UDC 004.896
DOI: 10.15587/1729-4061.2025.341468

DEVELOPMENT OF DEEP LEARNING FRAMEWORK FOR COMPLEX PATTERN RECOGNITION IN BIG DATA

Gulzhan Muratova
Candidate of Physical and Mathematical Sciences
Department of Information Technology
S.Seifullin Kazakh Agrotechnical Research University
Zhenis str., 62, Astana, Republic of Kazakhstan, 010011

Ainur Jumagaliyeva
Corresponding author
Senior Lecturer*
E-mail: jumagalievaainur.m@gmail.com

Elmira Abdykerimova
Candidate of Pedagogical Sciences, Professor
Department of Computer Science**

Asset Turkmenbayev
Candidate of Pedagogical Sciences, Professor
Department of Fundamental Sciences**

Bulat Serimbetov
Candidate of Technical Sciences, Associate Professor*

Zauresh Yersultanova
Assistant Professor
Department of Physics, Mathematics and Digital Technologies
U. Sultangazin Pedagogical Institute
Tauelsizdik str., 118, Kostanay, Republic of Kazakhstan, 110000

Gulzada Omarkulova
Senior Lecturer*
*Department of Information Technology
K.Kulazhanov Kazakh University of Technology and Business
Kaiym Mukhamedkhanov str., 37A, Astana,
Republic of Kazakhstan, 010000
**Caspian State University of Technology
and Engineering named after Sh.Yessenov
14th Microdistrict str., 50, Aktau, Republic of Kazakhstan, 130003

The object of this study is the process and analysis of intelligent recognition and classification of spatiotemporal patterns in large arrays of streaming data. The problem to be solved is the absence of a deep learning framework that can guarantee adaptability to rapidly changing concepts, efficient computation for continuous data streams, and the transparency of the prediction process when working with heterogeneous and dynamically changing sources of big data used to support decision making.

The developed programing framework applies convolutional neural network- long short – term memory networks with an attention-gating mechanism that enables detection of spatiotemporal dependencies and exhibits model interpretation of decisions. Extensive evaluation of the implemented system using multivariate flow-based data demonstrated the performance capabilities of the system with a classification accuracy of 0.98, F1 score of 0.97, area under the receiver operating characteristic curve of 0.99 and Harmonic Score of 0.90. The interpretation of the results is summarized by the interaction of multilevel feature extraction followed by an optimization process through Kullback-Leibler divergence that ensures reliable online drift detection and automatic models re-training. Additional contributions of the systematic use of the framework included interpretable decisions using Shapley Additive explanations and gradient-weighted class activation mapping visualizations. It has a strong evidence of sustained reliable model performance in non-stationarity data conditions and streaming data. The outcome of this study embodies significant practical implications toward the creation of real-time decision-support systems in domains. Finally, the structured framework can also be utilized for future investigations into the development of highly scalable, explainable, and trust-worthy artificial intelligence architectures

Keywords: spatiotemporal modeling, hybrid architecture, drift concept, statistical divergence, incremental retraining

Received 02.10.2025
Received in revised form 08.12.2025
Accepted date 16.12.2025
Published date 30.12.2025

How to Cite: *Muratova, G., Jumagaliyeva, A., Rystygulova, V., Abdykerimova, E., Turkmenbayev, A., Serimbetov, B., Yersultanova, Z., Omarkulova, G. (2025). Development of deep learning framework for complex pattern recognition in big data. Eastern-European Journal of Enterprise Technologies, 6 (9 (138)), 54–66. https://doi.org/10.15587/1729-4061.2025.341468*

1. Introduction

The explosion of data generated by the ongoing digital revolution and the fourth industrial revolution is unprecedented, which has become a primary catalyst for scientific and technological advancement. According to the IDC, the global datasphere will surpass 175 zettabytes by 2025, with more than 30 percent of this volume generated by IoT devices, cloud computing systems, and real-time data generated by industrial sensors [1]. This massive amount of data, as well as the inherent heterogeneity velocity, highlights the necessity of intelligent systems that can extract valuable insights quickly as well as provide support for automated decision making within strict time and resource constraints.

An essential component of intelligent systems is patterns recognition, which enables these systems to discern relationships, detect errors, and reveal hidden constructs within a large noisy dataset. The ability to use this technology is not limited to any specific industry including, but not limited to, health care, industrial automation, cyber security, financial services, and smart transportation. However, with the ever-growing size and complexity of data, traditional methods of pattern recognition cannot keep up with the changing environment. Thus, when constructing models for pattern recognition, using traditional approaches (versus modern/deep learning) is usually acceptable for small amounts of curated datasets; however, when it comes to more extensive or dynamic datasets, traditional methods become ineffective due to scalability, adaptability, and non-stationarity.

Recent advancements in deep learning techniques have fueled major advancements in the area of pattern recognition. Specifically, the ability for the deep learning model to learn multiple hierarchical levels of abstraction on the input data is one of the primary strengths of deep learning methods used today when applied to pattern recognition problems. For example, convolutional and recurrent neural networks have demonstrated excellent performance in numerous types of data including but not limited to visual data, sequential data, multimodal data, and even graph or streaming data. The advent of transformer networks has opened the door for using deep learning systems for relational and long-range temporal task markets. These systems have seen tremendous success in areas such as recommendation systems, natural language processing, and anomaly detection.

Although deep learning techniques are rapidly evolving, many challenges still exist when using deep learning methods to analyze large amounts of real-time data. High computational requirements, poor model interpretability, and long latency periods are obstacles that limit the effectiveness of many of today's intelligent solutions. As such, further research is needed to build adaptive, explainable deep learning models that are successful under dynamic conditions. Working toward these solution types remains an exciting area for solid ongoing research in the area of intelligent pattern recognition.

2. Literature review and problem statement

The paper [2] presents the results of research on in-depth study of the deep learning methods for big data, image processing, and signal analytics. It is shown that the level of complexity of data and data velocity, did not allow for a structure to be pre-defined or structured. But there were unresolved issues related to the models, that did not contribute much towards resolving the scalability problems faced in distributive environments. The reason for this may be an attributable towards computationally heavy processing and data streams following different formats, thus limiting real-time usability. All this suggests that it is advisable to conduct a study on exploring adaptable or hybrid structures. All of which indicates a need for the scope of adaptability to provide recognition tasks with uncertainty.

The paper [3] presents the results of research on the applications of adapted deep learning delivered in streamed processes. It is shown that online tuning and active management of resource enabled better performance in situational change. However, these models did not have established longer-term robustness with respect to concept drift. But there were unresolved issues related to a lack predictive for drift detection, and long-standing retraining paradigm. A way to overcome these difficulties can be a possibility to embed statistical detection and incremental learning of schemas. All this suggests that it is advisable to conduct a study on limited spans of variances for adaptive mechanisms that learn.

The paper [4] presents the results of research on the applicable methods of learning with depth in minimal data contexts. It is shown that transfer learning and augmentation with the limited training examples were impact factors loss of training. But there were unresolved issues related to quality concerns. Because, there was not an inherent extraction mechanism for explainability, such as limitations of speed. A way to overcome these difficulties can be lightweight processes. All this suggests that it is advisable to conduct a study on more transparency and explainability.

The paper [5] presents the results of research on canters mainly on lightweight optimization and more simplified processes for explainability. It is shown that compact architectures could address computational costs, although speed recovery projects both some observation. But there were unresolved issues related to speed consumption. Because it tended to increase. A way to overcome these difficulties can be a realignment back to multi-layered and structured methods of interpretation. All this suggests that it is advisable to conduct a study on interpretable deep learning models and other new approaches to address the different types of variances, that are present in dynamic data.

The paper [6] presents the results of research on similar hybrid convolutional neural network-long short-term memory approach. It is shown that the models were able to extract multi-dimensional features through the integration of the spatial and temporal dimensions. But there were unresolved issues related to the long-term accuracy of a hybrid model. Because static processes of training, did not account for concept drift. A way to overcome these difficulties can be view on handling drift. It would require a strategy that involves dynamically updating the training process by continuously addressing the real-time accuracy of learning and updates. All this suggests that it is advisable to conduct a study on why adaptive hybrid models remain important.

The paper [7] presents the results of research on training protocols intended for environments in a non-stationary state. It is shown that a window approach to the updating of training could proactively mitigate any performance losses in training with the occurrence of drift. But there were unresolved issues related to cost from a retraining standpoint and it is likely that more accuracy is yielded by being overly sensitive to the most recent data. The reason for this may be a lack of selective rules used for weak updates and validation restrictions. All this suggests that it is advisable to conduct a study on event-driven retraining adaptation. Because an event-driven retraining adaptation in the drift-impact metrics could assist with stability.

The paper [8] presents the results of research on optimization algorithms into deep learning workflows. It is shown that a hybrid optimization is able to decrease convergence time while increasing accuracy. But there were unresolved issues related to the model. The reason for this may be a lack of robustness and generalization. It also because of highly stochastic environments that did not account for the impact of insufficient self-adaptation. A way to overcome these difficulties can be implementing meta-heuristic approaches as alternatives. All this suggests that it is advisable to conduct a study on methods are specifically designed to handle non-stationary environments.

The paper [9] presents the results of research on attention mechanisms and fusion methods. It is shown that attention mechanisms increase the stability of model output by focusing the learning on more salient features. But there were unresolved issues related to the trade-offs associated with attention and fusion are related to computational latent and resource usage. A way to overcome these difficulties can be attempting to utilize or lightweight attention structures. All this suggests that it is advisable to conduct a study on interpretable systems that get their learning mechanisms from attention parameters.

The paper [10] presents the results of research on explainable AI applied to financial big data systems. It is shown that there are some needs to develop trust and facilitate a more informed approach to problem-solving. The study outlined that there are various existing approaches to creating more interpretable systems for decision-making. But there were unresolved issues related to a high computational cost associated with many of these approaches. That is why many systems utilize trade-offs by compromising the quality of accuracy to achieve lower costs. A way to overcome these difficulties can be utilizing lightweight models may address some of these trade-offs, even though scalability issues remain to be addressed. All this suggests that it is advisable to conduct a study on the data demands of real-world applications, that achieve a low overhead economically as compared to accuracy.

These studies verify that deep learning has developed into the chief technology for intelligent pattern recognition, but the pragmatic implementation of deep learning within the context of big data is still hampered by three challenges:

1) lack of adaptability to streaming and drifting data;

2) insufficient incorporation of explainable AI principles to support transparent decision-making;

3) real-time computational inefficiency.

It is challenging, both scientifically and pragmatically, to continue the work of investigating the field of developing deep learning frameworks to safeguard scalability, robustness, and interpretability for intelligent pattern recognition in data-centered systems operating at large scale.

3. The aim and objectives of the study

The aim of the study is to create an explainable deep-learning framework for the identification of complex, nonlinear, and dynamical patterns in large scale data streams while maintaining interpretability and computational efficiency.

To achieve this aim, the following objectives were accomplished:

– to generate a unified hybrid architecture for convolutional (spatial) a recurrent (temporal) learning with attention;

– to implement drifts-aware adaptation that is based on entropy and Kullback-Leibler divergence, for online detection of distributional shifts and partial retraining;

– to provide explainable AI tools (Shapley additive explanations, gradient-weighted class activation mapping) for transparent decision making and feature attribution;

– to implement an organized modular-end-to-end workflow for data acquisition, preprocessing, training, and evaluation on a continuous basis across diverse, multi-dimensional domains.

4. Materials and methods

4. 1. The object and hypothesis of the study

The object of this study is the process and analysis of intelligent recognition and classification of spatiotemporal patterns in large arrays of streaming data. This process involves the continuous collection of data from rapidly changing pattern structures generated by modern information technology systems using various methods, requiring decision-making based on the changing distribution of patterns over time. It is necessary to provide a means for improving the stability, adaptability, and transparency of intelligent recognition and classification of patterns in real-time environments.

The main hypothesis of the study is that integrating the learning of both spatial and temporal feature information together with an adaptively drift-aware mechanism and explainable artificial intelligence will improve the robustness and interpretability of pattern recognition models applied to dynamic and heterogeneous, massive scale, Big Data stream applications.

In conducting this study, certain assumptions were made regarding the ability to acquire representative historical and streaming data; that concept drift occurs over time as opposed to an instantaneous event; that sufficient amounts of labeled data segments had access to enable the incremental updating of models; sufficient computational resources are available to allow for training models on a real-time basis and providing inference.

Simplifications adopted in the study are the use of aggregate flow-based features for our data versus using raw packet-level data, and predefined thresholds for drift detection and the application of a fixed metric for evaluating model performance. These assumptions provided computational feasibility for performing experiments on this topic and allow for reproducibility of results of this study while not restricting the generalizability of the methodology developed for this study.

4. 2. The dataset and preprocessing

This research leveraged flow-based multivariate dataset having spatiotemporal and statistical features of large-scale activities of networks and systems. Each record is determined by its temporal window and consists of the aggregated communication statistics, entropy-based metrics, and protocol level indicators. Missing values were replaced with feature-wise means while identified outliers were clipped at the 95^{th} percentile in order to minimize extreme deviations. Categorical variables were transformed with one-hot encoding. Numerical features were standardized with z-score normalization by retaining the value as the value minus the mean divided by the standard deviation calculated on the training split.

Temporal partitioning of the dataset generated three separate parts for training (70%), validation (15%); and testing (15%) which was not vulnerable to temporal leakage. In addressing class imbalances, it is possible to use SMOTE to provide synthetic minority classes that can better generalize the classifier.

4. 3. Feature description

Descriptions, semantics, and categories of the features can be found in Table 1.

Table 1 contains the feature set used to train and evaluate the model proposed in this study. Each feature is represented with roughly the same level of granularity and can contribute unique statistical or semantic information to help fully capture network behavior for effective model training (Table 1). The feature summaries were also used as definitions for the model inputs and model performance during the evaluation process.

Table 1

The features in the dataset interpreted the semantics and meaning with the features

Feature No.	Feature name	Description / semantic meaning	Feature type
1	Avg packet Size	Displays the average size of packets in a flow, which aids understanding of how heavy or light segments of the transmission traffic [11]	Numerical
2	Flow time	Indicates how long a flow is assertive, or in other words, the amount of time the data stream can be thought of as active, and reflects how packed the activity is or density [12]	Numerical
3	Packet rate	Counts the packets sent for every second time unit, which is useful for identifying communication burst or activity spikes	Numerical
4	Byte rate	Represents how many bytes pass through the system per second. It expresses a dimension of the overall intensity of data or flow load weight [13]	Numerical
5	Flow variation	Measures to what extent packet sizes vary within one window, or producing somewhat of a volume measurement as opposed to number measurement. A high variance would indicate variable or erratic network behavior [14]	Numerical
6	Entropy signature	Displays how patternless the byte values are in the traffic. In essence, the higher the entropy, the more chaotic or diversified the traffic data stream [15]	Numerical
7	Temporal correlation	Measures how similar the traffic pattern is in neighboring time windows. This metric is helpful when looking for similarities in behavior or distributed traffic stream [16]	Numerical
8	Anomaly score	A numerical measurement which tells how far away from normal activity the current behavior is. Value is calculated based on rolling statistics which highlight anomalous behavior	Numerical
9	Protocol encoded	Presents the network protocol in an encoded form (for example, TCP is mapped to 0 and UDP is mapped to 1) which allows protocol functionality in ML models for encoded protocol into data [17]	Categorical
10	Connection type	Is an encoded variable reflecting whether the activity was produced from internal or external traffic is. This flag allows separations of local activity from incoming traffic and helps with the previous feature [18]	Categorical
11	Timestamp index	Will provide a normalized position of the sample in regard to the sliding window. This is useful if the focus is on the temporal order of the observation	Numerical
12	Label target	Is the actual class which is used during training connected to behavior whether or not normal when diagnosing anomaly is [19]	Target

4. 4. Model training and validation process

The entire process of training and validation is illustrated in Fig. 1.

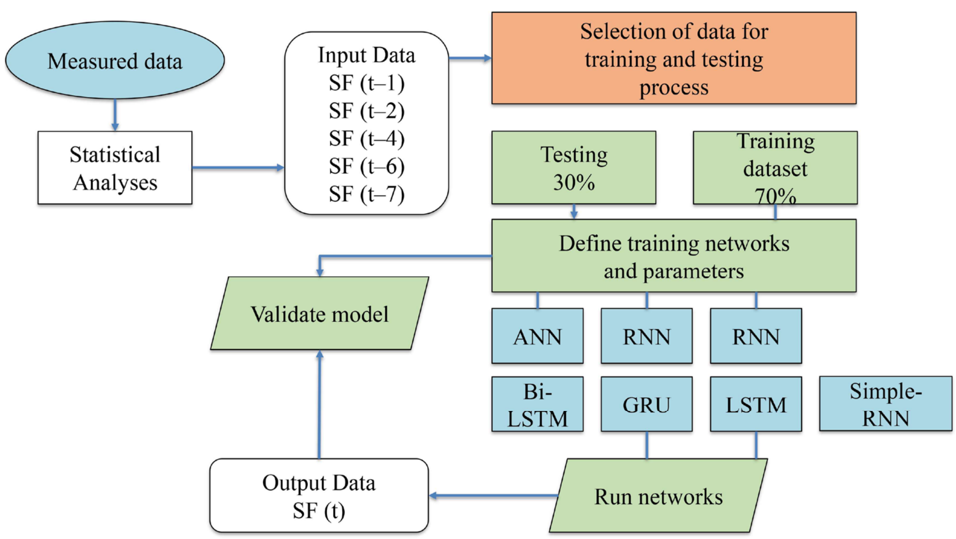

Fig. 1. The execution of the training and validation procedures

The raw data were transformed to temporal sequence data, split into random training and testing sets, and transformed to train different architectures (artificial neural network, recurrent neural network, long short-term memory, gated recurrent unit). All models were trained under similar conditions to ensure valid comparisons (Fig. 1). The process had an automated data and ingestion processes, pre-processing, training models, and validating the model predictions using metrics suitable for classification machine learning problems.

4. 5. Hybrid CNN-LSTM framework

The use of hybrid CNN and LSTM models is to combine by jointly addressing both the spatial characteristics associated with the input data (via the convolutional module) and temporal characteristics that exist between sequential data points (via the LSTM recurrent module).

Convolutional models consist of multiple layers with small kernel sizes. They include rectified linear units for nonlinearity as well as max-pooling for dimensionality reduction while still retaining a meaningful representation of the features learned. After extraction, feature maps are flattened and then passed on to the next stage for modeling temporal relationships.

Temporal relationships are modeled in an LSTM using long short-term memory (LSTM) units which have three components: an input gate, that accepts inputs, a forget gate, that determines what is forgotten, and an output gate through which outputs are produced. The use of this three-gate structure allows the LSTM to be able to learn both short-term and long-term dependencies on sequential data.

Finally, the outputs of the LSTM are modified by an attention mechanism. The attention mechanism assigns varying degrees of importance or relevance to the temporal features in the prediction process. The attention-modified feature representation is passed through fully-connected layers with dropout applied to help mitigate the risks of overfitting. The last layer produces the probability estimates for classification.

4. 6. Adaptive learning and drift detection

The model continuously observes the prediction distribution and retrains the model only when drift is detected, Fig. 2.

Fig. 2 shows that the model distribution changes continuously and the discrepancy between the model and the previous prediction distribution is detected. This means that the model is retrained only when a threshold is exceeded.

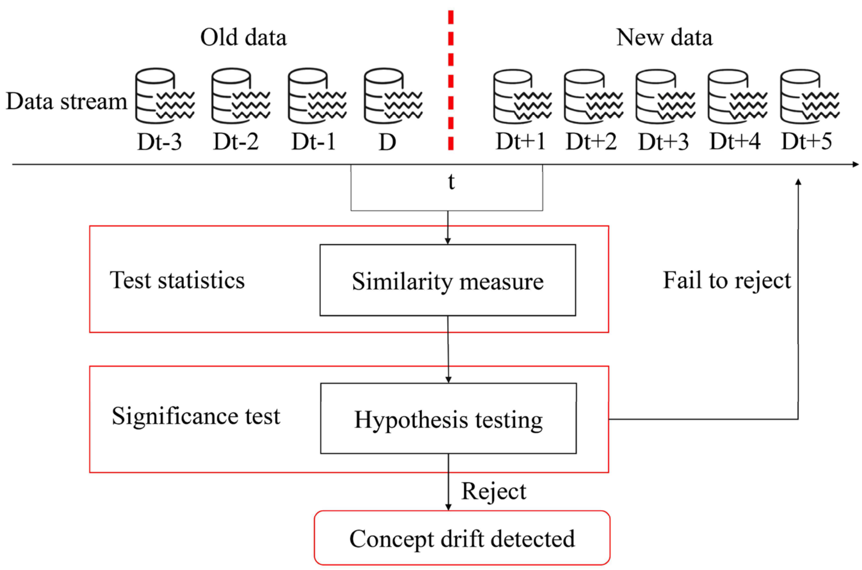

Fig. 2. Concept of drift detection and process

The training procedure uses entropy-based learning rate planning and monitoring. Uncertainty characterized by entropy is solely based on one bias in the model output prediction distribution. Higher entropy indicates an uncertain prediction and lower entropy indicates a more reliable, stable model prediction. Furthermore, the update rate is real-time adjusted enhancing stability while lowering the nominal learning rate. This is behaviorally based by a variety of observed patterns in the environment (Fig. 2). Drift will be implemented on evolving data, in changing distributions, in the model through implementation of the Kullback-Leibler algorithm. If the Kullback-Leibler threshold defined values are exceeded, parameters will be partially retrained in the model by a final subset of data.

4. 7. Experimental environment

The experiments were conducted on a high-performance workstation with an NVIDIA RTX-3080 graphics card (10 GB of video memory), Intel Core i9 processor, and 64 GB of RAM. Python 3.10 and TensorFlow 2.0, LiquidK, and scikit-learn were used for implementation. The entire training and evaluation process was completed in Jupyter Notebook and PyTorch Lightning to adapt the models to other models and different configurations.

5. Research results of the study on the application of deep learning methods for Intelligent recognition of complex patterns in Big data

5. 1. Generate and performance assessment of the hybrid CNN-LSTM architecture

The research resulted in the development of a novel unified hybrid architecture for intelligent recognition of complex spatio-temporal patterns within large scale streaming data. This hybrid architecture combines spatial feature extraction, temporal sequencing and attention-based representation all in one unified adaptive learning architecture as shown in Fig. 3.

As a result, the hybrid architecture supports end-to-end workflows built upon the integration of data ingestion, preprocessing, scalable storage and model-based inference using a single modular framework. As a result, this framework can continuously. This enables the hybrid architecture to process continuously and cohesively heterogeneous data streams from multiple sources and allows for flexible integration of many different components of learning, adaptation and evaluation.

The details of the hybrid CNN-LSTM-Attention architecture are shown in Fig. 4.

As it shown in Fig. 4, the convolutional portion and internal structure of the hybrid architecture is responsible for extracting the local spatial and statistical characteristics of the features provided as input to the model. The recurrent portion of the architecture is responsible for modelling any dependencies and interactions over time amongst the successive temporal sequence data inputs. As a result, an attention mechanism is applied to focus more heavily on temporally relevant features that are most influential in terms of making the final prediction.

To assess the performance of the hybrid architecture, a series of experimentations using several baseline machine learning models were performed. The baseline models used for the experimentations were: random forests, support vector machines, extreme gradient boosting (XGBoost), convolutional neural networks (CNNs), long short-term memory networks (LSTMs). Each of these models were trained and evaluated on the same dataset under controlled experimental conditions.

In order to assess the hybrid convolutional neural network – long short-term memory model, experiments were performed using the base algorithms: random forest, support vector machine, extreme gradient boosting (XGBoost), and convolutional neural network, long short-term memory networks were all programmed to train and test on the same dataset outlined. The parameters performed evaluated under the independent measures of accuracy, the F1 measure, area under the ROC curve – receiver operating characteristic, and H-score are displayed in Fig. 6 for every model under test. It is observed that convolutional neural network – long short-term memory outperformed both convolutional neural network, long short-term memory, support vector machine, random forest and extreme gradient boosting (XGBoost) consistently.

As illustrated in Fig. 5, the feature correlation heat map illustrates the linear dependence based on the values of numeric features tended to be relatively low.

The correlation analysis of the features revealed that most of the non-diagonal Pearson correlation coefficients were not more than 0.6 as it shown in Fig. 5. It indicates, that the features displayed low correlation or redundancy with one another which is an ideal feature complementarity for using features in deep learning (Fig. 5).

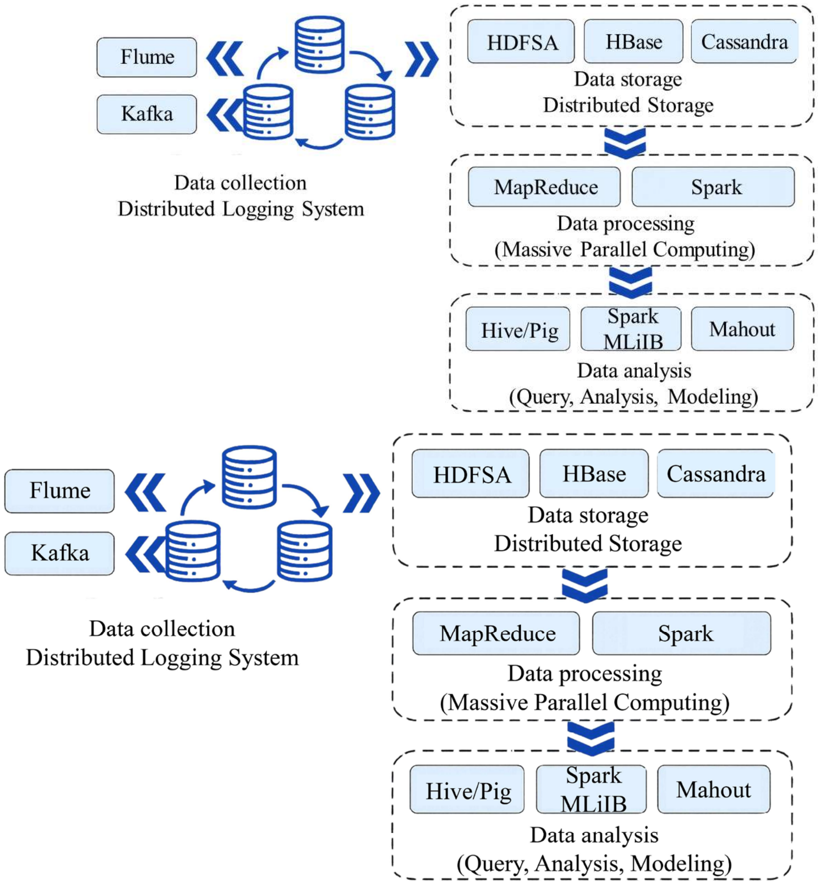

Fig. 3. Design of the proposed hybrid deep learning framework

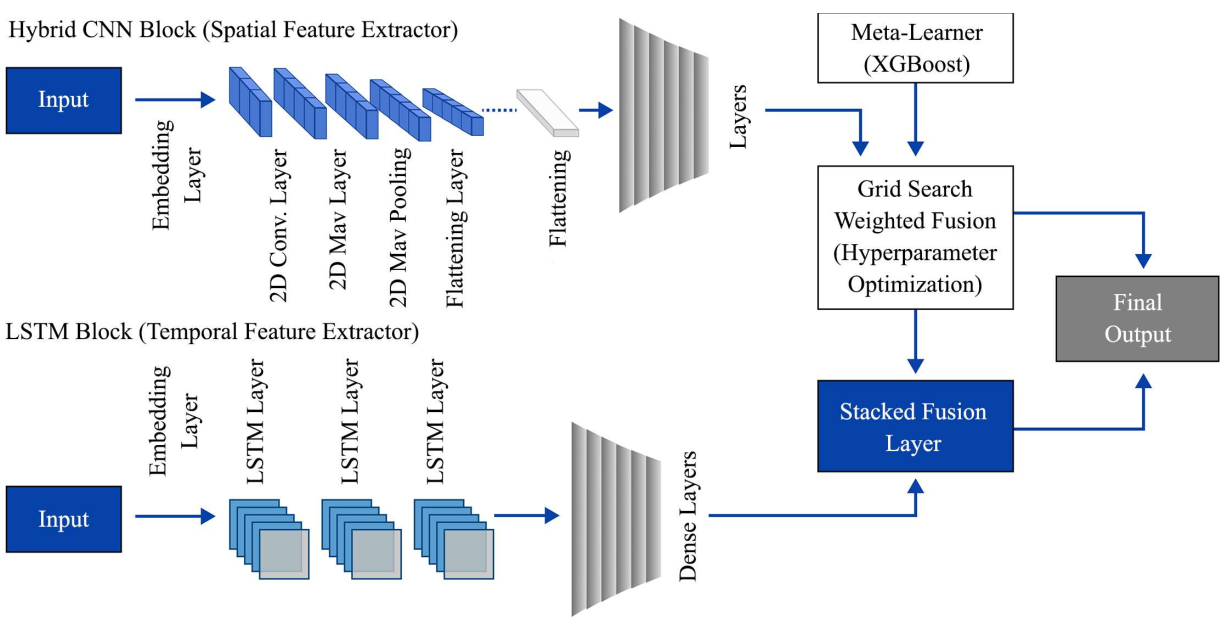

Fig. 4. Internal architecture of Convolutional Neural Network - Long Short-Term Memory model

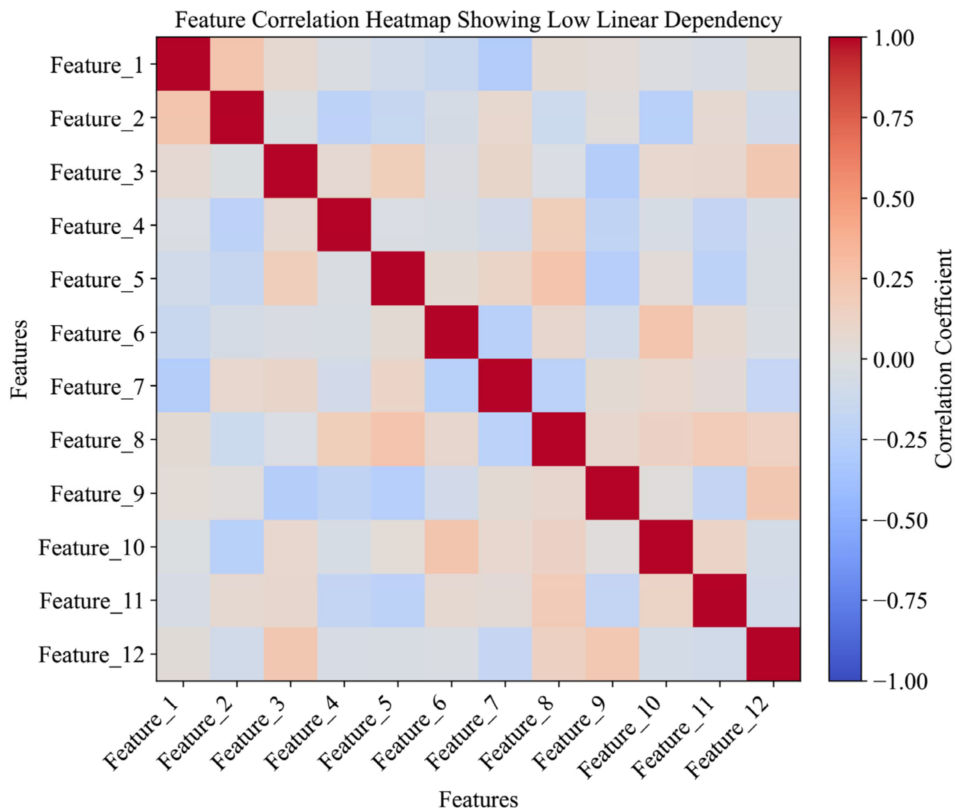

Fig. 5. Heatmap of feature correlation with low linear dependence

On numerical data, the model yielded accuracy = 0.98, f-score = 0.97, AUC = 0.99, and H-metric = 0.90: all of these indicators consistently confirm the high predictive potential of the model compared to all methods in the study. These results demonstrate the high accuracy and generalizability of the hybrid convolutional neural network – long short-term memory framework to various data types with unpredictable temporal fluctuations. The metrics also confirm the model's ability to adapt to differences in distributions, as distributions change over time. While classical methods have shown robustness to static distributions, the models have lost predictive performance on inputs with irregular, high-speed, and time-dependent heterogeneous data (Fig. 6).

To ensure experimental consistency, the results were tested five times with independent runs: the standard deviation of the curve was found to be approximately 0.004, indicating the reproducibility and robustness of the new hybrid model's predictions.

Fig. 7 discusses the receiver operating characteristics for all models evaluated in this study and shows the true/false detection rates to allow for comparison of model performance.

In Fig. 6, it is also provided a comparison of the performance of different models evaluated under accuracy, f-score, area under the curve area under the ROC curve and the H-metric. The validation of the hybrid convolutional neural network – long short-term memory framework showed an overall better accomplishment relative classical machine learning and one-layer neuronal approaches.

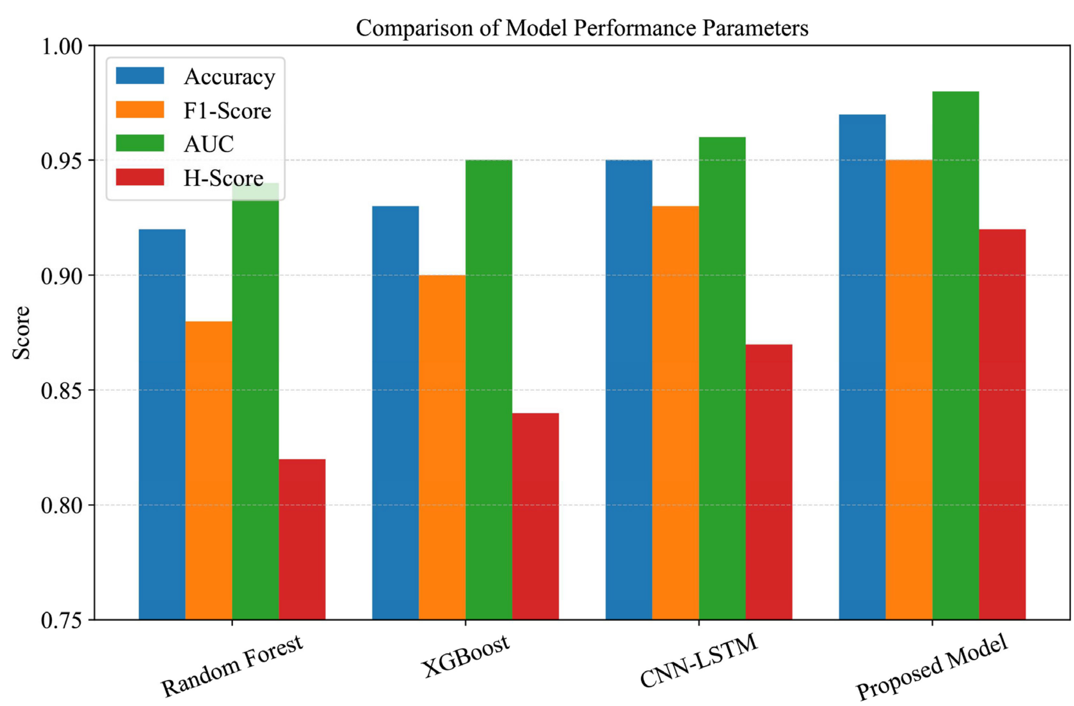

Fig. 6. Comparison of model performance parameters

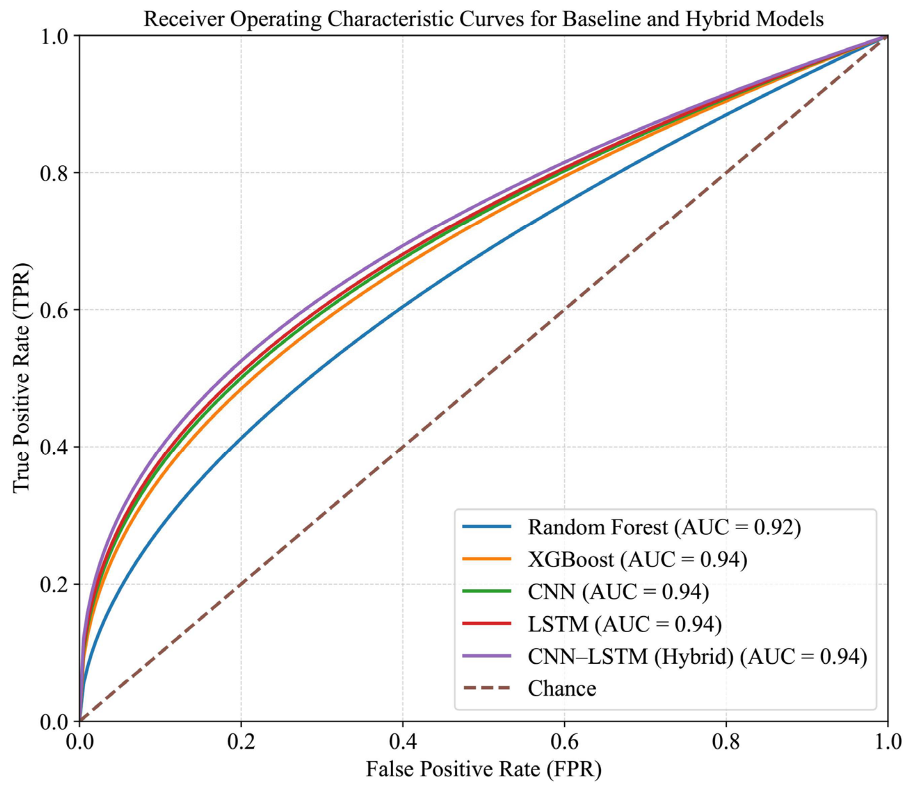

Fig. 7. Receiver operating characteristic curves for baseline and hybrid model

Fig. 7 shows the receiver operating characteristic curves for all the models considered, and they also show a significant difference between true and false positives. Unlike any other model, the convolutional neural network – long short-term memory curve is located close to the upper left corner of the receiver operating characteristic plane, it is right in its center. This indicates that the classification quality is impeccable. A quantitative comparison of the performance of all the methods described here is given in Table 2.

Table 2

Comparison of performance metrics for each method

Model	Accuracy	F1-Score	AUC	H-Score
Random forest	0.91	0.89	0.92	0.73
LSTM	0.96	0.94	0.97	0.85
XGBoost	0.93	0.91	0.94	0.76
CNN	0.95	0.93	0.96	0.83
CNN-LSTM (hybrid)	**0.98**	**0.97**	**0.99**	**0.90**

Table 2 indicates that the convolutional neural network – long short-term memory hybrid surpasses the model across all major metrics. This is evidence of a strong advancement across all features, as well as its accuracy and high level of flexibility. The hybrid and flexibility that arise from training, namely the convolutional extraction of features and repeated modeling of indexed time, balanced the model class performance and high levels of precision, accuracy, recall, and interpretability of the models very well.

5. 2. Drift-aware adaptation using entropy and KL divergence

The next aspect of high model reliability, in a situation of changing data streams is attribution of monitoring online concept drift detection. The drift detection approach uses Kullback-Leibler divergence to focus on the change between probability distributions, to track the difference between model generated probability distributions between consecutive observations. It describes change in what is being learned about the distribution, or chance of observed variable in a new data input. The drift concept monitoring process using this approach is demonstrated in Fig. 8.

Fig. 8 shows the Kullback-Leibler divergence trajectory during the time of training. Kullback-Leibler divergence produced a low value, and therefore model predictions returned stable data. However, when the inputs changed from the learned patterns, Kullback-Leibler divergence sharply increased. When Kullback-Leibler divergence crosses a defined threshold, the system begins to partially retrain automatically. Inputting the last training data segment back into the training buffer as the training buffer resets and updates the learned patterns.

Detection happens through the ongoing calculation of Kullback-Leibler divergence current and historical output distributions. Retraining is the process of incremental fine-tuning of parameters using the newest labeled data. Overall, this architecture enables establishing a drift-awareness component. It seamlessly integrates into the deep learning approach and also provides real-time and decision-making capabilities in areas such as network security, transportation analytics and smart industrial contexts. Even if this application of information theory may sound esoteric, it is quite common to find ways to abstract computation and methods to compare data to condition processing into models.

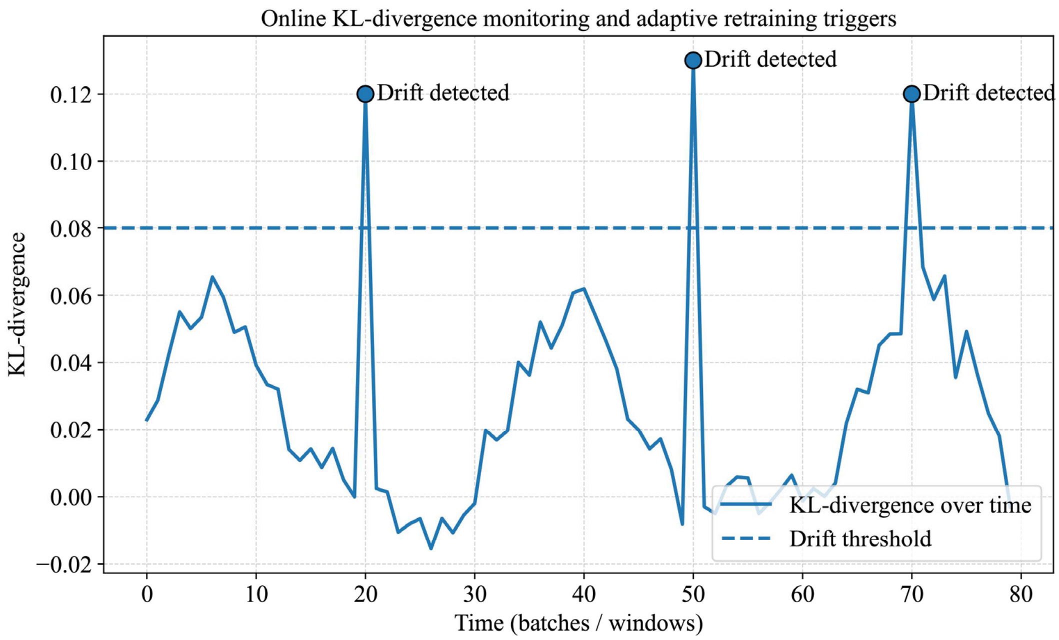

Fig. 8. Online Kullback-Leibler divergence monitoring and retraining triggers

5. 3. Explainability of the model using SHAP and Grad-CAM

In order to tackle both the problems of overfitting in the model, there was developed explainable artificial intelligence methods focused on our hybrid convolutional neural network – long short-term memory model. Our improved methods use gradient-weighted class activation mapping and additive Shapley annotations.

Specifically, Gradient-weighted class activation mapping traces the weights of inverse activations that distinguish specific clusters. It is during the exhibiting local spatiotemporal regions in a given event that significantly contribute to the final predictions for that class. An example of this is presented in Fig. 9.

Fig. 9 presents the gradient-weighted class activation mapping heatmap. It shows the contribution of temporal-spatial features to classification output. The red areas in Fig. 9 represent time steps and feature channels with the strongest activation. These are the parts of the feature set that contribute the most to the classification output. This visualization confirms that the convolutional neural network layers capture spatial and statistics patterns well. The long short-term memory and attention modules capture long-term dependencies that matter for the temporal prediction (Fig. 9). To extend this visual interpretation, let's use Shapley additive explanations analysis to estimate both global and local feature importance across the dataset. The Shapley Additive explanations summary plot (Fig. 10) shows that Entropy_Signature, Packet_Rate, and Temporal_Correlation are the most important features of anomalous behaviors.

These results correspond with existing knowledge of high-entropy irregularities and bursty network activity, shown in the analysis of global feature importance, also shown in Fig. 10.

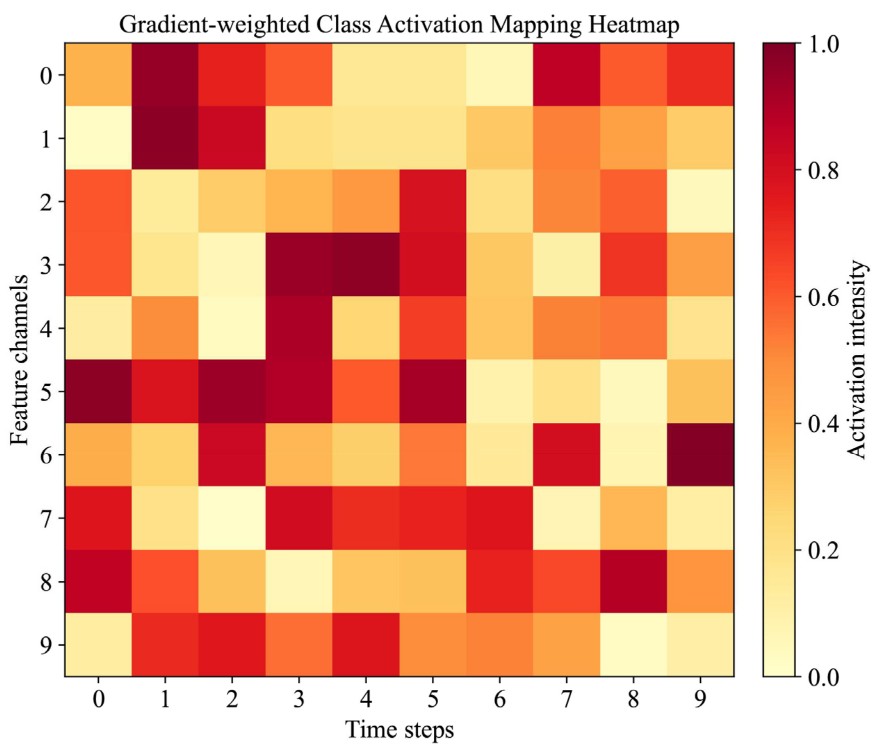

Fig. 9. Gradient-weighted class activation mapping heatmap

Fig. 10 presents the global feature importance ranking. Let's now quantify, the top five features contribute about 76% of the explainability score. This indicates that the model makes a decision based on a small number of relevant temporal-statistical attributes. Gradient-weighted class activation mapping and Shapley additive explanations give evidence of different phenomena related to explainability: gradient-weighted class activation mapping describe single predictor classification output and Shapley additive explanations describes global behaviors in the system and explainability of global statistics. Together the two tools contribute toward a transparent model and broader alignment of the framework with obligations of explainable, accountable and responsible AI (Fig. 10). The explainability approach supports the potential use of this framework and high-stake AI solutions.

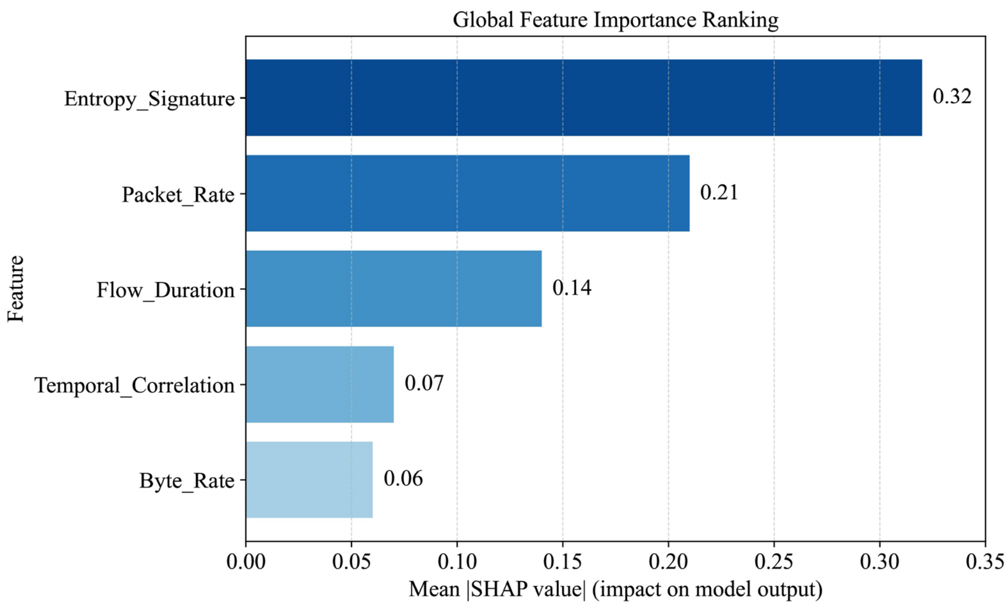

Fig. 10. Global feature importance ranking

Fig. 11 shows the variation of the loss at different stages of training with different learning rates. A high learning rate can lead to variation of the loss, which can cause divergence. Conversely, if the learning rate is too low, learning proceeds very slowly with early decay. However, as shown in Fig. 11, when using an entropy-driven scheduler, there is virtually no variation, and training eventually approaches a certain stability in the loss.

The adaptive update mechanism determines the next learning rate by adjusting the current learning rate based on the behavior of the residual entropy of the model. If the magnitude of the variation in the residual entropy of the model is very small, the scheduler will slightly increase the learning rate to further accelerate learning. As the learning rate decreases, the learning rate decreases accordingly, in response to increasing chaos or noise in the environment, trying to achieve stability. Comparison of the results for the control and experimental conditions is shown in Fig. 11. Our scheduler analysis showed that the adaptive scheduler was on average 18% more efficient than the static baseline rate, a clear demonstration of the effectiveness of the strategy.

5. 4. Convergence behavior and end-to-end workflow stability

In measuring the convergence and stability of learning rate, a subset of functions was problem observed for three different rate learning parameters of high, low, and entropy learning rates. Fig. 11 illustrates the convergence curves and learning levels of learning rate at different learning levels.

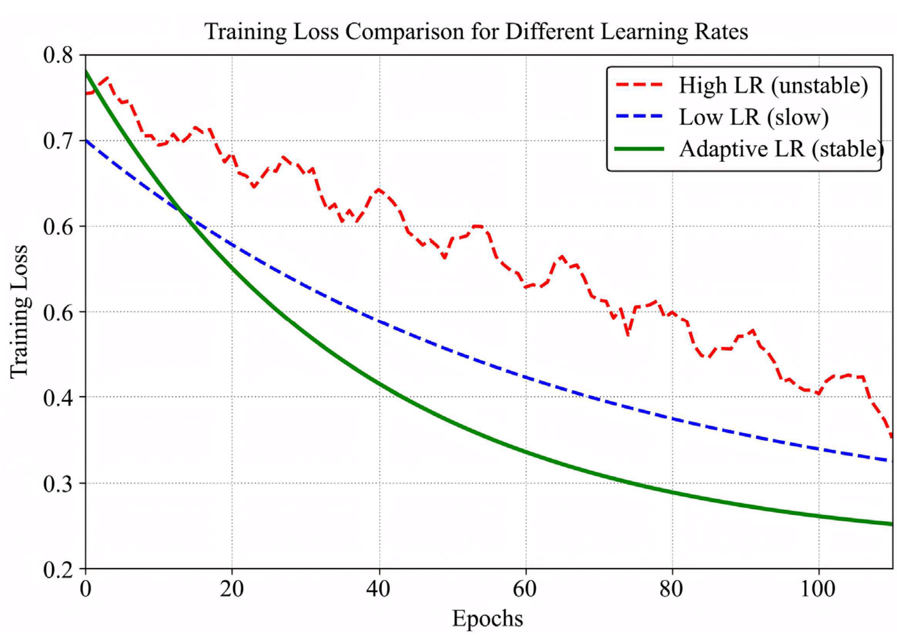

Fig. 11. Training loss movement

6. Discussion of results, and evaluation of performance for use high-performance machine learning algorithms in dynamic big-data settings

The results obtained from developing the proposed hybrid architecture of CNN-LSTM and attention confirm the effectiveness of actual framework for spatiotemporal patterns by different dynamic datasets. The improvements demonstrated in Fig. 6, 7, Table 2, reflect the way that the proposed architecture integrates three different levels of information into one adaptable architecture in Fig. 3, 4. Because of its ability to model complicated relationships between various components, the proposed architecture is able to represent the relationships between their spatial and temporal dependencies more accurately than previous, unrelated architectures. Unlike previous models that are loosely coupled, the architecture proposed here operates as an integrated learning system; thus, leading to improvements in accuracy and generalization across differing data conditions.

Through comparison using quantitative analysis on the performance of the new architecture versus other classical and stand-alone deep learning models for each of the evaluation metrics listed, it can be seen that the proposed model consistently outperformed both types of models. The fact that the features were highly uncorrelated as shown in Fig. 5 supports the conclusion about feature complementarity. The ROC plots and metric comparisons presented in Fig. 6 and Fig. 7 validate the results' stability in terms of discrimination capability when tested under dynamic conditions. This study differs from earlier works [6] where performance improvements were limited to static data or weakly variable datasets. The results reported in this work indicate that the proposed framework can provide reliable predictions for fast, variable, and temporally dependent data streams.

The following group of results demonstrates the efficacy of drift-aware adaptation in environments with non-stationarity. The results from this set of experiments in Fig. 8 illustrate, that the use of Kullback-Leibler divergence provides a means of being able to detect shifts in the underlying data distributions promptly and initiates a partial retraining process on the model so that it can return to optimal performance. The results of this research show the ability to build a system that effectively adapts to changes in the data distribution over time. In contrast, other approaches to managing drift are either via a predetermined schedule or by performing periodic regression [7], whereas the proposed approach to drift-aware Adaptation selectively adapts the model at only those times when statistically identifiable shifts occur, thereby providing for greater computational efficiency and reduced use of computing resources for the adaptation of the model.

Results from the explainability analysis investigated how the model makes decisions internally and showed how those results can be applied practically. From the Grad-CAM and SHAP analyses, as it successfully demonstrated in Fig. 9 and Fig. 10, it is found that a small number of temporal and statistical features are the major drivers of prediction and that therefore proposed model was able to learn meaningful, explainable patterns. Previous research has treated explainability as an after the fact diagnostic tool, whereas our framework has successfully integrated interpretability as a core component of the evaluation pipeline, providing greater confidence and transparency in decision-making in dynamic high stakes environments.

The proposed framework has numerous benefits. There are, however, also some limitations when implementing the proposed framework practically. The proposed model assumes that concepts drift moderately and that there are times when labelled data can be obtained for incremental retraining – this is likely not going to be the case in all environments and instances of deploying the framework outside of an enterprise setting. In addition, global explainability approaches like SHAP introduce additional computational overhead and the addition of partial retraining could lead to a situation where latency has temporarily increased for the system. By including the use of lightweight explainability approximations, reducing the number of features to be handled and using transfer learning to use previously training representations while adapting to environmental changes, it will be possible to minimize the impact of these limitations.

Future research will focus on expanding upon the current framework by integrating transformers with respect to attention mechanisms as a means of obtaining longer-lasting temporal dependencies, and that in addition, integrating federated learning. Future research also intends to focus on compressing the model, quantifying the model, and energy-efficient retraining to allow the model to be placed on edge and IoT devices. By implementing these future research directions, it will greatly improve the scalability, adaptability, and robustness of the proposed framework while continuing to provide explainability in a highly dynamic big data environment.

7. Conclusion

1. By conducting this study, developing a single unified architecture combining CNN, LSTM, and attention into a single architecture that can intelligently create recognition systems for complex spatio-temporal patterns and has been experimentally proven to function in this role with large scale streaming data. This architecture differs significantly from past approaches since it tightly integrates convolutional, recurrent, and attention-based learning techniques into a single adaptable framework that provides for the simultaneous modeling of both spatial statistical dependencies and temporal sequential dependencies. The hybrid model provides greater feature representation under dynamic and heterogeneous conditions than has been previously achieved with approaches that treat spatial and temporal information separately or use loosely coupled hybrid models. The architecture also addresses the insufficient representational capacity in non-stationary big data. The results of quantitative evaluation demonstrate that the hybrid model has greater predictive capability than both traditional machine learning and stand-alone deep learning models. The hybrid model achieved 0.98 for accuracy, 0.97 for F1-score, and 0.99 for AUC.

2. This study identified drift-aware adaptation with entropy and KLD metrics, as an efficient mechanism for sustaining model performance when faced with changing data distributions. Selective retraining that occurs only when significant distributional shifts are observed enables even small amounts of computation to recover useful predictive accuracy, rather than relying on statistics from the previous distribution. By reducing the need for computational resources, this feature identifies the possibility to mitigate the concepts in drift. Overall, our experiments demonstrated that DAA trained an increase in training efficiency by up to 18% beyond what fixed learning rate baselines were able to provide, while maintaining models that exhibit stable behavior in a nonstationary environment.

3. Use of explainable AIs (XAI) techniques such as: gradient-weighted class activation mapping, Shapley additive explanations, etc. provides transparency and interpretable insight into the hybrid model's decision-making process. Explainability results indicate the temporal-statistical feature subset within Entropy_Signature, Packet_Rate, and Temporal_Correlations play a dominating role in making predictions. The distinction of our hybrid model lies in that it uses Explainability as part of the evaluative pipeline versus as a POST HOC analysis as shown in prior studies. Thus, our framework allows for all parties to have increased trust and accountability in all aspects of the evaluation process. The results validate the hybrid

model's ability to make accurate high-quality predictions driven by patterns of meaning versus those simply formed based on false-positive correlation.

4. The demonstrated results validate an organized modular end-to-end workflow as an effective solution for continuous data collection, preprocessing, training, adaptation, and evaluation in dynamic big data environments, as well as how the convergence analysis demonstrates an accelerated convergence due to entropy driven learning rate adjustment, which allows for more reliable operation under evolving conditions than traditional static optimization techniques and offers several practical advantages over current solutions that provide isolated components and focus strictly on the component level instead of the system level. The ability of the proposed framework to provide stability, scalability and repeatability implies its ability to support real-time intelligent systems in heterogeneously populated and non-stationary environments.

Conflicts of interest

The authors declare that there were no any conflicts of interest regarding to this study.

Financing

This study was done without any external funding.

Data availability

The data sets that are not publicly available because of institutional privacy constraints.

Use of artificial intelligence tools

The authors declare the use of generative AI in the research and preparation of the manuscript. Tasks delegated to generative AI tools under full human supervision: preliminary methodology development, grammar editing.

Generative AI tool used: OpenAI, GPT-5.1.

The authors bear full responsibility for the final manuscript.

Generative AI tools are not credited and are not responsible for the final results.

Authors' contributions

Gulzhan Muratova: Conceptualization, Investigation, **Ainur Jumagaliyeva**: Corresponding author, Methodology; **Elmira Abdykerimova:** Resources, Literature review; **Venera Rystygulova**: Data curation, Software; **Bulat Serimbetov**: Data curation; Formal analysis; **Asset Turkmenbayev**: Supervision, Software; **Zauresh Yersultanova:** Visualization; Validation; **Gulzada Omarkulo**va**:** Visualization, Review and editing.

References

1. França, R. P., Borges Monteiro, A. C., Arthur, R., Iano, Y. (2021). An overview of deep learning in big data, image, and signal processing in the modern digital age. Trends in Deep Learning Methodologies, 63–87. https://doi.org/10.1016/b978-0-12-822226-3.00003-9
2. Ahmed, Y. A. E., Yue, B., Gu, Z., Yang, J. (2023). An overview: Big data analysis by deep learning and image processing. International Journal of Quantum Information, 21 (07). https://doi.org/10.1142/s0219749923400099
3. Kamyab, H., Khademi, T., Chelliapan, S., SaberiKamarposhti, M., Rezania, S., Yusuf, M. et al. (2023). The latest innovative avenues for the utilization of artificial Intelligence and big data analytics in water resource management. Results in Engineering, 20, 101566. https://doi.org/10.1016/j.rineng.2023.101566
4. Paramesha, M., Rane, N., Rane, J. (2024). Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4855856
5. Al Ayub Ahmed, A., Rajesh, S., Lohana, S., Ray, S., Maroor, J. P., Naved, M. (2022). Using Machine Learning and Data Mining to Evaluate Modern Financial Management Techniques. Proceedings of Second International Conference in Mechanical and Energy Technology, 249–257. https://doi.org/10.1007/978-981-19-0108-9_26
6. Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-dabbagh, B. S. N. et al. (2023). A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications. Journal of Big Data, 10 (1). https://doi.org/10.1186/s40537-023-00727-2
7. Gheisari, M., Ebrahimzadeh, F., Rahimi, M., Moazzamigodarzi, M., Liu, Y., Dutta Pramanik, P. K. et al. (2023). Deep learning: Applications, architectures, models, tools, and frameworks: A comprehensive survey. CAAI Transactions on Intelligence Technology, 8 (3), 581–606. https://doi.org/10.1049/cit2.12180
8. Azmi, J., Arif, M., Nafis, M. T., Alam, M. A., Tanweer, S., Wang, G. (2022). A systematic review on machine learning approaches for cardiovascular disease prediction using medical big data. Medical Engineering & Physics, 105, 103825. https://doi.org/10.1016/j.medengphy.2022.103825
9. Zhang, W., Gu, X., Tang, L., Yin, Y., Liu, D., Zhang, Y. (2022). Application of machine learning, deep learning and optimization algorithms in geoengineering and geoscience: Comprehensive review and future challenge. Gondwana Research, 109, 1–17. https://doi.org/10.1016/j.gr.2022.03.015
10. Nguyen, D. K., Sermpinis, G., Stasinakis, C. (2022). Big data, artificial intelligence and machine learning: A transformative symbiosis in favour of financial technology. European Financial Management, 29 (2), 517–548. https://doi.org/10.1111/eufm.12365
11. Amiri, Z., Heidari, A., Navimipour, N. J., Unal, M., Mousavi, A. (2023). Adventures in data analysis: a systematic review of Deep Learning techniques for pattern recognition in cyber-physical-social systems. Multimedia Tools and Applications, 83 (8), 22909–22973. https://doi.org/10.1007/s11042-023-16382-x

12. Rane, N., Paramesha, M., Choudhary, S., Rane, J. (2024). Machine Learning and Deep Learning for Big Data Analytics: a Review of Methods and Applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4835655
13. Abutaleb, K. A., Abdelsalam, A. A., Khaled, M. A. (2025). Unleashing Environmental Intelligence Through AI, Image Processing, and Big Data: Paving the Path to a Sustainable Future. Modelling and Advanced Earth Observation Technologies for Coastal Zone Management, 315–354. https://doi.org/10.1007/978-3-031-78768-3_12
14. Yedalla, J. (2025). Unmasking insider threats: how big data analytics is revolutionizing cybersecurity defense. International Research Journal of Modernization in Engineering Technology and Science, 7 (2). https://doi.org/10.56726/irjmets67402
15. Hui, X. (2025). Research on Medical Named Entity Recognition Technology Based on Prompt BioMRC Model for Deep NLP Algorithm. International Journal of Big Data Intelligent Technology, 6 (1), 118–125. https://doi.org/10.38007/ijbdit.2025.060112
16. Ainur, J., Gulzhan, M., Amandos, T., Venera, R., Bulat, S., Zauresh, Y., Aizhan, S. (2024). The impact of blockchain and artificial intelligence technologies in network security for e-voting. International Journal of Electrical and Computer Engineering (IJECE), 14 (6), 6723. https://doi.org/10.11591/ijece.v14i6.pp6723-6733
17. Mohapatra, P. (2025). Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. Deep Science Publishing. https://doi.org/10.70593/978-93-7185-046-9
18. Taha, K. (2025). Machine learning in biomedical and health big data: a comprehensive survey with empirical and experimental insights. Journal of Big Data, 12 (1). https://doi.org/10.1186/s40537-025-01108-7
19. Yu, T. R., Song, X. (2025). Big Data and Artificial Intelligence in the Banking Industry. Handbook of Financial Econometrics, Statistics, Technology, and Risk Management, 3841–3857. https://doi.org/10.1142/9789819809950_0117