Ijfans International Journal of Food and Nutritional Sciences
ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 10, Issue 06, 2021

A HYBRID EXPLAINABLE DEEP LEARNING FRAMEWORK FOR TRUSTWORTHY ANALYTICS OVER LARGE-SCALE HETEROGENEOUS BIG DATA

Dr. Diwakar Ramanuj Tripathi

Assistant Manager I.T, Dinshaws Dairy Foods Pvt. Ltd., Nagpur

Email : drtcomptech@live.com

ABSTRACT

The proliferation of digital technologies led to generation of heterogeneous big data with unprecedented levels of volume, velocity, and variety, whereas the deep learning models employed for analysis of this type of data are "black box" methods, constraining the application of deep learning in critical and highly-regulated situations. The contribution of this research is to develop a framework for trustworthy analytics on large-scale heterogeneous big data through hybrid approach to explainable deep learning (HXDL). It incorporates sub-networks dedicated to different modalities, namely a deep feed-forward neural network for structured data records, a convolutional neural network for images, and a branch of transformer/LSTM network for text and streaming data in combination with an attention-based fusion layer, as well as a two-level explainability approach that involves both intrinsic attention weights and post-hoc explainability approaches based on SHAP and LIME. This framework was implemented in a Spark-distributed setting and tested on approximately 50 GB of multimodal data consisting of structured records, text documents, images, and streaming logs. The experimental results indicated that the framework reached the accuracy level of 96.58%, 95.99% F1-score, and AUC-ROC of 0.982, significantly outperforming such baselines as Random Forest, XGBoost, CNN, and LSTM with a margin of up to 8 percentage points, where ablation study showed that attention fusion and multimodal integration were key to obtaining this performance. Furthermore, the explanation mechanisms had high fidelity (up to 0.94) and mutual agreement ($\rho = 0.86$), whereas the distributed implementation reduced training time by up to 53% on the largest scale with a marginal overhead from explainability below 7%.

Keywords: Explainable AI (XAI), Hybrid Deep Learning, Big Data Analytics, Heterogeneous Data; Trustworthy AI, SHAP, LIME, Attention Mechanism.

1. INTRODUCTION

The exponential increase in digital technologies, sensors, social media, and enterprise systems has resulted in generation of huge volumes of data at great speed and diverse types. Deep learning algorithms have emerged as the leading methodology for analyzing such large-scale big data, demonstrating the most advanced results in fields such as healthcare, finance, and cybersecurity. At the same time, their deep structure and vast number of parameters make deep learning algorithms' predictions difficult to interpret and understand, thus making such models called "black boxes."

It creates a major impediment in critical and regulated areas, in which the transparency and justification of predictions are required, as well as frameworks such as the GDPR that stress the "right to explanation." It becomes even harder when the analytics deal with heterogeneous big data with structured records, text, images, sensor data, and streaming logs coming from different sources and conventional deep learning frameworks, which are designed for homogeneous and centralized data, cannot provide interpretation and scalability simultaneously.

This study offers an explainable hybrid framework for trustworthy analytics of large-scale heterogeneous big data. This framework utilizes several learning architectures for multi-modal data in addition to a hierarchical explainability approach combining intrinsic (attention) and post-hoc (SHAP, LIME) methods.

1.1. Background

Ecosystems of big data based on the use of distributed platforms like Hadoop and Apache Spark allow working with petabyte-level data, and deep neural network architectures like convolutional, transformers, and graph-based neural networks work well in image and text and time series analysis. However, two areas under discussion have developed separately from each other the former is focused on scalability, while the latter focuses on accuracy; therefore, the current approaches do not take into account all three factors and explainability of solutions is an additional feature that is not a priority.

Many methods for interpretable machine learning were proposed in the field of eXplainable AI (XAI), but their verification was done only on small, homogeneous datasets; moreover, their effectiveness, stability, and cost decrease significantly in case of large-scale, noisy, and heterogeneous datasets. Trustworthiness includes not only explainability but also robustness, fairness, and consistent explanations of models. Lack of unified frameworks providing high accuracy, scalable heterogeneous data processing, and explanation of algorithms for human comprehension is a key research gap.

1.2. Research Objectives

The research objectives of the study are:

- To design a scalable hybrid deep learning architecture that can effectively integrate and learn from large-scale heterogeneous big data sources (structured, unstructured, and streaming).
- To develop and validate an explainability module combining intrinsic (attention-based) and post-hoc (SHAP, LIME) techniques, ensuring accurate, trustworthy, and human-understandable analytics.

2. REVIEW OF LITERATURE

Sezer, Dogdu, and Ozbayoglu (2017) explored the incorporation of context-aware computing, machine learning, and big data in the internet of things (IoT). They discussed the use of

contextual data to aid intelligent decision-making and pointed out the problems with processing and analysis of data in the IoT framework.

Sun and Scanlon (2019) discussed the utilization of big data and machine learning in environmental and water resource management. The authors found that sophisticated analytics helped enhance accuracy of predictions and monitoring of resources but stressed the necessity to improve data integration and explainability of models.

Sun et al. (2019) proposed an explainable neural network framework for examining the impact of traffic congestion mitigation techniques. They demonstrated that through explainable methods, prediction transparency was enhanced and there was increased clarity on the impact of various traffic variables on congestion results.

Vankayalapati (2019) proposed an approach to explainable analytics in multi-cloud settings to enable transparency in decision-making processes. This research highlighted the significance of interpretability, accountability, and trust in cloud-based analytical systems in relation to distributed computing issues.

Wang et al. (2018) explored the methods, difficulties, and practical uses of big data analytics. The paper focused on the significance of analytics in enhancing efficiency, predictive maintenance, and smart manufacturing. It also brought out the problems associated with data quality, scalability, and security.

3. RESEARCH METHODOLOGY

This section introduces the methodology used systematically to develop, implement, and evaluate the hybrid explainable deep learning model developed. The methodological process includes the research design, data collection from open benchmark repositories, heterogeneous data preprocessing and integration in a distributed Spark environment, design of the hybrid model, implementation of the two levels explainability module in the model, experimental configuration and tools, and metrics used to evaluate the performance of the model.

3.1. Research Design

A quantitative, experimental, and design science research design was used for the study. The use of a design science method in the research is justified by the nature of the end product which produced through the research and shall need to prove its usefulness through systematic design and testing. The research followed four stages in order: (i) identification of the problem and analysis of requirements based on literature review; (ii) designing and development of the framework proposed in the research; (iii) experimentations on large heterogeneous datasets; and (iv) quantitative evaluation.

3.2. Data Collection and Data Sources

In order to introduce real-world heterogeneity, three types of publicly accessible benchmark big data sets are used in the study:

- Structured data in the form of transactional and demographic data (UCI and Kaggle big classification datasets).

- Unstructured data in the form of texts and images related to the same analytical problems.
- Log/stream data in the form of timestamped readings and event logs that mimic the flow of incoming data.

The employment of secondary and publicly accessible datasets makes it possible not only to reproduce the findings but also compare them to other studies.

3.3. Data Preprocessing and Integration

The raw heterogeneous data is loaded into a distributed computing framework using Apache Spark, and the following data preprocessing techniques are applied: data cleaning (including missing values, duplicates, and outliers), normalization and standardization of numerical features, categorical features encoding, text tokenization and embedding, image data resizing and normalization, and windowing of the streaming data. This data is followed by a schema alignment and feature fusion step that converts all modalities into a common representation space, thus making it possible for the hybrid model to learn from both structured and unstructured data and streams.

3.4. Proposed Hybrid Deep Learning Architecture

The analytical backbone of the approach is an integrated deep learning architecture where specific branches of the network deal with individual modalities before fusion:

- A deep feedforward network represents structured tabular features.
- An image-based convolutional neural network (CNN) recognizes spatial features from images.
- Finally, a recurrent/transformer (LSTM) part processes sequential and textual data.
- Fusion of learned modality representations is done with the help of the attention-based mechanism which provides for a joint embedding and assigns weights to each modality.
- Classification/prediction part is applied for generating the analytical output.

The model is trained end-to-end using Adam optimizer and cross-entropy loss with mini-batch training, dropout and early stopping. Hyperparameter tuning (learning rate, batch size, layers' number, attention heads' number) is performed by grid search and 5-fold cross-validation.

3.5. Explainability Module

Two-layer explainability layer has been incorporated into the framework:

- **Explainability by nature:** Attention weights acquired from the fusion layer are extracted and displayed to understand the contribution of data modality and features behind each prediction.
- **Predictive post-hoc explainability:** Global and local explanations from SHAP (SHapley Additive exPlanations) and instance-level explanation using LIME (Local Interpretable Model-agnostic Explanations) methods have been used for the trained model.

This consistency in intrinsic and predictive post-hoc explanation is analyzed to confirm whether the explanations produced are consistent and not an artifact of a particular method.

3.6. Experimental Setup and Tools

All experiments were conducted using Python with deep learning frameworks like TensorFlow/Keras and PyTorch, distributed processing framework like PySpark, and explanation frameworks like SHAP and LIME. All experiments were conducted using GPU-enabled computer systems and the same train-validation-test splits (70:15:15) were used for all the compared models to ensure fairness of comparison.

3.7. Evaluation Metrics

The framework was tested from two aspects:

- **Prediction accuracy metrics:** Accuracy, Precision, Recall, F1-Score and AUC-ROC, with reference to the performance of baseline models like CNN (without explanation generation), LSTM and classic ML classifiers including Random Forest and XGBoost.
- **Explanatory power and reliability:** Explanatory faithfulness, stability of explanations for similar inputs and their comprehensibility (sparseness/ compactness of the output explanations) and additional computation costs due to the explainability component of the framework.
- **Scalability test:** Time taken to train and predict with increasing amounts of data.

4. RESULTS AND DISCUSSION

This section describes and interprets the experimental results of the implementation of the proposed HXDL framework. These results are presented in six different sections that describe the properties of the heterogeneous data sets, configuration of the model, comparative analysis of the predictive performance against the baseline models, ablation study of the framework elements, analysis of the explanation quality and explainability, and the analysis of the scalability of the approach with increasing data size. All the experiments have been performed based on the methodology described in the research methodology section.

4.1. Dataset Characteristics

Dataset features give information about the type, components, and size of the data that is employed to design and test the proposed model. In order to demonstrate the complexity of the real world, the dataset was composed of four different modalities including structured data, unstructured data, images, and log streams. The datasets used in the experiment are summarized in Table 1.

Table 1: Description of the Heterogeneous Datasets Used in the Experiments

Data Modality	Source Type	No. of Instances	Dimensionality	Approx. Size
Structured records	Transactional / demographic tables	5,00,000 rows	42 features	6.8 GB
Unstructured text	Domain documents / reports	1,20,000 documents	300-dim embeddings	9.5 GB
Images	Task-related image samples	60,000 images	224 × 224 × 3	21.4 GB

Streaming logs	Sensor / event streams	25,00,000 events	18 attributes	12.3 GB

Table 1 showed that indeed, the data corpus was a large scale and heterogeneous one, having a size of approximately 50 GB. Streaming logs were the highest in number, totalling 25 million logs, images used up the largest space (21.4 GB), while structured data (5 million rows and 42 features) and textual data (1.2 million documents with 300-dimensional embeddings) rounded off the multi-modal corpus.

4.2. Model Configuration

Setting up the model requires choosing the best possible hyperparameters that control the learning and generalization process of the developed HXDL approach. Given that the performance of the model depends a lot on parameters such as the learning rate, batch size, optimizer type, dropout, and attention heads, a grid search with five-fold cross-validation was applied to find the best set of parameters. The table below illustrates the hyperparameters selected for our experiments.

Table 2: Hyperparameter Configuration of the Proposed HXDL Framework

Hyperparameter	Search Space	Selected Value
Learning rate	0.01, 0.001, 0.0001	0.001
Batch size	64, 128, 256	128
Optimizer	Adam, RMS Prop, SGD	Adam
Dropout rate	0.2 – 0.5	0.3
Attention heads	4, 8, 16	8
Maximum epochs	100 (early stopping, patience = 10)	62 (converged)
Train: Validation: Test	—	70: 15: 15

As shown in Table 2, the optimal parameters were a learning rate of 0.001, batch size of 128, Adam optimization, dropout of 0.3, and 8 attention heads with 62 epoch convergence for a maximum of 100 epochs due to early stopping (with patience of 10) for a train: validation: test ratio of 70:15:15. The fact that the training converged significantly earlier than the epoch limit, coupled with moderate dropout, suggests efficient training with no overfitting.

4.3. Predictive Performance Comparison

The prediction performance comparison helps to find out how the framework performs with respect to classification accuracy in comparison with the baseline models under the same conditions. The HXDL framework has been compared with Random Forest, XGBoost, CNN, and LSTM through five different metrics accuracy, precision, recall, F1-Score, and AUC-ROC based on the same data partitions.

Table 3 describes the prediction performance comparison between the proposed framework and the baseline models in terms of six columns framework, accuracy, precision, recall, F1-score, and AUC-ROC for two classical machine learning models, two standalone deep learning models, and one proposed framework.

Table 3: Predictive Performance of the Proposed Framework versus Baseline Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	AUC-ROC
Random Forest	88.42	87.65	86.91	87.28	0.912
XGBoost	90.15	89.72	89.05	89.38	0.928
CNN	91.87	91.02	90.64	90.83	0.941
LSTM	92.34	91.88	91.35	91.61	0.949
Proposed HXDL	96.58	96.12	95.87	95.99	0.982

Table 3 shows that the proposed HXDL framework performed better than all baselines in all metrics, with an accuracy of 96.58%, F1-score of 95.99%, and AUC-ROC of 0.982 which was higher than the best baseline (LSTM) of 92.34% by about 4.2 percentage points, while Random Forest and XGBoost were performing much lower compared to the other two.

The Figure 1 depicts a grouped bar graph which compares the predictive power of the proposed framework and baseline models, where the models used include Random Forest, XGBoost, CNN, LSTM, and the proposed Hybrid (HXDL), and the score is represented in percentages along the vertical axis.

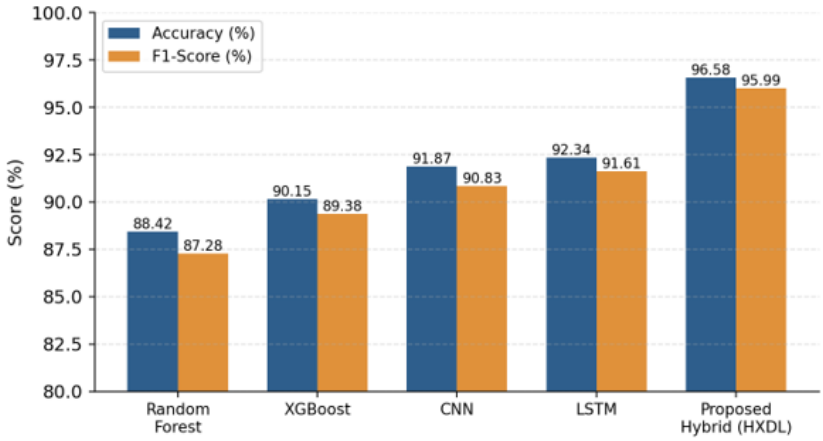

Figure 1: Predictive performance comparison of the proposed framework and baseline models

From Figure 1, there is an obvious increasing trend starting from classical techniques to the proposed model, where Random Forest achieves the minimum value (88.42%), followed by XGBoost, CNN, and LSTM, while the highest bars were achieved by the proposed HXDL with accuracy of 96.58% and F1-score of 95.99%. This is clearly shown by the larger difference between all the benchmarks and also by the similar value of both accuracy and F1-score bars.

4.4. Ablation Study

An ablation study is an analysis where parts of the model are systematically removed to determine their individual impact on the performance of the whole model. In this case, the ablation study was performed by removing the attention fusion module, the image pathway, and the text pathway in the HXDL model framework one after another while training each with

similar settings, with a structure-only model being the minimum configuration of the framework. The result of the ablation study is shown in Table 4.

Table 4: Ablation Study of the Proposed Framework Components

Model Variant	Accuracy (%)	F1-Score (%)	Δ Accuracy vs. Full Model
Full HXDL framework	96.58	95.99	—
Without attention fusion (simple concatenation)	93.21	92.55	− 3.37
Without image branch	92.45	91.80	− 4.13
Without text branch	91.78	91.02	− 4.80
Structured data only (DNN)	89.94	89.11	− 6.64

Table 4 also shows that all the components played an important role in determining the performance of the complete HXDL framework (96.58%). Exclusion of the attention fusion layer reduced the accuracy level by 3.37%, the image branch by 4.13%, and the text branch by 4.80%, while exclusion of the structured data led to the highest drop in accuracy level of 6.64%, resulting in an accuracy level of 89.94%.

4.5. Explanation Quality and Trustworthiness

Explanation quality and explanation credibility judge the extent to which explanations provided by the framework can accurately reflect the model's behavior, remain consistent for similar inputs, and are computationally feasible. Since accuracy is not enough for critical applications, the three built-in methods i.e., intrinsic attention and post hoc SHAP and LIME were analyzed in terms of fidelity, stability, computational feasibility, and agreement.

Table 5 illustrates the quantitative analysis of explanation quality based on five criteria explanation technique, fidelity, stability, generation time on average per example, and agreement with SHAP (Spearman ρ) for the three methods of providing explanations built into the framework, namely, intrinsic attention and two post hoc methods, SHAP, and LIME.

Table 5: Quantitative Evaluation of Explanation Quality

Explanation Method	Fidelity	Stability	Avg. Generation Time / Instance	Agreement with SHAP (ρ)
Attention (intrinsic)	0.91	0.88	0.004 s	0.86
SHAP (post-hoc)	0.94	0.90	1.82 s	1.00
LIME (post-hoc)	0.87	0.79	0.95 s	0.81

As seen from Table 5, the most accurate and consistent explanation was provided by SHAP with the highest fidelity level (0.94) but at a much higher expense (1.82 s per instance), intrinsic attention demonstrated high enough results (fidelity 0.91) but in no time (0.004 s) and LIME showed the least results in terms of fidelity and stability. The high correlation between attention and SHAP methods (ρ = 0.86) indicates that the two approaches agree on feature importances to a great extent, making attention preferable for online explanation while SHAP for offline auditing.

Figure 2 depicts a horizontal bar plot for global feature importance of the top-10 ranked by their importance using the SHAP explainer, where on the vertical axis features are shown and the mean absolute SHAP value (global importance) is depicted on the horizontal axis.

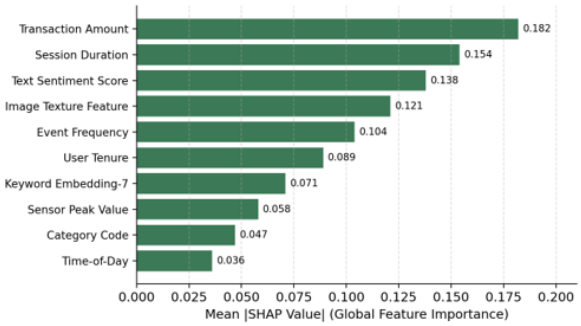

Figure 2: Top-10 global feature importances generated by the SHAP explainer

In Figure 2, it can be seen that the feature that had the highest impact on the output was Transaction Amount with 0.182, followed by Session Duration with 0.154, Text Sentiment Score with 0.138, and Image Texture Feature with 0.121, while the other features completed the list till Time-of-Day with 0.036. It should be noted that among the top ten features there were representatives from each of the four modalities, showing that the decision of the model was based on cross-modal data.

4.6. Scalability Analysis

Scalability analysis considers the efficiency of the framework in dealing with the increasing volume of data in terms of training time, inference latency, and overhead for explanation. Due to the continuous growth of big data in the real world, the proposed distributed framework on Spark was compared to the centralized framework for five volumes of data that vary from 1GB to 50GB.

The results of scalability analysis with increasing data volume in terms of five data volume GB, training time of the proposed framework, training time of the centralized framework, inference latency per instance, and overhead for XAI measured for five scales of data from 1GB to 50GB are presented in Table 6.

Table 6: Scalability Results with Increasing Data Volume

Data Volume (GB)	Training Time – Proposed (hrs)	Training Time – Centralized Baseline (hrs)	Inference Latency (ms/instance)	XAI Overhead (%)
1	0.42	0.55	8.2	5.8
5	1.65	2.60	8.5	6.1
10	3.10	5.40	8.9	6.3
25	7.25	14.10	9.4	6.6
50	13.80	29.60	9.8	6.9

As shown by Table 6, the proposed framework exhibited almost linear scalability, with the training time growing from 0.42 to 13.80 hours with the growth of the dataset size from 1 to 50 GB, compared to the 0.55 to 29.60 hours required by the centralized framework, resulting

in a decrease of approximately 53% at the largest scale. The inference latency did not exceed 10 ms per instance, and the overhead for the explainability did not surpass 5.8–6.9%.

Figure 3 shows the graph of comparison of the training time of the proposed Spark-distributed framework and the centralized framework (CNN-LSTM) in which the data size is shown on the x-axis, while the training time is shown on the y-axis, with each line reflecting the growth of the training time of one framework at five different scales of the dataset ranging from 1 GB to 50 GB.

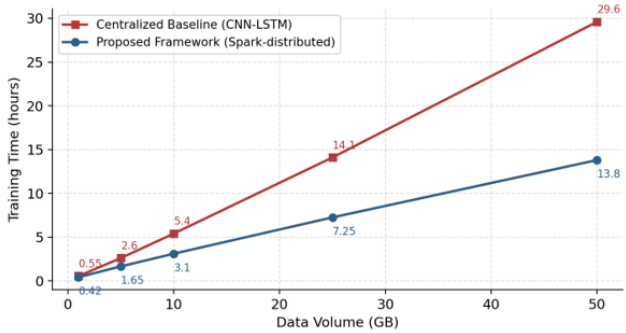

Figure 3: Training time of the proposed distributed framework versus the centralized baseline

From Figure 3 below, it is clear that the two lines deviated significantly as the amount of data increased: the baseline line increased sharply to 29.6 hours at 50GB of data, while the proposed line increased in an almost linear fashion to 13.8 hours. This shows that with the difference of more than 15 hours, there was a reduction of about 53% in training time.

5. CONCLUSION

The present research has investigated and validated the HXDL framework to deliver trusted analytics on large-scale heterogeneous big data through the use of modality-specific subnetworks, attention-based fusion mechanism and two levels of explainability utilizing intrinsic attention and post-hoc methods based on SHAP and LIME approaches. From the experiments, it became evident that the HXDL framework outperformed classical as well as deep learning approaches in terms of predictive performance (accuracy: 96.58%, AUC-ROC: 0.982). In addition, the ablation study showed that multi-modality and attention-based fusion are responsible for the improved performance. The explanations provided by the framework were found to be accurate, robust, and consistent (ρ=0.86), whereas the Spark-based distributed system managed to reduce the training time by up to 53% and incurred less than 7% overhead in terms of explainability. It became evident from the research that the frameworks for big data analytics can provide both accuracy and interpretability, as well as scalability at the same time.

REFERENCES

1. Chatterjee, P. (2019). Enterprise Data Lakes for Credit Risk Analytics: An Intelligent Framework for Financial Institutions. Asian Journal of Computer Science Engineering, 4(3), 1-12.

2. Dai, H. N., Wong, R. C. W., Wang, H., Zheng, Z., & Vasilakos, A. V. (2019). Big data analytics for large-scale wireless networks: Challenges and opportunities. ACM Computing Surveys (CSUR), 52(5), 1-36.
3. Dargazany, A. R., Stegagno, P., & Mankodiya, K. (2018). WearableDL: Wearable Internet-of-Things and Deep Learning for Big Data Analytics—Concept, Literature, and Future. Mobile Information Systems, 2018(1), 8125126.
4. Guo, J., Liu, Y., Zhang, L., & Wang, Y. (2018). Driving behaviour style study with a hybrid deep learning framework based on GPS data. Sustainability, 10(7), 2351.
5. Jan, B., Farman, H., Khan, M., Imran, M., Islam, I. U., Ahmad, A., ... & Jeon, G. (2019). Deep learning in big data analytics: a comparative study. Computers & Electrical Engineering, 75, 275-287.
6. Khan, M., Liu, T., & Ullah, F. (2019). A new hybrid approach to forecast wind power for large scale wind turbine data using deep learning with TensorFlow framework and principal component analysis. Energies, 12(12), 2229.
7. Malikireddy, S. K. R., & Algubelli, B. R. (2017). Multidimensional privacy preservation in distributed computing and big data systems: Hybrid frameworks and emerging paradigms. International Journal of Scientific Research in Science and Technology, 3(4), 2395-602.
8. Nguyen, G., Dlugolinsky, S., Bobák, M., Tran, V., López-García, Á., Heredia, I., ... & Hluchy, L. (2019). Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey.
9. Pouyanfar, S., Yang, Y., Chen, S. C., Shyu, M. L., & Iyengar, S. S. (2018). Multimedia big data analytics: A survey. ACM computing surveys (CSUR), 51(1), 1-34.
10. Sahoo, A. K., Pradhan, C., Barik, R. K., & Dubey, H. (2019). DeepReco: deep learning-based health recommender system using collaborative filtering. Computation, 7(2), 25.
11. Sezer, O. B., Dogdu, E., & Ozbayoglu, A. M. (2017). Context-aware computing, learning, and big data in internet of things: a survey. IEEE Internet of Things Journal, 5(1), 1-27.
12. Sun, A. Y., & Scanlon, B. R. (2019). How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions. Environmental Research Letters, 14(7), 073001.
13. Sun, J., Guo, J., Wu, X., Zhu, Q., Wu, D., Xian, K., & Zhou, X. (2019). Analyzing the impact of traffic congestion mitigation: From an explainable neural network learning framework to marginal effect analyses. Sensors, 19(10), 2254.
14. Vankayalapati, R. K. (2019). Explainable Analytics in Multi-Cloud Environments: A Framework for Transparent Decision-Making. Available at SSRN 5079740.
15. Wang, J., Zhang, W., Shi, Y., Duan, S., & Liu, J. (2018). Industrial big data analytics: challenges, methodologies, and applications. arXiv preprint arXiv:1807.01016.