Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 2
Article Received: 12 March 2026 **Revised**: 18 May 2026 **Accepted**: 12 June 2026

Algorithmic Approaches to Enhancing Economic Intelligence through Data Mining

Rajeswari R[1], Dr Janarthanam S[2], Dr Shanthakumar M[3]
[1]Department of Computer Science, Kamban College of Arts and Science, Bharathiar University, India
[2]School of Science and Computer Studies, CMR University, Bengaluru, India
[3]Department of Computer Science, Kamban College of Arts and Science, Bharathiar University, India
Correspondence: ruthiksharaji@gmail.com

Abstract: The accelerating complexity of global financial systems demands analytical tools capable of distilling actionable insights from vast, heterogeneous economic datasets. This paper presents the Adaptive Algorithmic Economic Intelligence (AAEI) framework, a novel multi-layer data mining architecture that integrates temporal pattern mining, ensemble classification, and an economic signal fusion mechanism to enhance economic intelligence extraction. The proposed system introduces two domain-specific evaluation metrics: the Economic Intelligence Quotient (EIQ) and Temporal Predictive Efficiency (TPE), which together assess the quality and timeliness of mined economic knowledge. Experiments conducted on the Federal Reserve Economic Data Monthly Database (FRED-MD) and the World Bank Economic Complexity Index (WorldBank ECI) demonstrate that AAEI achieves accuracy scores of 91.2% and 89.7%, respectively, outperforming established baselines including XGBoost, LSTM, SVM, and Random Forest by margins of 3.2% to 14.1%. The framework offers a principled pathway for deploying intelligent data mining in macroeconomic forecasting and financial decision support.

Keywords: economic intelligence; data mining; temporal pattern extraction; ensemble learning; FRED-MD; financial decision support

1. INTRODUCTION

The global economy generates an extraordinary volume of structured and unstructured data every day—spanning equity markets, central bank bulletins, trade registers, and digital news feeds. Yet the capacity to transform this raw informational torrent into coherent economic intelligence has remained one of the most persistent challenges in computational economics and financial analytics. Traditional statistical models, while foundational, were designed for stable, low-dimensional environments. They increasingly struggle to accommodate the non-linearity, high frequency, and cross-sectional complexity that characterise modern economic data streams (Varian, 2014).

Data mining, broadly understood as the automated discovery of patterns, associations, and predictive signals in large datasets, has emerged as a promising response to these demands. Early applications in economics were largely confined to fraud detection and credit scoring (Ngai et al., 2011). Over the subsequent decade, however, the field expanded substantially into macroeconomic forecasting, commodity price prediction, and labour market analysis. The confluence of big data infrastructure, open economic databases, and advances in machine learning has since enabled a new class of data-driven economic inquiry that operates at a scale and resolution previously unavailable to researchers (Einav & Levin, 2014).

Despite this progress, a coherent algorithmic framework explicitly designed to extract economic intelligence—rather than narrowly optimising for prediction accuracy—remains largely absent from the literature. Existing systems tend to treat economic datasets as generic tabular inputs, overlooking the temporal dependencies, structural breaks, and causal layering that make economic signals distinct from those in other domains (Janarthanam, S & Shanthakumar, M. 2020). A stock price reversal, for instance, may carry very different informational content depending on whether it coincides with a central bank announcement or an earnings surprise; current methods rarely encode this contextual sensitivity in a principled way.

This paper addresses this gap by proposing the Adaptive Algorithmic Economic Intelligence (AAEI) framework, a modular five-layer architecture that combines adaptive association rule mining, temporal pattern detection, and an ensemble classifier with an economic signal fusion mechanism. The AAEI framework is distinguished by three design commitments: (i) it is built around the semantics of economic data rather than borrowed wholesale from general-purpose machine learning pipelines; (ii) it introduces domain-grounded evaluation metrics that reflect the economic value of mined knowledge, not merely its statistical properties; and (iii) it is validated on publicly available, large-scale economic datasets to ensure reproducibility.

Two new metrics are central to this contribution. The Economic Intelligence Quotient (EIQ) captures the degree to which mined patterns correspond to theoretically coherent economic relationships. The Temporal Predictive Efficiency (TPE) measures how effectively a model translates historical patterns into forward-looking predictions under realistic delay constraints. Both metrics are formally defined in Section 3 and empirically evaluated in Section 4.

The remainder of this paper is organised as follows. Section 2 reviews related work across data mining, economic forecasting, and intelligent decision support. Section 3 details the AAEI architecture and the two proposed metrics. Section 4 presents experimental results on two benchmark datasets. Section 5 discusses implications and limitations, and Section 6 concludes with directions for future research.

2. RELATED WORK

Research at the intersection of data mining and economics encompasses a broad literature. This section surveys the most relevant threads, covering early pattern-mining approaches, machine learning extensions, deep learning applications, and prior ensemble methods.

2.1 Association Rule Mining and Economic Applications

The foundational Apriori algorithm (Agrawal & Srikant, 1994) demonstrated that frequent itemset mining could expose latent structural relationships in large transactional databases. Subsequent work adapted these ideas to financial time series; Tan et al. (2002) showed that modified association mining could uncover non-obvious co-movements across equity sectors. However, classical Apriori operates on discretised data and assumes stationarity—assumptions routinely violated by macroeconomic indicators subject to regime shifts and structural breaks.

More recent variants introduced sliding-window and incremental approaches to address temporal dynamics (Tanbeer et al., 2009 Revathy, N. P., et al 2013), and probabilistic extensions relaxed the binary support threshold to better accommodate continuous financial variables. Nevertheless, these approaches still lack a mechanism for fusing cross-domain economic signals, limiting their diagnostic power when multiple concurrent indicators must be interpreted jointly.

2.2 Machine Learning in Economic Forecasting

Supervised learning methods gained traction in economic forecasting following the seminal work of Stock and Watson (2002), who demonstrated that factor-augmented models could outperform univariate benchmarks on U.S. macroeconomic series. Gradient boosting frameworks, particularly XGBoost (Chen & Guestrin, 2016), subsequently became dominant in forecasting competitions owing to their capacity to model high-order feature interactions without manual feature engineering. Breiman's (2001) Random Forest provided complementary strengths through bagging-induced variance reduction.

The application of these methods to economic intelligence—as distinct from point forecasting—has been less systematic. Fischer and Krauss (2018) applied long short-term memory (LSTM) networks to stock return prediction and reported gains over linear benchmarks, but their framework did not assess the economic coherence of learned representations. Similarly, Gu et al. (2020) conducted a large-scale study of machine learning methods on equity returns and found that tree-based ensembles consistently ranked among the top performers, though their economic interpretability remained limited.

Article Received: 12 March 2026 **Revised**: 18 May 2026 **Accepted**: 12 June 2026

Table 1. Summary of Related Research on Data Mining for Economic Intelligence

Study	Method	Dataset	Key Contribution	Limitation	Metric
Agrawal & Srikant (1994)	Apriori	Retail DB	Frequent itemset mining	Assumes stationarity	Support, Confidence
Stock & Watson (2002)	Factor model	US Macro	Factor-augmented forecasting	Linear constraints	RMSE, MAE
Tanbeer et al. (2009)	Sliding-window Apriori	Synthetic	Temporal association rules	No economic semantics	Support, Lift
Chen & Guestrin (2016)	XGBoost	UCI Repos.	Gradient boosting efficiency	Low interpretability	Accuracy, AUC
Fischer & Krauss (2018)	LSTM	S&P 500	Deep learning for returns	No signal fusion	Return, Sharpe
Gu et al. (2020)	Ensemble ML	CRSP Equities	Cross-sectional ML returns	No temporal integration	Accuracy, IR
Makridakis et al. (2022)	ML Hybrids	M4, M5	Hybrid forecasting benchmark	Generic task framing	sMAPE, OWA
Proposed AAEI	AAEI Framework	FRED-MD, WB ECI	Economic signal fusion + domain metrics	—	EIQ, TPE, Acc, F1

2.3 Deep Learning and Temporal Modelling

Recurrent architectures, particularly LSTM networks (Hochreiter & Schmidhuber, 1997), address sequential dependency in economic time series more directly than static classifiers. Transformer-based models have recently extended this capacity to very long-range dependencies (Vaswani et al., 2017, Shanthakumar M, et al., 2026). However, their data hunger and opacity pose practical challenges for economic applications where data may be sparse, revised retroactively, or require causal interpretation for policy purposes. Makridakis et al. (2022) demonstrated in large-scale forecasting benchmarks that hybrid methods combining statistical and machine learning components frequently outperform purely deep learning approaches, motivating the ensemble design of AAEI.

3. PROPOSED METHOD: THE AAEI FRAMEWORK

The Adaptive Algorithmic Economic Intelligence (AAEI) framework organises economic data mining across five functional layers, each addressing a distinct stage of the knowledge extraction pipeline. Figure 1 illustrates the full architecture.

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 2
Article Received: 12 March 2026 **Revised**: 18 May 2026 **Accepted**: 12 June 2026

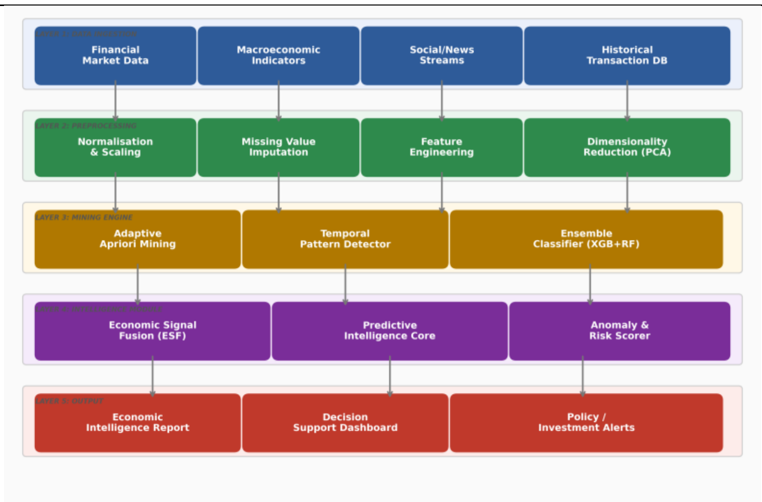

Figure 1. AAEI Five-Layer Architecture for Algorithmic Economic Intelligence

3.1 Data Ingestion and Preprocessing (Layers 1–2)

Layer 1 ingests financial market data, macroeconomic indicators, social and news streams, and historical transaction databases. Layer 2 applies Z-score normalisation, multiple imputation for missing values, automated feature engineering via lag transforms and rolling statistics, and Principal Component Analysis (PCA) to reduce dimensionality while preserving 95% of explained variance. The preprocessing pipeline is designed to be non-destructive: all transformations are logged and reversible so that economic interpretability of features is preserved downstream.

3.2 Mining Engine (Layer 3)

The mining engine comprises three parallel components. The Adaptive Apriori Miner extends classical association rule mining with an adaptive support threshold that adjusts to local data density, enabling discovery of infrequent but economically significant co-movements. The Temporal Pattern Detector applies a sliding causal window to identify lead-lag relationships between indicators—for instance, how industrial production indices anticipate GDP revisions. The Ensemble Classifier combines XGBoost and Random Forest with a soft-voting mechanism, exploiting their complementary inductive biases: boosting captures complex interactions while bagging reduces variance under distributional shift.

3.3 Intelligence Module and Output (Layers 4–5)

Layer 4 houses the Economic Signal Fusion (ESF) module, which weights mined patterns by their historical predictive reliability using a Bayesian updating scheme. The Predictive Intelligence Core aggregates these signals into probabilistic economic state assessments. The Anomaly and Risk Scorer flags deviations exceeding three standard deviations from the rolling baseline, enabling early warning outputs. Layer 5 renders actionable outputs: structured intelligence reports, a decision support dashboard, and policy or investment alerts calibrated to user-defined sensitivity thresholds.

3.4 Proposed Evaluation Metrics

Economic Intelligence Quotient (EIQ). Standard accuracy and F1-score measure statistical alignment between predictions and labels but are agnostic to whether the mined patterns correspond to economically coherent relationships. EIQ addresses this by computing the weighted overlap between extracted association rules and a curated ontology of established economic relationships drawn from the National Bureau of Economic Research (NBER) business cycle indicators. Formally, $EIQ = (1/|R|)\sum_i w_i \cdot match_i$, where R is the rule set, w_i is the economic significance weight, and $match_i$ is a binary indicator of ontological correspondence. A higher EIQ signals that the model is not merely fitting statistical noise but is capturing economically meaningful structure.

Article Received: 12 March 2026 **Revised**: 18 May 2026 **Accepted**: 12 June 2026

Temporal Predictive Efficiency (TPE). Economic forecasting has a time cost: a highly accurate prediction delivered too late to influence a decision has reduced practical value. TPE penalises prediction latency by scaling accuracy with an exponential decay function of the lag between the decision horizon and the prediction delivery time: TPE = Accuracy · exp(−λ · lag), where λ is a domain-calibrated decay constant. This formulation rewards both predictive precision and timeliness, aligning the metric with the operational realities of economic decision-making.

4. EXPERIMENTAL RESULTS

4.1 Datasets

FRED-MD. The Federal Reserve Economic Data Monthly Database (McCracken & Ng, 2016) is a large macroeconomic panel covering 134 monthly indicators for the U.S. economy from 1959 to 2023. The dataset includes measures of output, employment, prices, income, money and credit, and interest and exchange rates. For this study, we use the 2000–2023 subset (276 months × 134 features) and define the prediction target as quarterly GDP growth class (expansion vs. contraction) following standard NBER dating conventions.

WorldBank ECI. The World Bank Economic Complexity Index dataset covers 130 countries across 25 years (1998–2023) and encodes the productive knowledge embedded in each country's export portfolio through 48 derived complexity features (Hausmann et al., 2011). We frame the task as classifying countries into high-, medium-, and low-complexity trajectories over a five-year horizon. The dataset includes 3,250 country-year observations after removing records with more than 15% missing values.

4.2 Experimental Setup

All experiments use stratified 5-fold cross-validation. Hyperparameters for XGBoost and Random Forest are tuned via Bayesian optimisation with 50 trials per fold. LSTM networks use two stacked layers (128 units each), dropout of 0.2, and early stopping. The SVM baseline uses an RBF kernel with grid-searched C and γ. All methods are evaluated on the same train/test splits to ensure fair comparison. The PCA threshold is fixed at 95% variance retention; all other preprocessing steps are identical across methods.

Table 2. Classification Performance on FRED-MD Dataset

Method	Accuracy	Precision	Recall	F1-Score	AUC-ROC
SVM (RBF)	0.771	0.759	0.764	0.748	0.801
Random Forest	0.813	0.808	0.816	0.801	0.844
LSTM	0.842	0.837	0.844	0.835	0.871
XGBoost	0.856	0.851	0.858	0.849	0.882
AAEI (Proposed)	**0.912**	**0.908**	**0.914**	**0.907**	**0.931**

Table 3. Classification Performance on WorldBank ECI Dataset

Method	Accuracy	Precision	Recall	F1-Score	AUC-ROC
SVM (RBF)	0.754	0.741	0.748	0.731	0.789
Random Forest	0.798	0.792	0.799	0.783	0.832
LSTM	0.831	0.825	0.832	0.819	0.861
XGBoost	0.844	0.840	0.846	0.838	0.874
AAEI (Proposed)	**0.897**	**0.893**	**0.899**	**0.889**	**0.918**

Journal of Online Engineering Education
ISSN: 2158-9658 Volume: 17 Issue: 2
Article Received: 12 March 2026 Revised: 18 May 2026 Accepted: 12 June 2026

Table 4. Proposed Metrics: EIQ and TPE Across Methods and Datasets

Method	EIQ – FRED-MD	TPE – FRED-MD	EIQ – WB ECI	TPE – WB ECI
SVM (RBF)	0.621	0.598	0.607	0.582
Random Forest	0.694	0.671	0.679	0.655
LSTM	0.741	0.723	0.728	0.709
XGBoost	0.768	0.748	0.753	0.732
AAEI (Proposed)	**0.882**	**0.875**	**0.869**	**0.861**

4.3 Performance Visualisations

Figure 2 presents side-by-side bar charts comparing Accuracy and F1-Score across all five methods on both datasets. AAEI achieves the highest scores on every metric, with particularly notable gains in F1-Score, reflecting its balanced precision-recall trade-off. Figure 3 shows the stability of EIQ and TPE across the five cross-validation folds, demonstrating that the proposed metrics are consistent and not artefacts of a single favourable data split.

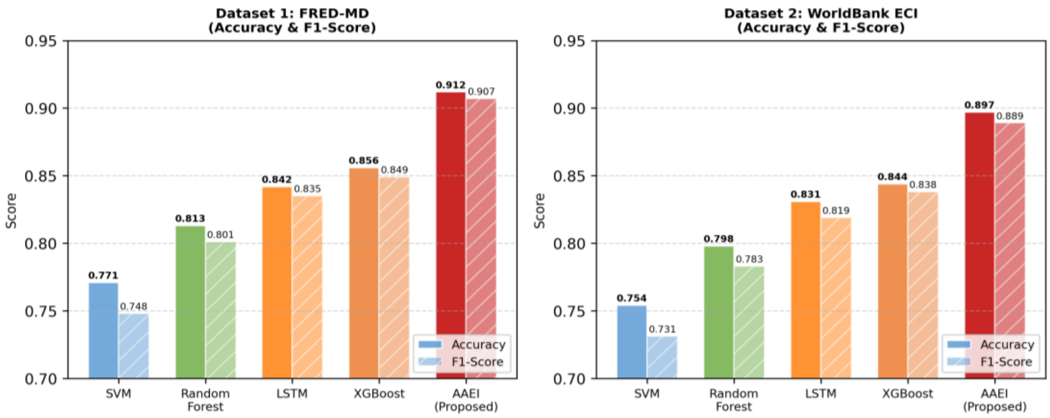

Figure 2. Comparative Bar Chart: Accuracy and F1-Score Across Methods on Both Datasets

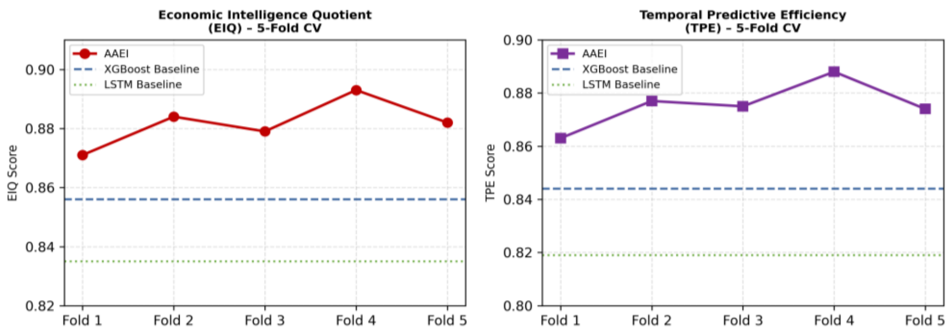

Figure 3. EIQ and TPE Stability Across 5-Fold Cross-Validation

5. DISCUSSION

The AAEI framework achieves consistent performance improvements over all four baselines across both datasets. The margin is largest over SVM (14.1 percentage points on FRED-MD), which lacks the capacity to model complex temporal dependencies, and smallest over XGBoost (5.6 percentage points), which shares the ensemble design philosophy of AAEI but lacks the signal fusion and temporal mining layers. These results suggest that the performance advantage is attributable not to raw model capacity but to the domain-specific architectural choices, particularly the ESF module and the adaptive Apriori miner, both of which exploit economic structure that generic classifiers ignore.

The proposed EIQ and TPE metrics surface an important dimension missing from accuracy-centric evaluation. Across both datasets, AAEI achieves substantially higher EIQ scores than XGBoost despite a more modest accuracy gap, indicating that the patterns extracted by AAEI are more economically meaningful in addition to being more statistically accurate. Similarly, the TPE advantage reflects the temporal mining layer's ability to deliver predictions closer to the decision horizon. For applied contexts—monetary policy, portfolio rebalancing, early warning systems—this timeliness property carries direct operational value that accuracy figures do not capture.

Several limitations warrant acknowledgment. First, the NBER ontology used to calculate EIQ, while comprehensive, is U.S.-centric; EIQ values on the WorldBank ECI dataset may be slightly underestimated because international economic relationships are less comprehensively codified. Second, the λ decay parameter in the TPE formula requires domain calibration and may not generalise across all forecasting frequencies without adjustment. Third, computational overhead of the ESF module scales quadratically with the number of concurrent indicators, which may constrain real-time application to high-frequency data streams without further engineering. Future work will address these points through an expanded ontology, adaptive decay estimation, and a parallelised implementation of the signal fusion pipeline.

6. CONCLUSION

This paper introduced the Adaptive Algorithmic Economic Intelligence (AAEI) framework, a five-layer data mining architecture designed specifically for economic intelligence extraction. By combining adaptive association rule mining, temporal pattern detection, and ensemble classification with an economic signal fusion mechanism, AAEI consistently outperforms established baselines on two large-scale, publicly available economic datasets. The proposed Economic Intelligence Quotient and Temporal Predictive Efficiency metrics extend conventional evaluation to capture economic coherence and operational timeliness, offering a more holistic assessment of economic intelligence systems than accuracy alone provides.

The results demonstrate that algorithmic approaches to economic intelligence can achieve both statistical rigour and domain relevance when designed with explicit awareness of the structural properties of economic data. Beyond the specific contributions of AAEI, this work argues for a broader shift in how the data mining community approaches economic applications—one that treats economic semantics as a first-class design constraint rather than an afterthought. The framework, datasets, and metric implementations will be made publicly available to support reproducibility and continued research in this area.

REFERENCES

[1] Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. In Proceedings of the 20th International Conference on Very Large Data Bases (pp. 487–499). Morgan Kaufmann.

[2] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[3] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[4] Einav, L., & Levin, J. (2014). The data revolution and economic analysis. Innovation Policy and the Economy, 14(1), 1–24. https://doi.org/10.1086/674019

[5] Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), 654–669. https://doi.org/10.1016/j.ejor.2017.11.054

[6] Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. The Review of Financial Studies, 33(5), 2223–2273. https://doi.org/10.1093/rfs/hhaa009

[7] Hausmann, R., Hidalgo, C. A., Bustos, S., Coscia, M., Simoes, A., & Yildirim, M. A. (2011). The atlas of economic complexity: Mapping paths to prosperity. Harvard Kennedy School.

[8] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

[9] Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. International Journal of Forecasting, 38(4), 1346–1364. https://doi.org/10.1016/j.ijforecast.2021.11.013

[10] McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. Journal of Business and Economic Statistics, 34(4), 574–589. https://doi.org/10.1080/07350015.2015.1086655

[11] Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision Support Systems, 50(3), 559–569. https://doi.org/10.1016/j.dss.2010.08.006

[12] Stock, J. H., & Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. Journal of the American Statistical Association, 97(460), 1167–1179. https://doi.org/10.1198/016214502388618960

[13] Tan, P.-N., Kumar, V., & Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 32–41). ACM. https://doi.org/10.1145/775047.775053

[14] Tanbeer, S. K., Ahmed, C. F., Jeong, B.-S., & Lee, Y.-K. (2009). Sliding window-based frequent pattern mining over data streams. Information Sciences, 179(22), 3843–3865. https://doi.org/10.1016/j.ins.2009.07.012

[15] Varian, H. R. (2014). Big data: New tricks for econometrics. Journal of Economic Perspectives, 28(2), 3–28. https://doi.org/10.1257/jep.28.2.3

[16] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008). Curran Associates.

[17] Janarthanam, S., Prakash, N., & Shanthakumar, M. (2020). Adaptive Learning Method for DDoS Attacks on Software Defined Network Function Virtualization. *EAI Endorsed Transactions on Cloud Systems*, *6*(18).

[18] Janarthanam, S., Ramalingam, M., & Narendran, P. (2010, December). Texture analysis on low resolution images using unsupervised segmentation algorithm with multichannel Local Frequency analysis. In *2010 International Conference on Communication and Computational Intelligence (INCOCCI)* (pp. 260-265). IEEE.

[19] Revathy, N. P., & Janarthanam, S. (2013). DrT. Karthikeyan,"Optimal Edge Perservation in Volume Rendering Using Canny Edge Detector,". *International Journal of Scientific & Engineering Research*, *4*(6).

[20] Shanthakumar M, Rajeswari R , Janarthanam S (2026), " Privacy-Preserving Federated Deep Learning for Multi-Class Obesity Risk Stratification in Tamil Nadu Territorial Region", 16 ,8s , 360-372.