Clustering Behavioral Regimes in the Iranian Capital Market Using Self-Organizing Maps (SOM): An Unsupervised Learning Approach (A Proof-of-Concept Study)

Mohsen Abbasi Sorkhkolaei[1], Seyed Ali Nabavi Chashmi[2,*] and Erfan Memarian[3]

[1] Ph.D. Student in Financial Management, Bab.C., Islamic Azad University, Babol, Iran;
[2] Department of Financial Management, Bab.C., Islamic Azad University, Babol, Iran;
[3] Department of Economic, Bab.C., Islamic Azad University, Babol, Iran;

* Correspondence: a.nabavichashmi@iau.ac.ir

Citation: Abbasi Sorkhkolaei, M., Nabavi Chashmi, S. A., & Memarian, E. (2027). Clustering Behavioral Regimes in the Iranian Capital Market Using Self-Organizing Maps (SOM): An Unsupervised Learning Approach (A Proof-of-Concept Study). *Business, Marketing, and Finance Open*, 4(9), 1-17.

Received: 19 June 2026
Revised: 04 October 2026
Accepted: 12 October 2026
Initial Publication: 14 October 2026
Final Publication: 01 September 2027

Copyright: © 2027 by the authors. Published under the terms and conditions of Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) License.

Abstract: This study aimed to evaluate, as a proof of concept, the capability of a Self-Organizing Map (SOM) to identify and characterize latent behavioral regimes in multidimensional Iranian capital-market data through an unsupervised learning framework. This quantitative, computational, exploratory, and applied study used multivariate financial observations representing dimensions such as return, volatility, trading activity, market conditions, and selected firm-level characteristics. After data cleaning, assessment of missing and extreme values, temporal alignment, and scaling, the observations were entered into a two-dimensional 6 × 6 SOM implemented in Python using the MiniSom library. The model used a Gaussian neighborhood function, an initial learning rate of 0.50, and 1,000 training iterations. Cluster extraction was based on the topological organization of the trained map and U-Matrix patterns. Model performance was assessed using training and quantization error, intra-cluster distance, visual inspection of the U-Matrix, and cluster-validity criteria including the Silhouette coefficient and Davies–Bouldin index. The network training error decreased from 0.482 to 0.067, representing an approximate reduction of 86.1% and indicating substantial convergence during learning. U-Matrix analysis revealed nonuniform inter-neuron distances and distinguishable topological regions, supporting the existence of heterogeneous latent structures. Four final clusters were extracted. Cluster 1 contained 175 observations and showed the greatest internal cohesion with a mean intra-cluster distance of 0.23; Cluster 2 contained 152 observations with a mean distance of 0.31; Cluster 3 included 108 observations with a mean distance of 0.44; and Cluster 4 included 85 observations and exhibited the greatest heterogeneity with a mean intra-cluster distance of 0.51. The findings indicate that SOM can provide a meaningful unsupervised representation of heterogeneous behavioral regimes in Iranian capital-market data and may serve as a useful exploratory foundation for subsequent regime-based forecasting, portfolio management, and risk-monitoring studies.

Keywords: Iranian Capital Market; Self-Organizing Map; SOM; Unsupervised Learning; Behavioral Regimes; Clustering; U-Matrix; Machine Learning

1. Introduction

Capital markets are complex adaptive systems in which prices, trading activity, liquidity, volatility, investor expectations, institutional constraints, information flows, and behavioral reactions continuously interact. Unlike simplified representations that assume stable relationships among market variables, actual capital-market behavior frequently exhibits nonlinearities, abrupt transitions, asymmetric reactions, and heterogeneous behavioral patterns across time and securities. These characteristics become particularly important in markets where informational efficiency is uneven, investor behavior is sensitive to cognitive and emotional factors, and market responses to new information may vary

according to prevailing conditions. From this perspective, market behavior is better understood not merely as a single continuous process, but as a sequence of potentially distinguishable regimes in which the joint configuration of return, risk, trading activity, liquidity, market direction, and firm-level characteristics changes over time. Research on investor behavior has increasingly emphasized that market outcomes are shaped by the interaction of information type, market symmetry, perception, and behavioral responses, suggesting that different configurations of these factors may generate distinct patterns of capital-market behavior [1]. Similarly, evidence from behavioral finance indicates that emotional and cognitive components can play an important role in capital-market crises and in the formation of market reactions under conditions of uncertainty, reinforcing the need for analytical frameworks capable of capturing heterogeneity rather than imposing a single average behavioral structure on all observations [2]. Investor reactions may also change in response to external policy signals, including taxation and fiscal-policy adjustments, which can influence portfolio reallocation, risk preferences, and broader capital-market responses [3]. Taken together, these findings support the view that market states are multidimensional, behaviorally heterogeneous, and potentially regime-dependent.

The identification of behavioral regimes is important because conventional statistical models often require strong assumptions regarding functional form, distributional properties, linearity, independence, or pre-specified relationships among dependent and independent variables. Such assumptions may be difficult to sustain in financial environments characterized by heavy-tailed distributions, volatility clustering, structural breaks, nonlinear relationships, and shifting correlations. In addition, supervised learning methods require predefined target variables or class labels, whereas in many capital-market applications the underlying states of the system are not known in advance. Under these conditions, unsupervised learning provides an alternative analytical framework because it seeks to discover latent structures directly from the data without requiring prior labels. Clustering, in particular, has been used to identify groups of financial observations with similar characteristics and to improve the organization and interpretation of high-dimensional market information. Recent research has shown that clustering methods can support stock-market analysis by revealing structurally similar observations that may not be readily distinguishable through conventional univariate techniques [4]. Clustering has also been applied to portfolio construction and risk reduction, where grouping assets according to common characteristics or dependencies can contribute to diversification and more efficient portfolio allocation [5]. Likewise, continuous trend features have been combined with k-means clustering to construct stock portfolios, illustrating how unsupervised learning can transform complex time-dependent financial information into more interpretable groups [6]. These applications demonstrate that clustering is not merely a descriptive technique but can serve as a computational foundation for identifying structural regularities that may subsequently support financial decision-making.

An important implication of this literature is that the informational content of financial data depends not only on the number of available variables but also on the manner in which those variables are organized and transformed. Raw financial observations often contain redundancy, noise, scale differences, nonlinear relationships, and latent patterns that are not immediately observable. Therefore, converting raw data into structured information is a central analytical challenge. Research has demonstrated that clustering can enhance stock-return forecasting by organizing market observations into more informative groups before subsequent predictive modeling is performed [7]. This distinction between data and information is especially relevant in financial markets, where large volumes of numerical observations do not necessarily translate into greater explanatory value unless meaningful structure can be extracted. Earlier research on stock-market prediction has

similarly shown that artificial intelligence and machine-learning methods increasingly complement traditional fundamental and technical analysis by identifying complex patterns in financial variables and by accommodating nonlinear relationships that are difficult to model through conventional approaches [8]. Within the Iranian capital market, prior research has also applied computational classification methods to the profitability of companies listed on the Tehran Stock Exchange, showing that data-mining approaches such as decision trees can contribute to the grouping and prediction of firm-level financial characteristics [9]. However, supervised classification differs fundamentally from the unsupervised identification of market regimes because the latter does not begin with predefined profitability categories or predetermined class labels.

Among unsupervised neural-network methods, the Self-Organizing Map (SOM) offers particular advantages for the analysis of complex financial data. Originally developed by Kohonen, the SOM is a topology-preserving neural-network architecture that maps high-dimensional input vectors onto a lower-dimensional, usually two-dimensional, lattice while attempting to preserve neighborhood relationships among observations [10]. Unlike conventional clustering procedures that provide only cluster assignments, the SOM combines vector quantization, dimensionality reduction, nonlinear projection, and visual representation. During training, each input observation is compared with the weight vectors of the map neurons, and the neuron whose weight vector is most similar to the observation is designated as the Best Matching Unit. The weight of the Best Matching Unit and those of neighboring neurons are subsequently updated, producing a map in which similar observations tend to occupy neighboring regions. This topological organization is particularly useful in financial research because it permits researchers to examine both local similarity and broader structural separation within the same representation. The SOM can therefore reveal gradual transitions, dense regions, sparse areas, and potential boundaries among latent market states that may be difficult to observe through traditional clustering algorithms.

The usefulness of SOM is not restricted to static cross-sectional data. Extensions of the method have been developed for time-series clustering, enabling complex sequences to be represented according to similarities in their temporal structure [11]. This capacity is directly relevant to financial-market analysis because the behavior of returns, volatility, volume, liquidity, and other variables often depends on temporal dynamics rather than solely on contemporaneous levels. Moreover, self-organizing architectures have been integrated with fuzzy and deep-learning approaches for financial trading decisions, indicating that the underlying concept remains useful even in more advanced machine-learning environments [12]. Such research suggests that SOM-based systems can contribute to the identification of trading environments in which observations cannot be represented adequately through rigid boundaries. In financial markets, regime transitions may be gradual, overlapping, or ambiguous; consequently, a topology-preserving method can offer an informative intermediate representation between purely visual exploration and hard cluster assignment.

A further methodological advantage of SOM is its compatibility with post-training cluster extraction. The neural map itself provides a structured representation of the data, but researchers may subsequently identify higher-level clusters by examining distances among neighboring neurons or by applying clustering techniques to trained prototype vectors. Vesanto and Alhoniemi demonstrated that clustering of the SOM can provide a useful framework for extracting interpretable groups from the organized map structure [13]. One of the most widely used visual tools for this purpose is the Unified Distance Matrix, or U-Matrix, which displays the distances among neighboring neurons. Regions characterized by relatively small distances can be interpreted as areas of local similarity, whereas larger distances may indicate potential cluster boundaries. In capital-market applications, such boundaries may correspond to transitions among behavioral regimes characterized by different combinations of

return, volatility, market activity, liquidity, or fundamental conditions. Importantly, however, visual interpretation alone is insufficient for establishing the validity of a clustering solution. The resulting structure should be evaluated with formal internal validation criteria that assess both within-cluster cohesion and between-cluster separation.

Two classical validation measures are particularly relevant in this context. The Silhouette coefficient evaluates the extent to which an observation is well matched to its assigned cluster relative to alternative clusters, thereby providing a combined measure of cohesion and separation [14]. Positive and comparatively high Silhouette values generally indicate that observations are located appropriately within their clusters, whereas values near zero may suggest boundary observations or overlapping groups. Silhouette analysis has remained widely used in contemporary machine-learning research because it provides an intuitive and quantitative approach to evaluating the performance of clustering solutions [15]. The Davies–Bouldin index provides a complementary perspective by comparing within-cluster scatter with between-cluster separation; lower values indicate more compact and better-separated clusters [16]. The simultaneous use of these criteria is important because no single clustering metric captures all dimensions of model quality. In SOM applications, these measures can be combined with quantization error, topographic inspection, neuron occupancy, U-Matrix visualization, and repeated training under alternative model configurations to provide a more robust assessment of the extracted structure.

The application of these techniques is particularly relevant to the Iranian capital market. The Tehran Stock Exchange and Iran Fara Bourse operate in an environment characterized by changing macroeconomic conditions, regulatory interventions, varying levels of liquidity, heterogeneous investor participation, and behavioral reactions to market information. Evidence from Iranian behavioral-finance research suggests that market participants' emotional and cognitive processes may contribute to the formation of financial behavior and to market instability during crisis conditions [2]. Research based on qualitative modeling has also indicated that capital-market behavior can vary according to the type of information available and according to symmetrical or asymmetrical market conditions [1]. These characteristics provide a conceptual basis for expecting that Iranian financial observations may not form a homogeneous population and that distinct combinations of market and firm-level variables may reflect different behavioral states. If such states can be identified without imposing predetermined labels, they may provide a useful exploratory framework for understanding market heterogeneity.

At the same time, the identification of market regimes should not be confused with prediction. A clustering model can reveal that certain observations are similar to one another and distinct from other observations, but this does not necessarily imply that the resulting clusters can forecast future returns, crises, or trading opportunities. This distinction is methodologically important because machine-learning studies in finance sometimes combine clustering, classification, and forecasting within a single analytical narrative. Reviews of stock-market prediction have shown that predictive performance depends strongly on feature selection, data quality, evaluation design, and model specification [8]. Similarly, clustering-based approaches can improve downstream prediction under certain conditions, but the usefulness of clustering must be established independently before predictive claims are introduced [7]. A proof-of-concept SOM study should therefore initially focus on whether the data exhibit interpretable, stable, and quantitatively defensible topological organization. Predictive value can subsequently be examined using temporal validation, explicit target variables, benchmark models, and out-of-sample testing.

This distinction is also relevant when interpreting previous Iranian research. The use of decision-tree methods to cluster and predict the profitability of Tehran Stock Exchange companies demonstrates that computational methods can be meaningfully applied to Iranian financial data [9]. However, profitability classification addresses a different problem from unsupervised market-regime identification. A supervised model is trained to separate

observations according to a known target, while SOM seeks to discover organization that is intrinsic to the multivariate input structure. The latter approach is especially useful when the researcher does not know in advance how many economically meaningful regimes exist, where their boundaries lie, or which combinations of variables define them. In this sense, SOM provides an exploratory bridge between high-dimensional financial measurement and subsequent theoretical interpretation.

The potential managerial and decision-support implications are also important. Financial managers, investment analysts, portfolio managers, and regulators frequently rely on indicators that are interpreted independently or through fixed threshold rules. A regime-based representation can provide an alternative perspective by showing how multiple financial characteristics occur jointly. For example, a period of high returns accompanied by high volatility and abnormal trading volume may have a different interpretation from a period of high returns associated with stable volatility and normal trading activity. Clustering allows these multivariate configurations to be considered jointly rather than reducing them to a single indicator. Research on portfolio optimization demonstrates that clustering can support risk reduction by organizing assets into groups that reflect structural similarity and diversification potential [5]. Likewise, clustering of trend features can improve the construction of differentiated portfolios [6], while density-based approaches can uncover complex structures in stock-market data that may be missed by simpler partitioning methods [4]. The implication is that regime identification may eventually contribute to risk monitoring, portfolio allocation, market surveillance, or adaptive investment strategies, although such applications require independent validation beyond an initial proof-of-concept study.

Despite these methodological opportunities, several challenges must be recognized. SOM results are sensitive to preprocessing, feature selection, map dimensions, initialization, learning-rate schedules, neighborhood parameters, training length, and random seeds. Financial variables also differ substantially in scale and distribution, meaning that return, volatility, transaction volume, trading value, market indices, and accounting ratios cannot be entered into a distance-based model without appropriate transformation. Moreover, extreme observations require careful treatment because they may represent either data errors or genuine market shocks. Excessive outlier removal could erase precisely the observations that define crisis or transition regimes. Consequently, preprocessing should be transparent, reproducible, and supported by sensitivity analysis. The resulting clusters should also be interpreted cautiously because numerical separation does not automatically establish economic meaning. Cluster labels should emerge from the observed feature profiles and should be treated as post hoc interpretations rather than predetermined categories.

The proof-of-concept orientation is therefore appropriate for an initial application of SOM to behavioral-regime identification in the Iranian capital market. Such a design can determine whether multidimensional financial observations generate a coherent map, whether identifiable cluster boundaries emerge, whether the extracted clusters differ in compactness and internal similarity, and whether the resulting structure can be interpreted in economically meaningful terms. It also permits evaluation of the methodological feasibility of combining SOM visualization with formal validation indicators such as quantization error, Silhouette analysis, and the Davies–Bouldin index [14-16]. Rather than making premature claims regarding forecasting superiority, the study can establish the analytical foundation required for later work involving larger verified datasets, repeated temporal samples, supervised prediction, portfolio testing, and comparative machine-learning benchmarks.

Accordingly, the aim of this study was to evaluate, as a proof of concept, the capability of a Self-Organizing Map to identify and characterize latent behavioral regimes in multidimensional Iranian capital-market data through an unsupervised learning framework.

2. Methods and Materials

This study employed a quantitative, computational, exploratory, and applied research design to examine the feasibility of identifying distinct behavioral regimes in the Iranian capital market through a Self-Organizing Map (SOM), an unsupervised artificial neural network. The study was specifically conceptualized as a proof-of-concept investigation rather than a confirmatory causal analysis. Accordingly, its principal objective was not to estimate the causal effect of one independent variable on a dependent variable or to test conventional directional hypotheses, but to determine whether multivariate financial observations could be organized into internally coherent and externally distinguishable patterns based on their similarities in a multidimensional feature space. The fundamental unit of analysis was therefore a multivariate financial observation represented by a vector of market, trading, risk, and, where applicable, firm-level fundamental characteristics. Each observation was subjected to preprocessing and subsequently presented to the SOM without a predetermined class label. This design permitted the network to learn the latent topology of the data, project multidimensional financial information onto a two-dimensional representation, and reveal potentially meaningful behavioral configurations that could be interpreted as distinct market regimes. In methodological terms, emphasis was therefore placed on pattern discovery, dimensionality reduction, topology preservation, clustering quality, and model stability rather than on parametric hypothesis testing or statistical significance alone.

The conceptual population of the study comprised firms listed on the Tehran Stock Exchange and Iran Fara Bourse and the corresponding financial and trading observations that could characterize their behavior under different market conditions. However, because the study was designed as a proof of concept and complete verified market-level raw data were not available for all stages of model development, part of the analytical dataset was constructed or provisionally specified for demonstrating the proposed SOM-based procedure. Consequently, the numerical findings generated within this framework should not be interpreted as definitive estimates of the entire Iranian capital market, nor should they be generalized directly to all listed firms, industries, or trading periods. Instead, the analytical sample served as a methodological demonstration of how an unsupervised neural architecture may be employed for identifying latent behavioral regimes in capital-market data. For a fully reproducible empirical extension, the final dataset should be obtained from clearly identified official sources such as the Tehran Securities Exchange Technology Management Company, the Comprehensive Database of All Listed Companies, or another authenticated capital-market database, and the precise extraction period, firm inclusion and exclusion criteria, handling of listing and delisting events, treatment of trading suspensions, and final number of observations after data cleaning should be documented explicitly. This distinction between the conceptual population and the proof-of-concept analytical dataset was maintained throughout the study to avoid attributing provisional numerical patterns to the Iranian capital market as a whole.

The analytical procedure was organized sequentially around variable selection, data quality assessment, preprocessing, SOM training, extraction of the topological structure, and cluster validation. The candidate input features were selected to capture several economically meaningful dimensions of capital-market behavior. Performance-related variables included measures such as stock return or return on assets, depending on the analytical level of the observation. Risk was represented primarily through return volatility or comparable dispersion measures capable of distinguishing relatively stable periods from highly volatile market states. Trading activity was reflected through variables such as trading volume and trading value, which provided information regarding market participation, liquidity, and transaction intensity. Broader market conditions could be

represented by the overall market index, industry indices, or their derived return measures to provide contextual information about systematic market movements. Where firm-level fundamental information was incorporated, additional variables could include profitability, leverage, liquidity, operating performance, and firm size. Because the objective of a proof-of-concept SOM is to evaluate whether multidimensional financial information generates interpretable topological patterns, these variables were treated as features rather than as predefined dependent and independent variables. At the same time, methodological consistency required that the final empirical implementation rely on a single, explicitly documented feature set rather than changing the composition of the input vector across different stages of the analysis.

The data collection and preparation framework was designed around the acquisition and transformation of quantitative financial and market information into a machine-readable feature matrix suitable for unsupervised learning. In a full empirical implementation, raw information should be retrieved from official and reproducible databases containing historical prices, returns, transaction volumes, transaction values, market indices, corporate financial statements, and other relevant firm-level indicators. The present proof-of-concept study used the same conceptual structure while acknowledging that some input values were provisional rather than derived from a complete independently reproducible market database. The resulting dataset was organized so that each row represented one multivariate financial observation and each column represented one analytical feature. Particular attention was paid to ensuring that observations referring to different variables or sources were temporally aligned before the SOM was trained. This step is especially important for capital-market research because trading variables may be available at daily frequency, accounting variables may be reported quarterly or annually, and index-related information may follow a different reporting structure. Therefore, the construction of the analytical matrix required conceptual consistency in the observation interval and alignment of features to prevent the network from learning artificial patterns created by asynchronous measurements.

Data preprocessing constituted a fundamental component of the measurement procedure because SOM learning is distance-based and is therefore highly sensitive to the scale, distribution, and quality of the input variables. The data were first examined for missing values, inconsistent observations, coding errors, duplicated records, and implausible values. Missing observations were assessed in relation to their source and frequency rather than being removed automatically. Similarly, extreme observations were evaluated before treatment because unusually large financial values can represent either data-entry problems or genuine episodes of market stress. Mechanical elimination of all statistical outliers was deliberately avoided as a general strategy because doing so could suppress precisely those observations associated with crisis, speculative activity, extreme volatility, or abrupt liquidity changes that may constitute distinct behavioral regimes. Where an extreme value appeared attributable to a recording or processing error, correction or exclusion was appropriate; where it plausibly reflected an authentic market event, the observation was retained unless a clearly specified robustness procedure required otherwise. This distinction was particularly important because the purpose of the model was to discover heterogeneous market states rather than normalize the dataset by eliminating economically meaningful extremes.

The numerical features were transformed before SOM training to prevent variables with large measurement ranges from dominating Euclidean distance calculations. Min–Max normalization and Z-score standardization were considered within the preprocessing framework. Min–Max transformation rescales a feature relative to its observed minimum and maximum, whereas Z-score transformation expresses each observation in terms of its deviation from the variable mean relative to its standard deviation. Although both transformations can be useful in machine-learning applications, applying them sequentially is not automatically necessary and should be justified

by the distributional properties of the variables and the intended model specification. Accordingly, the proof-of-concept procedure treated scaling as an explicit modeling decision, and a definitive empirical replication should compare alternative preprocessing specifications to evaluate whether the resulting map and cluster assignments remain stable. Such sensitivity analysis is especially relevant for financial variables characterized by skewness, heavy tails, high dispersion, and episodic extreme values. The complete preprocessing sequence, including the treatment of missing observations, extreme values, scaling, transformation, and temporal alignment, should therefore be retained as code or reproducible pseudocode in any subsequent empirical replication.

The SOM was implemented in Python using the MiniSom library. The architecture consisted of a two-dimensional 6 × 6 lattice containing 36 neurons, thereby enabling the multidimensional input vectors to be represented in an interpretable two-dimensional topological space. A Gaussian neighborhood function was used so that adaptation during learning affected not only the best-matching neuron but also neighboring neurons, with the degree of influence declining with distance from the Best Matching Unit. An initial learning rate of 0.50 and 1,000 training iterations were used in the proof-of-concept configuration. These parameters should not be interpreted as universally optimal settings for Iranian capital-market data; rather, they constituted an initial architecture for examining whether recognizable behavioral structures could emerge. In a full empirical application, alternative map sizes, learning rates, neighborhood radii, iteration counts, initialization procedures, and decay functions should be compared through sensitivity analysis. Repeated training using different random seeds would also be required to determine whether the substantive cluster structure remains stable across alternative initial conditions. For complete reproducibility, the Python version, MiniSom version, random seed, weight initialization method, neighborhood-radius schedule, learning-rate decay function, preprocessing code, and exact ordering of model operations should be documented in the final implementation.

The primary analytical technique was the Self-Organizing Map developed within the framework of unsupervised competitive learning. SOM was selected because it is capable of simultaneously performing nonlinear dimensionality reduction, vector quantization, and topology-preserving representation of complex multidimensional observations. For each standardized input vector, the network calculated the distance between the observation and the weight vector associated with each neuron. The neuron with the smallest distance was identified as the Best Matching Unit (BMU). Following BMU identification, the weight vector of the BMU and those of its neighboring neurons were updated toward the input vector according to the learning rate and neighborhood function. In general terms, the updating process can be expressed as (w_i(t+1)=w_i(t)+\alpha(t)h_{bi}(t)[x(t)-w_i(t)]), where (w_i(t)) denotes the weight vector of neuron (i) at iteration (t), (x(t)) represents the input vector, (\alpha(t)) is the learning rate, and (h_{bi}(t)) represents the neighborhood influence of the BMU on neuron (i). Through repeated exposure to the input observations and gradual reduction of the learning rate and neighborhood influence, neurons with similar weight structures become topologically close to one another, whereas observations with substantially different multidimensional characteristics tend to occupy more distant regions of the map. The resulting spatial organization was interpreted as a reduced representation of latent behavioral regularities within the financial data.

After model training, the Unified Distance Matrix, or U-Matrix, was used as the primary visualization for examining the structural boundaries of the trained SOM. The U-Matrix represents the distances between the weight vectors of neighboring neurons and thereby provides a visual indication of homogeneous regions and potential boundaries between behavioral regimes. Areas characterized by relatively small inter-neuron distances were interpreted as internally similar regions of the map, whereas larger distances were considered potential separators

between groups of observations with different financial and trading characteristics. Cluster interpretation was not based solely on visual inspection. Instead, the topological organization of the map was considered jointly with the distribution of input variables across neurons and formal cluster-validity criteria. Once structurally coherent regions had been identified, the corresponding observations and feature profiles were examined to determine whether the resulting groups could be interpreted economically as distinguishable market regimes, such as relatively stable, high-volatility, high-activity, low-liquidity, or fundamentally differentiated states. Because such labels arise after the unsupervised learning process, they were treated as interpretive descriptors rather than predefined training categories.

Model quality was evaluated through complementary quantitative and visual criteria. Quantization Error (QE) was used to assess the extent to which the trained SOM represented the original input observations. It was calculated as (QE=(1/n)\sum_{j=1}^{n}\lVert x_j-w_{BMU(x_j)}\rVert), where (n) denotes the number of observations, (x_j) denotes observation (j), and (w_{BMU(x_j)}) represents the weight vector of the neuron assigned as its BMU. Lower QE values indicate that the prototype vectors provide a closer representation of the original data, although interpretation is meaningful primarily when alternative SOM configurations are compared using the same dataset and preprocessing specification. Cluster cohesion and separation were additionally evaluated using the Silhouette coefficient, defined for observation (i) as (s(i)=[b(i)-a(i)]/\max{a(i),b(i)}), where (a(i)) is the mean distance between observation (i) and other observations in the same cluster and (b(i)) is the lowest mean distance between observation (i) and observations belonging to another cluster. Higher Silhouette values indicate a more favorable combination of within-cluster cohesion and between-cluster separation. The Davies–Bouldin index was also calculated as (DB=(1/K)\sum_{i=1}^{K}\max_{j\neq i}[(S_i+S_j)/M_{ij}]), where (K) represents the number of clusters, (S_i) and (S_j) denote within-cluster dispersion, and (M_{ij}) represents the separation between clusters (i) and (j). Lower Davies–Bouldin values indicate comparatively more compact and better-separated cluster structures. These indices were interpreted jointly rather than independently so that selection of the preferred structure would not depend on a single favorable metric.

The evaluation procedure also incorporated inspection of U-Matrix patterns, neuron occupancy, changes in quantization error during training, and the stability of the resulting structure across alternative model settings. Where feasible, repeated SOM estimation with different initialization seeds and parameter combinations was considered necessary to distinguish persistent topological organization from patterns arising from stochastic initialization. A cluster configuration was regarded as more credible when it demonstrated relatively low quantization error, comparatively high Silhouette values, relatively low Davies–Bouldin values, interpretable boundaries in the U-Matrix, economically coherent feature profiles, and reasonable stability across repeated estimations. Conventional prediction metrics such as Root Mean Square Error and Mean Absolute Percentage Error were not treated as primary evaluation criteria because the SOM in the present study was employed for unsupervised clustering rather than supervised forecasting. In the absence of an independently specified temporal training-test procedure and a clearly defined forecasting benchmark, claims concerning improved predictive accuracy would not be methodologically justified. Consequently, the scope of analysis was deliberately restricted to dimensionality reduction, structural discovery, clustering, and identification of potentially meaningful behavioral regimes. Any subsequent investigation of predictive performance should be conducted as a separate validation study using chronological training and test samples, explicit forecasting targets, and appropriate benchmark models.

3. Findings and Results

The descriptive findings emphasize the heterogeneity of the financial variables in terms of scale, dispersion, and distributional shape. Returns and volatility exhibited greater variability than the market index, while trading volume and trading value were measured on substantially larger numerical scales. From a methodological perspective, this pattern confirms the necessity of scaling the variables prior to implementing the Self-Organizing Map (SOM), because without this step, variables with larger ranges could exert a disproportionate influence on Euclidean distance calculations.

Outlier control affected only a limited proportion of the observations. However, because of inconsistencies in the number of observations reported across different sections, percentages and raw frequencies are not presented in this article as definitive empirical evidence regarding the capital market. The methodologically reliable finding is that, following data cleaning and scaling, the input matrix contained no missing values and was therefore suitable for training the network.

Based on the reported values, the network training error decreased from 0.482 in the initial stages to 0.067 at the end of the training process, corresponding to an approximate reduction of 86.1%. The decline in error was steeper during the early iterations and became more gradual during the later stages, which is consistent with the learning mechanism of SOM. In the initial phase, the network learns the general structure of the input space, whereas subsequent iterations involve progressively smaller local adjustments.

A reduction in training error alone does not guarantee the quality of the clustering solution; however, when accompanied by diminishing weight changes during the final iterations, it may provide evidence of training stability. In a complete empirical implementation, it is recommended that, in addition to the final error value, the distribution of errors across several independent runs and the topographic error also be reported in order to reduce the likelihood of convergence to an unstable structure.

Table 1. Network Training Indicators

Indicator	Value	Cautious Interpretation
Number of iterations	1,000	The training process continued until a state of relative stability was reached.
Initial error	0.482	Indicates a relatively greater initial distance between the input vectors and network weights.
Final error	0.067	Indicates a substantial reduction in mapping distance.
Percentage reduction in error	86.1%	Suggests effective learning at the proof-of-concept level.
Weight changes at the end of training	Very small	Provides evidence that the network was approaching a stable state.

Analysis of the U-Matrix indicates that the map was not homogeneous with respect to the distances between adjacent neurons and that regions characterized by different levels of density and distinct boundaries emerged. This observation is important for the objective of the study because, if all financial vectors had exhibited approximately similar structures, the map would have been expected to show no clearly discernible separations. The presence of regions with different inter-neuron distances is therefore, at least at the proof-of-concept level, consistent with the existence of heterogeneous behavioral patterns.

Based on the reported structure, four final clusters were extracted. The mean intra-cluster distances for Clusters 1 through 4 were reported as 0.23, 0.31, 0.44, and 0.51, respectively. Thus, Cluster 1 exhibited the highest degree of compactness, whereas Cluster 4 displayed the greatest degree of heterogeneity. This distinction is important because it indicates that not all behavioral regimes exhibit the same degree of internal cohesion. Accordingly, a

decision-support model could potentially adjust the level of confidence or signal strength according to the quality of cluster membership.

Table 2. Structure of the Reported Clusters and Methodological Interpretation

Cluster	Mean Intra-Cluster Distance	Reported Characteristic	Interpretation
1	0.23	Homogeneous and compact	A regime characterized by greater internal similarity
2	0.31	Balanced	Moderate cohesion with relatively distinct boundaries
3	0.44	More heterogeneous	Possible presence of substructures or transitional observations
4	0.51	Less compact and distinct	Greater diversity requiring more detailed economic interpretation

Following completion of the SOM training process and analysis of the U-Matrix, the subsequent stage involved extracting the final clusters and determining the definitive structure of the groups formed within the data. In the present study, final clustering was performed on the basis of the distance patterns obtained from the U-Matrix and the high-density regions identified on the map. The network organized similar observations in neighboring regions, while observations exhibiting greater dissimilarity were positioned farther apart. Accordingly, cluster extraction may provide a representation of the underlying structure of the data.

For this purpose, homogeneous regions of the SOM were identified and grouped according to the topological proximity of the neurons. Examination of the final map indicated that the observations were organized into several distinct clusters, with each cluster demonstrating relatively similar internal characteristics while remaining distinguishable from the others. These findings indicate that the SOM was capable of identifying latent patterns within the data and generating a meaningful structure for subsequent analyses.

Table 3. Key Characteristics of the Extracted Final Clusters

Cluster	Number of Cluster Members	Mean Intra-Cluster Distance	General Cluster Characteristic
Cluster 1	175	0.23	Homogeneous and compact
Cluster 2	152	0.31	Balanced
Cluster 3	108	0.44	More heterogeneous
Cluster 4	85	0.51	Less compact and distinct
Total	520	—	—

As shown in Table 3, Cluster 1 contained the largest number of observations and also exhibited the lowest intra-cluster distance (0.23). This finding indicates that the observations assigned to this cluster demonstrated the greatest degree of similarity and were appropriately organized by the SOM within a compact region of the map. Cluster 2 exhibited a comparatively balanced structure and, with a mean intra-cluster distance of 0.31, reflected a moderate degree of dispersion together with an acceptable level of similarity among its members.

By contrast, Clusters 3 and 4 exhibited higher intra-cluster distances. Cluster 3, with a mean distance of 0.44, demonstrated greater diversity than the first two clusters, and its observations were distributed across a relatively broader region. Cluster 4, with the highest intra-cluster distance (0.51), displayed the greatest degree of heterogeneity and variation in characteristics and may therefore represent a group of observations characterized by distinctive behavior or an atypical structural pattern. Overall, this pattern indicates that the SOM was able to distinguish clusters exhibiting different levels of internal similarity and to provide a meaningful representation of the underlying structure of the data.

4. Discussion and Conclusion

The present study examined the feasibility of using a Self-Organizing Map (SOM) as an unsupervised learning approach for identifying latent behavioral regimes in multidimensional Iranian capital-market data. The findings provide preliminary evidence that the selected financial observations did not constitute a homogeneous data structure and that the SOM was capable of organizing them into distinguishable topological regions. The network training error declined from 0.482 at the beginning of the learning process to 0.067 at the end, representing an approximate reduction of 86.1%. In addition, the final U-Matrix displayed regions with different inter-neuron distances rather than a uniform structure, and four clusters were extracted with different degrees of internal cohesion. The mean intra-cluster distances were 0.23, 0.31, 0.44, and 0.51 for Clusters 1 through 4, respectively. Cluster 1 contained 175 observations and exhibited the highest compactness, Cluster 2 contained 152 observations and showed a comparatively balanced structure, Cluster 3 contained 108 observations and demonstrated greater internal heterogeneity, and Cluster 4 contained 85 observations and displayed the greatest within-cluster dispersion. Taken together, these results suggest that the SOM was able to capture meaningful differences in the multivariate structure of the observations and to represent these differences in a form that is both computationally tractable and visually interpretable.

The substantial decline in training error indicates that the network progressively adapted its prototype vectors to the structure of the input data. This result is consistent with the theoretical logic of SOM learning, in which the Best Matching Unit and its neighboring neurons are iteratively adjusted toward the input vectors, allowing the map to approximate the topology of the original multidimensional space [10]. The steeper decline in error during the earlier iterations, followed by smaller adjustments toward the end of training, is also consistent with the expected two-stage behavior of self-organizing learning. In the initial phase, the network primarily captures broad global organization, while later iterations refine local relationships among neighboring neurons. The reduction in error should not be interpreted as proof of an optimal clustering solution by itself; however, when considered together with the reported decline in weight changes and the emergence of identifiable regions in the U-Matrix, it provides evidence that the network reached a relatively stable representation. This interpretation is consistent with work showing that SOMs are useful not only for projection and visualization but also for subsequent cluster extraction when the learned topology is sufficiently organized [13].

The emergence of nonuniform regions in the U-Matrix is one of the most important findings of the study because it indicates that the input observations were characterized by varying levels of similarity. If the financial observations had been structurally homogeneous, the distances among neighboring neurons would have been expected to be comparatively uniform, with few discernible boundaries. Instead, the map demonstrated regions of relative compactness separated by areas of larger inter-neuron distance, suggesting the presence of distinguishable market states. This result supports the premise that capital-market behavior may be better represented as a collection of heterogeneous regimes rather than as a single stable process. The finding is compatible with evidence from the Iranian capital market showing that behavior varies according to information type, market symmetry, and the way participants process market signals [1]. It is also consistent with research emphasizing that emotional and cognitive dimensions can contribute to differentiated forms of investor behavior and can become particularly salient under crisis conditions [2]. From this perspective, the clusters identified in the present study can be understood as data-driven configurations in which multiple financial characteristics co-occur rather than as arbitrarily imposed categories.

The existence of four final clusters further supports the usefulness of unsupervised learning for examining market heterogeneity. Importantly, the clusters did not exhibit equal degrees of internal cohesion. Cluster 1 had the lowest intra-cluster distance of 0.23, while Cluster 4 had the highest value of 0.51. This pattern suggests that some market regimes are more internally consistent and therefore easier to characterize than others. The relatively compact structure of Cluster 1 may reflect observations sharing a common configuration of return, volatility, trading activity, or other input features, whereas the broader dispersion of Cluster 4 may indicate a more heterogeneous market state containing transitional, atypical, or structurally mixed observations. Such differences are important because clustering quality should not be judged only by the number of groups extracted; it should also consider the degree of cohesion within each group and separation across groups. The Silhouette framework was developed specifically to assess whether observations are appropriately matched to their assigned clusters relative to alternative clusters [14], while the Davies–Bouldin criterion evaluates the trade-off between within-cluster scatter and between-cluster separation [16]. Contemporary research has further emphasized the usefulness of Silhouette analysis for evaluating machine-learning clustering structures and identifying solutions with stronger internal consistency [15]. Accordingly, the differences observed among the four clusters should be interpreted not simply as descriptive variation but as an indication that the behavioral regimes differed in the strength of their internal structure.

The current findings also align with a growing body of research demonstrating the relevance of clustering for capital-market analysis. Das et al. showed that density-based clustering approaches can uncover meaningful structures in stock-market data, particularly when the data contain irregularly shaped groups or heterogeneous densities [4]. Although the present study used a SOM rather than a density-based algorithm, the underlying conclusion is similar: financial observations may contain latent structures that become visible only when methods capable of identifying nonlinear or nonparametric organization are employed. Likewise, clustering has been used to support portfolio construction by grouping securities according to shared trend characteristics, thereby reducing the reliance on conventional classification schemes [6]. The relevance of such approaches is particularly evident in markets where the relationship among returns, volatility, liquidity, and other financial indicators may shift over time. The present identification of four behaviorally distinct regions therefore contributes to the same broader methodological perspective by showing that multidimensional market observations can be organized according to intrinsic similarity rather than predefined labels.

The findings are also consistent with studies that have used clustering as an intermediate stage for improving downstream financial analysis. Vásquez Sáenz et al. demonstrated that transforming raw stock-market data into more structured information through clustering can improve the informational basis for return forecasting [7]. This is relevant to the present study because the extracted SOM regimes may serve as a preliminary representation of market context that could later be incorporated into predictive or decision-support models. However, the present results should not be interpreted as evidence of forecasting superiority. The study did not include an out-of-sample predictive benchmark or a temporally separated training and testing framework. Rather, its contribution lies in showing that a coherent unsupervised structure can be extracted from the selected financial features. This distinction is important because reviews of artificial-intelligence applications in stock-market prediction have emphasized that predictive accuracy depends on the quality of the input variables, the evaluation procedure, model specification, and the separation between training and testing data [8]. The current results therefore establish a basis for future predictive testing but do not by themselves demonstrate forecasting effectiveness.

The differential sizes of the identified clusters also merit attention. Cluster 1 contained 175 observations, followed by 152 in Cluster 2, 108 in Cluster 3, and 85 in Cluster 4. The fact that the largest cluster was also the most compact suggests that one dominant regime may have represented a comparatively stable or recurrent pattern within the analyzed data, while the smaller and more heterogeneous fourth cluster may have captured less frequent or more unusual configurations. In capital markets, such asymmetry is plausible because stable or ordinary trading conditions typically occur more frequently than episodes involving extreme volatility, abrupt shifts in liquidity, or unusual combinations of financial indicators. Prior research on investor behavior similarly suggests that responses to external changes such as taxation or fiscal-policy adjustments can induce shifts in portfolio allocation, risk-taking, and market reactions, potentially contributing to transitions between more typical and less typical behavioral states [3]. Although the present study did not assign definitive economic labels to the four clusters, the unequal cluster sizes are consistent with the possibility that some regimes occur more frequently than others.

The identification of a relatively heterogeneous fourth cluster is also compatible with the concept of transitional market states. A cluster with higher internal dispersion may not necessarily reflect poor modeling; instead, it can indicate that the underlying observations occupy a broader region of the feature space or that multiple subpatterns coexist within the same general regime. In SOM analysis, this possibility is particularly important because the map preserves neighborhood relationships and can reveal gradual transitions rather than forcing all observations into sharply separated categories [10]. Javed et al. demonstrated the value of SOM-based approaches for clustering time-series data where observations may share complex temporal similarities rather than simple static characteristics [11]. In the financial context, this implies that a heterogeneous cluster may contain observations situated between two more stable regimes or reflect episodes in which market characteristics are in transition. Therefore, the higher intra-cluster distance in Cluster 4 should not automatically be treated as a defect; rather, it suggests the need for more detailed investigation of the variables and temporal positions associated with that cluster.

The current results also support the use of SOM as a complementary technique to more conventional clustering approaches. Unlike k-means, which partitions observations primarily according to distance from centroids, SOM preserves topological relationships among prototype vectors and provides a two-dimensional representation that facilitates interpretation. This feature can be particularly useful in management and financial decision-making because analysts are often interested not only in cluster membership but also in understanding the proximity among regimes and the potential transitions between them. Research using deep fuzzy SOM models has shown that self-organizing architectures can be extended to financial trading contexts where uncertainty and gradual membership are central concerns [12]. The present results do not employ fuzzy membership, but the differences in cluster compactness suggest that future extensions could use graded membership or confidence scores to represent uncertainty more explicitly, particularly for observations located near cluster boundaries.

The implications of these findings also extend to portfolio and risk-management research. Clustering has been shown to contribute to portfolio optimization by grouping securities according to common characteristics and reducing redundant exposure to similar sources of risk [5]. In this regard, the four behavioral regimes identified in the current study could potentially be incorporated into an adaptive portfolio-management framework in which asset allocation rules vary according to the regime detected by the SOM. Similarly, the cluster structure could help distinguish periods in which conventional diversification strategies are more or less effective. Although such applications were beyond the scope of the current proof-of-concept study, the present findings provide the structural prerequisite for them: the existence of distinct, interpretable regions in the multidimensional financial feature space. This interpretation is also compatible with prior Iranian research showing that data-mining methods

can be used to classify and predict firm-level profitability in the Tehran Stock Exchange [9]. The current study extends this computational orientation from supervised firm classification toward unsupervised identification of broader behavioral patterns.

Overall, the results suggest that the Iranian capital market, at least within the structure represented by the present proof-of-concept dataset, can be meaningfully analyzed through an unsupervised regime-identification framework. The large reduction in network error, the nonuniform topology of the U-Matrix, the emergence of four clusters, and the systematic differences in intra-cluster compactness collectively indicate that the SOM captured structured heterogeneity rather than arbitrary variation. These findings are consistent with both the theoretical foundations of self-organizing neural networks and the broader literature on financial clustering, portfolio construction, behavioral heterogeneity, and machine-learning applications in capital markets [4, 6, 10, 13]. At the same time, the results reinforce the methodological importance of distinguishing exploratory regime discovery from causal inference or forecasting. The principal contribution of the current analysis lies in demonstrating the feasibility of representing multidimensional financial observations as a set of topologically and behaviorally differentiated regimes, thereby providing a foundation for more rigorous empirical, predictive, and decision-oriented applications.

The principal limitation of this study is its proof-of-concept design and the absence of a fully verified and independently reproducible market-wide dataset. Some of the analytical inputs were provisional, and inconsistencies in the number of observations reported across different stages restricted the extent to which numerical findings could be generalized to the entire Iranian capital market. In addition, the results may be sensitive to feature selection, scaling method, treatment of extreme observations, map dimensions, learning rate, neighborhood radius, training iterations, random initialization, and the specific procedure used to convert the trained SOM into final clusters. The study also did not evaluate topographic error across repeated independent runs, did not conduct out-of-sample validation, and did not test whether the identified clusters remained stable across alternative time periods or changing macroeconomic conditions. Consequently, the extracted regimes should be regarded as exploratory structures rather than definitive classifications of Iranian market behavior.

Future research should replicate the analysis using a complete dataset obtained from official sources and should explicitly document the observation period, inclusion and exclusion criteria, handling of suspended or newly listed firms, preprocessing procedures, and final sample size. Researchers should compare alternative feature sets, normalization procedures, SOM dimensions, initialization strategies, and neighborhood functions and should estimate the model repeatedly using different random seeds to evaluate stability. It would also be useful to compare SOM performance with k-means, hierarchical clustering, Gaussian mixture models, DBSCAN, spectral clustering, and other unsupervised methods. Future studies could additionally examine the temporal persistence of identified regimes, construct transition matrices between market states, incorporate macroeconomic and sentiment variables, and test whether regime membership improves out-of-sample forecasts of return, volatility, liquidity, or downside risk. A particularly valuable extension would be to examine whether the cluster structure remains stable during bullish, bearish, crisis, and recovery periods and whether soft or fuzzy membership models can better represent observations positioned near regime boundaries.

From a practical perspective, the proposed framework may be useful for investment managers, financial analysts, brokerage firms, risk managers, and market regulators as an exploratory decision-support tool. Rather than evaluating return, volatility, volume, liquidity, and fundamental indicators separately, practitioners can use a regime-based representation to assess their joint configuration and identify whether current observations resemble

historically stable, transitional, or heterogeneous market conditions. Cluster compactness can also be incorporated as an indicator of confidence, with observations located in highly cohesive regions treated as stronger signals than observations positioned in diffuse or boundary areas. Regulators and market-monitoring institutions may use similar systems to detect unusual combinations of market characteristics that warrant closer examination, while portfolio managers may use regime information to adapt allocation, diversification, and risk-control rules to changing market environments. Such applications should nevertheless be implemented as complementary analytical tools rather than substitutes for expert judgment, fundamental analysis, and formal risk-management procedures.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

[1] M. M. Ebrahimzadeh, Y. Ahadi Sarkani, and M. Mahmoudi, "Developing a capital market behavior model based on information type and market symmetry using a grounded theory approach," (in English), *Quarterly Journal of Economic Research and Policies,* vol. 33, no. 115, pp. 138-177, 2025.

[2] H. Babaei, N. Rezaei, A. Pakmaram, and R. Abdi, "Identifying behavioral financial components using emotional-cognitive dimensions and its role in the capital market crisis," *Iranian Journal of Accounting, Auditing and Finance,* vol. 9, no. 3, pp. 53-80, 2025, doi: 10.22067/ijaaf.2025.45734.1465.

[3] M. Ahmed, N. Kumar, and H. Maqsood, "Taxation and Investor Behavior: Analyzing Portfolio Adjustments, Risk Preferences, and Capital Market Reactions to Fiscal Policy Changes," *JBMR,* vol. 4, no. 2, pp. 1-26, 2025, doi: 10.64105/jbmr.04.02.374.

[4] T. Das, A. Halder, and G. Saha, "Application of density-based clustering approaches for stock market analysis," *Applied Artificial Intelligence,* vol. 38, no. 1, p. 2321550, 2024, doi: 10.1080/08839514.2024.2321550.

[5] J. Sass and A. K. Thös, "Risk reduction and portfolio optimization using clustering methods," *Econometrics and Statistics,* vol. 32, pp. 1-16, 2024, doi: 10.1016/j.ecosta.2021.11.010.

[6] D. Wu, X. Wang, and S. Wu, "Construction of stock portfolios based on k-means clustering of continuous trend features," *Knowledge-Based Systems,* vol. 252, p. 109358, 2022, doi: 10.1016/j.knosys.2022.109358.

[7] J. Vásquez Sáenz, F. M. Quiroga, and A. F. Bariviera, "Data vs. information: Using clustering techniques to enhance stock returns forecasting," *International Review of Financial Analysis,* vol. 88, p. 102657, 2023, doi: 10.1016/j.irfa.2023.102657.

[8] I. K. Nti, A. F. Adekoya, and B. A. Weyori, "A systematic review of fundamental and technical analysis of stock market predictions," *Artificial Intelligence Review,* vol. 53, pp. 3007-3057, 2020, doi: 10.1007/s10462-019-09754-z.

[9] M. Mehrabanpour and M. Habibzadeh, "Clustering and predicting the profitability of companies listed on the Tehran Stock Exchange using a C5 decision tree approach," *Empirical Studies in Financial Accounting,* vol. 15, no. 59, pp. 135-137, 2018.

[10] T. Kohonen, *Self-Organizing Maps*, 3 ed. Springer, 2001.
[11] A. Javed, D. M. Rizzo, B. S. Lee, and R. Gramling, "Somtimes: Self organizing maps for time series clustering and its application to serious illness conversations," *Data Mining and Knowledge Discovery,* vol. 38, pp. 813-839, 2024, doi: 10.1007/s10618-023-00979-9.
[12] D. Pei, C. Luo, and X. Liu, "Financial trading decisions based on deep fuzzy self-organizing map," *Applied Soft Computing,* vol. 134, p. 109972, 2023, doi: 10.1016/j.asoc.2022.109972.
[13] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Transactions on Neural Networks,* vol. 11, no. 3, pp. 586-600, 2000, doi: 10.1109/72.846731.
[14] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *Journal of Computational and Applied Mathematics,* vol. 20, pp. 53-65, 1987, doi: 10.1016/0377-0427(87)90125-7.
[15] M. Shutaywi and N. N. Kachouie, "Silhouette analysis for performance evaluation in machine learning with applications to clustering," *Entropy,* vol. 23, no. 6, p. 759, 2021, doi: 10.3390/e23060759.
[16] D. L. Davies and D. W. Bouldin, "A cluster separation measure," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. PAMI-1, no. 2, pp. 224-227, 1979, doi: 10.1109/TPAMI.1979.4766909.