and H. Li, "A novel domain adaptive deep recurrent network for multivariate time series prediction," *Eng. Appl. Artif. Intell.*, vol. 106, Nov. 2021, Art. no. 104498.
[30] K. Bisht and A. Kumar, "A portfolio construction model based on sector analysis using Dempster–Shafer evidence theory and Granger causal network: An application to national stock exchange of India," *Expert Syst. Appl.*, vol. 215, Apr. 2023, Art. no. 119434.
[31] M. Costola, O. Hinz, M. Nofer, and L. Pelizzon, "Machine learning sentiment analysis, COVID-19 news and stock market reactions," *Res. Int. Bus. Finance*, vol. 64, Jan. 2023, Art. no. 101881.
[32] Y. Ma, R. Mao, Q. Lin, P. Wu, and E. Cambria, "Multi-source aggregated classification for stock price movement prediction," *Inf. Fusion*, vol. 91, pp. 515–528, Mar. 2023.
[33] C. Zhong, W. Du, W. Xu, Q. Huang, Y. Zhao, and M. Wang, "LSTM-ReGAT: A network-centric approach for cryptocurrency price trend prediction," *Decis. Support Syst.*, vol. 169, Jun. 2023, Art. no. 113955.
[34] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu, "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," *Appl. Soft Comput.*, vol. 90, May 2020, Art. no. 106181.
[35] W. Chen, M. Jiang, W.-G. Zhang, and Z. Chen, "A novel graph convolutional feature based convolutional neural network for stock trend prediction," *Inf. Sci.*, vol. 556, pp. 67–94, May 2021.
[36] A. Hyvärinen, I. Khemakhem, and H. Morioka, "Nonlinear independent component analysis for principled disentanglement in unsupervised deep learning," *Patterns*, vol. 4, no. 10, Oct. 2023, Art. no. 100844.
[37] M. Wang, Y. Panagakis, P. Snape, and S. P. Zafeiriou, "Disentangling the modes of variation in unlabelled data," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 11, pp. 2682–2695, Nov. 2018.
[38] S. Yang, T. Guo, Y. Wang, and C. Xu, "Adversarial robustness through disentangled representations," in *Proc. AAAI Conf. Artif. Intell.*, May 2021, vol. 35, no. 4, pp. 3145–3153.
[39] H. Hwang, G.-H. Kim, S. Hong, and K.-E. Kim, "Variational interaction information maximization for cross-domain disentanglement," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 22479–22491.
[40] S. Hao, K. Han, and K.-Y. K. Wong, "Learning attention as disentangler for compositional zero-shot learning," 2023, *arXiv:2303.15111*.
[41] S. An, K. Song, and J.-J. Jeon, "Causally disentangled generative variational AutoEncoder," 2023, *arXiv:2302.11737*.
[42] H. M. Abdulsahib and F. Ghaderi, "GLAD: Global–local approach; disentanglement learning for financial market prediction," *IET Signal Process.*, vol. 2023, pp. 1–15, Nov. 2023.
[43] N. Tawara, A. Ogawa, T. Iwata, H. Ashikawa, T. Kobayashi, and T. Ogawa, "Multi-source domain generalization using domain attributes for recurrent neural network language models," *IEICE Trans. Inf. Syst.*, vol. 105, no. 1, pp. 150–160, 2022.
[44] X. Jiang, S. Pan, J. Jiang, and G. Long, "Cross-domain deep learning approach for multiple financial market prediction," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
[45] R. S. Rawlin, S. R. Ramachandran, and Rekha, "Causal relationships of the stock market with other asset classes: An Indian perspective," in *Proc. 13th Int. Conf. E-Bus., Manag. Econ.*, 2022, pp. 392–398.
[46] Q. Tian, P. Shang, and G. Feng, "The similarity analysis of financial stocks based on information clustering," *Nonlinear Dyn.*, vol. 85, no. 4, pp. 2635–2652, Sep. 2016.
[47] H. Hu, M. Tang, and C. Bai, "DATSING: Data augmented time series forecasting with adversarial domain adaptation," in *Proc. 29th ACM Int. Conf. Inf. Knowl. Manage.*, Oct. 2020, pp. 2061–2064.
[48] X.-Y. Liu, Z. Xia, J. Rui, J. Gao, H. Yang, M. Zhu, C. Wang, Z. Wang, and J. Guo, "FinRL-Meta: Market environments and benchmarks for data-driven financial reinforcement learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 35, 2022, pp. 1835–1849.

[49] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, ''Momentum contrast for unsupervised visual representation learning,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 9726–9735.
[50] C. Wang, Y. Chen, S. Zhang, and Q. Zhang, ''Stock market index prediction using deep transformer model,'' *Expert Syst. Appl.*, vol. 208, Dec. 2022, Art. no. 118128.
[51] H. Guo and R. Savickas, ''Idiosyncratic volatility, stock market volatility, and expected stock returns,'' *J. Bus. Econ. Statist.*, vol. 24, no. 1, pp. 43–56, Jan. 2006.
[52] C. Rong, ''Time series similarity evaluation based on Spearman's correlation coefficients and distance measures,'' in *Cloud Computing and Big Data*. Cham, Switzerland: Springer, 2015, p. 319.
[53] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, ''Attention is all you need,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.
[54] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi, ''CoST: Contrastive learning of disentangled seasonal-trend representations for time series forecasting,'' 2022, *arXiv:2202.01575*.
[55] T. Miller, ''Explanation in artificial intelligence: Insights from the social sciences,'' *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.
[56] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, ''Definitions, methods, and applications in interpretable machine learning,'' *Proc. Nat. Acad. Sci. USA*, vol. 116, no. 44, pp. 22071–22080, Oct. 2019.

HUMAM M. ABDULSAHIB received the Ph.D. degree from Tarbiat Modares University, Iran, in 2024. He did this research with the Human-Computer Interaction Laboratory, Computer Engineering Department, Faculty of Electrical and Computer Engineering, Tarbiat Modares University. His research interests include deep learning and financial technology.

FOAD GHADERI received the Ph.D. degree from Cardiff University, U.K., in 2010. From 2011 to 2014, he was with the University of Bremen and DFKI Bremen, Germany, and a Postdoctoral Researcher. He is currently with the Faculty of Electrical and Computer Engineering, Tarbiat Modares University, and the Director of the Human-Computer Interaction Laboratory. His research interests include machine learning, data mining, and statistical signal processing.