Ijfans International Journal of Food and Nutritional Sciences
ISSN PRINT 2319 1775 Online 2320 7876
 Journal UGC CARE Listed (Group-I) Volume 10, Issue 02, 2021

AN EXPLAINABLE FEDERATED MACHINE LEARNING FRAMEWORK FOR PRIVACY-PRESERVING PREDICTIVE ANALYTICS IN CLOUD-BASED BIG DATA ENVIRONMENTS

Dr. Diwakar Ramanuj Tripathi
Assistant Manager I.T, Dinshaws Dairy Foods Pvt. Ltd., Nagpur
Email : drtcomptech@live.com

Abstract

The adoption of cloud computing has led to increasing interest in privacy-preserving predictive analytics via federated machine learning. However, despite Federated Learning (FL)'s ability to train models without the need to share raw data, the existing methods lack sufficient privacy protection and model interpretability. In this research, we propose a new framework for explainable federated machine learning that combines differential privacy, secure aggregation and SHapley Additive exPlanations (SHAP). The main goal of our framework is to increase the level of data privacy and prediction explainability in cloud-based big data systems. We have developed our framework based on TensorFlow Federated and compared it with Centralized Machine Learning, Standard Federated Learning and Differentially Private Federated Learning on Adult Income dataset. The results have shown that our framework provides 85.4% accuracy, 0.730 F1-score and 0.891 ROC-AUC, while providing similar performance with Standard Federated Learning. Moreover, our method demonstrates good convergence properties when facing heterogeneity and increases robustness of the model with respect to privacy attacks due to combining differential privacy and secure aggregation methods. Also, SHAP-based explanation of features leads to higher interpretability of the model.

Keywords: Federated Learning; Explainable Artificial Intelligence; Differential Privacy; Secure Aggregation; SHAP; Privacy-Preserving Machine Learning; Cloud Computing.

1. INTRODUCTION

Cloud computing along with the Big Data has revolutionized the predictive analytics field by processing distributed data from different fields including healthcare, financial services, cyber security, and intelligent manufacturing. The traditional centralized machine learning models require all the necessary data to be gathered at the central server and thereby, it poses various issues related to data privacy, security, ownership, and regulations. All these challenges have prompted researchers to develop federated learning models which facilitate collaborative development of the model without exchanging sensitive data.

The FL technique is considered to be an efficient technique to facilitate collaborative training of shared models on distributed data without exposing sensitive data. Even with several advantages, federated learning is susceptible to inference attacks, communications costs, data heterogeneity, and model interpretability challenges. Thus, incorporating the privacy-preserving techniques with XAI like SHAP has gained more importance in cloud-based predictive analytics.

1.1 Background and Motivation

Cloud-based predictive analytics uses data produced by decentralized enterprises, IoT sensors, and edge computing frameworks. While federated learning does away with data sharing altogether, the update of model parameters can expose sensitive data without implementing further security measures. In addition, the black box character of some machine learning algorithms does not allow for user confidence in such systems. Consequently, the use of differential privacy, secure aggregation, and explainable AI could increase both the data security and model transparency of federated learning.

1.2 Research Gap and Objectives

Prior research efforts have predominantly centered on enhancing either the prediction or privacy in federated learning. Relatively less research effort has been put forth towards creating a framework that incorporates privacy preservation and explainability along with evaluating prediction, privacy, communication, and explainability aspects simultaneously.

In order to overcome such gaps in the prior work, this research effort intends to formulate and analyze an explainable privacy-preserving federated learning framework using the following objectives:

- To develop an explainable federated machine learning framework incorporating Differential Privacy, Secure Aggregation, and SHAP explainability for cloud-based predictive analytics.
- To evaluate the proposed framework using comparison analysis of the prediction performance, privacy preservation, communication efficiency, and model interpretability of the framework with the existing centralization and federated learning frameworks.

1.3 Major Contributions

This study presents a common framework incorporating Federated Learning, Differential Privacy, Secure Aggregation, and SHAP explainability for conducting predictive analysis in a privacy-preserving and interpretable manner. This framework is analyzed via extensive experiments based on various performance metrics in terms of prediction, privacy, and communication. As shown by the results of the study, the proposed method provides a successful trade-off between these aspects of machine learning. The rest of this paper deals with the literature review, research methods, experimental results, and conclusion of this study.

2. LITERATURE REVIEW

Advancements in cloud computing, federated learning, privacy-preserving machine learning, and explainable artificial intelligence technologies have made a substantial contribution to secure predictive analytics. Despite the fact that some previous works dealt with distributed learning, privacy, and model interpretability, there is little evidence on how these topics can be combined into a comprehensive framework.

Hashem et al. (2015) gave a review of the emerging cloud computing technology used for big data analytics and discussed major research challenges associated with scalability, security, and privacy of such systems.

Abadi et al. (2016) provided the first implementation of Differential Privacy for deep learning and proved that it was possible to protect the privacy of individuals during model training while preserving good predictive performance. In 2016, Ribeiro et al. (2016) designed LIME algorithm, which was aimed at explaining predictions of machine learning models.

Bonawitz et al. (2017) designed an efficient Secure Aggregation protocol for federated learning, which enabled model update without revealing any information about particular clients.

Antunes et al. (2018) underscored the significance of fairness and transparency in machine learning for trustworthy cloud services. Ramanathan (2018) stressed the importance of federated learning for the preservation of data privacy in cloud-based artificial intelligence systems.

Yang et al. (2019) explained the notions of and use cases for federated machine learning, illustrating the opportunities for collaboration in building models without exposing the raw data to each other. In turn, Atalor (2019) and Oloke (2019) investigated federated learning for privacy-preserving purposes in healthcare and finance.

The more recent literature is concerned with privacy-preserving techniques in predictive analytics. Hartley et al. (2020) summarized the secure data sharing approaches for big data analytics, Alkadi et al. (2020) developed the blockchain-based framework for collaborative intrusion detection in IoT and cloud networks, and Jani (2020) explored the application of federated learning in privacy-preserving healthcare.

Nevertheless, the examined literature reveals substantial advancements in the area of federated learning, differential privacy, secure aggregation, and explainable AI. On the other hand, the existing studies consider these aspects individually or within specific application areas. Consequently, the development of a framework incorporating Differential Privacy, Secure Aggregation, and explainable AI, which is expected to enhance privacy preservation, predictive accuracy, and model transparency, is required. This paper tries to fill the identified research gap by suggesting an explainable federated machine learning framework.

3. RESEARCH METHODOLOGY

This Research suggests a methodology for building an interpretable federated machine learning framework with privacy protection and prediction capability.

3.1 Overall Research Framework

The suggested framework uses a federated learning model where the clients work together to train the global model using their respective datasets without any data exchange among them. The client trains the model using the preprocessed dataset. The encrypted model parameters are sent to the central server where the Secure Aggregation and differential privacy are used for updating the global model. Finally, SHAP is used for providing global and local explanations about the prediction results. The complete process of the suggested framework is presented in Figure 1.

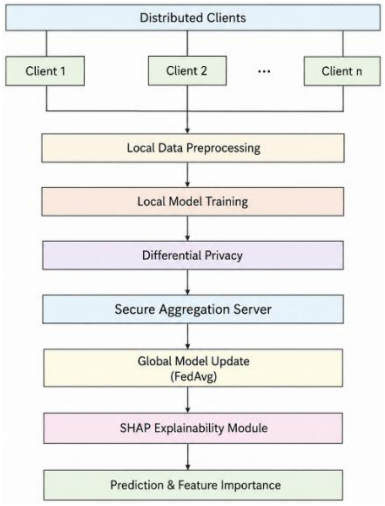

Figure 1. Overall workflow of the proposed explainable federated learning framework

This process involves five distinct steps:

- Preprocessing of data at distributed clients.
- Model training locally.
- Securely aggregating encrypted model parameters.
- Optimizing the global model using FedAvg algorithm.
- Explanation of prediction results using SHAP approach.

3.2 Proposed Explainable Federated Learning Framework

The proposed model makes use of Federated Learning, Differential Privacy, Secure Aggregation, and SHAP based Explainable Artificial Intelligence. The Federated Learning technique allows for training through collaboration without sharing the actual data, whereas Differential Privacy and Secure Aggregation ensure the privacy of local models during transmission. Once the global model is convergent, SHAP explains both local and global feature importances. The global model update is done using the Federated Averaging Algorithm:

$$W^{(t+1)} = \sum_{k=1}^{K} \frac{n_k}{N} W_k^{(t)}$$

where $W_k^{(t)}$ denotes the local model parameters of the kth client, n_k is the number of local training samples, N is the total number of samples, and K represents the participating clients.

3.3 Dataset and Data Preprocessing

Evaluation of the suggested framework was carried out using the Adult Income Dataset available in the UCI Machine Learning Repository. This dataset was preprocessed by removing the duplicate and missing values, using one-hot encoding for categorical data and Min-Max normalization for numerical data, before distributing the preprocessed data among the virtual clients for simulation purposes.

3.4 Experimental Setup

This framework was created using Python 3.11 along with TensorFlow Federated, Scikit-learn, SHAP, NumPy, and Pandas packages. Experiments were performed in Google Colab. Table 1 illustrates the configuration of the experiments.

Table 1. Experimental configuration

Parameter	Value
Programming language	Python 3.11
Framework	TensorFlow Federated
Explainability tool	SHAP
Number of clients	10
Communication rounds	100
Batch size	32
Learning rate	0.001
Optimizer	Adam
Local epochs	5

The following parameter values were chosen in order to achieve a compromise between predictivity, communication, computation costs, and model convergence.

3.5 Performance Evaluation Metrics

The proposed framework was evaluated using Accuracy, Precision, Recall, F1-score, and ROC-AUC. The evaluation metrics were computed as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Where TP, TN, FP, and FN refer to true positive, true negative, false positive, and false negative respectively. Apart from the predictive accuracy, the communication cost, the Differential Privacy budget (ε), and the SHAP feature importance have also been considered to determine the privacy preserving, communication efficiency, and explainability of the model. All these together offer a holistic evaluation of the proposed explainable federated learning paradigm for cloud based predictive analytics.

4. RESULTS AND DISCUSSION

The evaluation of the proposed explainable federated learning system is done on the basis of their predictive accuracy, convergence, explainability, and privacy-preserving capability. Proposed method was compared to Centralized Machine Learning, Standard Federated Learning (FL), and Differentially Private Federated Learning (DP-FL). All experiments were

done on the basis of the Adult Income dataset and with the same experimental setup and the results are averaged over five trials.

4.1 Predictive Performance Evaluation

Prediction accuracy of all examined models was checked using metrics such as Accuracy, Precision, Recall, F1-Score, and ROC-AUC. Results of comparison are shown in Table 2.

Table 2. Comparative predictive performance of the evaluated models

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Centralized machine learning	0.867	0.783	0.721	0.751	0.904
Standard Federated Learning	0.858	0.768	0.708	0.737	0.895
Differentially Private FL	0.842	0.744	0.681	0.711	0.878
Proposed framework	0.854	0.761	0.701	0.730	0.891

From Table 2, it can be seen that the performance of the centralized approach was the best since the approach made use of the entire dataset for training without any privacy restrictions. The conventional approach of federated learning also yielded similar results, proving that learning in a decentralized way is effective. While the DP-FL incurred the maximum performance degradation due to the differential privacy noise, the proposed approach managed to yield a performance of 85.4% accuracy, 0.730 F1-score, and 0.891 ROC-AUC score.

4.2 Model Convergence and Data Heterogeneity

In assessing the training stability of the models, the validation accuracy was observed for 100 rounds of communications. The behavior of the federated models during convergence is shown in Figure 2.

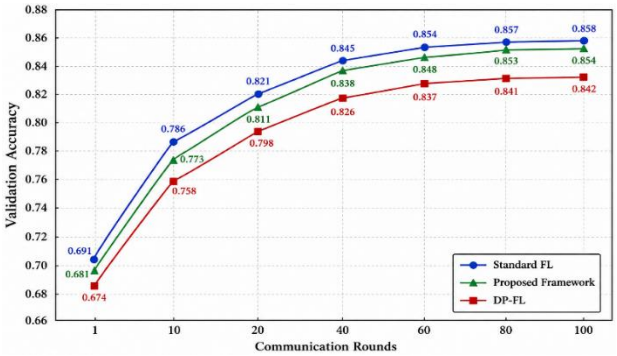

Figure 2. Validation accuracy across federated communication rounds

As is evident from Figure 2 below, convergence for all federated learning approaches occurred consistently. For Standard FL, convergence was quicker compared to DP-FL, which needed more iterations due to noise introduced for privacy reasons. Our approach yielded consistent convergence and surpassed the 85% mark for validation accuracy.

The next test case examined the robustness of the models on heterogeneous client distribution. The findings are provided in Table 3 below.

Table 3. Performance under IID and non-IID data distributions

Model	IID Accuracy	Non-IID Accuracy	Accuracy Reduction	Non-IID F1-score
Standard Federated Learning	0.858	0.839	0.019	0.712
Differentially Private FL	0.842	0.816	0.026	0.683
Proposed framework	0.854	0.836	0.018	0.707

However, as seen from Table 3, there was a slight drop in performance for all the models in question under non-IID data. The proposed model had a performance drop of just 1.8%, proving its robustness in heterogeneous client environment, similar to the performance of Standard FL.

4.3 SHAP-Based Explainability Analysis

In order to analyze the interpretability of the model, SHAP was used to determine the contribution of each predictor variable to the prediction outcome. The feature importance values are provided in Table 4.

Table 4. Global SHAP feature importance results

Rank	Feature	Mean Absolute SHAP Value	General Influence
1	Capital gain	0.184	Higher values increased the likelihood of income above the threshold
2	Marital status	0.143	Married categories generally increased the predicted probability
3	Education number	0.131	Higher education increased the probability of higher income
4	Age	0.112	Middle and higher age groups positively influenced prediction
5	Hours worked per week	0.096	Longer working hours generally increased predicted income
6	Occupation	0.083	Professional and managerial occupations had a positive influence
7	Relationship status	0.071	Household relationships affected prediction outcomes
8	Capital loss	0.059	Higher values produced a moderate positive influence
9	Work class	0.047	Employment category affected prediction probability
10	Sex	0.035	Produced a comparatively limited model contribution

According to Table 4, Capital Gain, Marital Status, Number of Educations, Age, and Hours Worked per Week proved to be the most important predictors of income, which means that this model depended heavily on the socioeconomic characteristics.

This measure of SHAP feature importance is visualized in Figure 3.

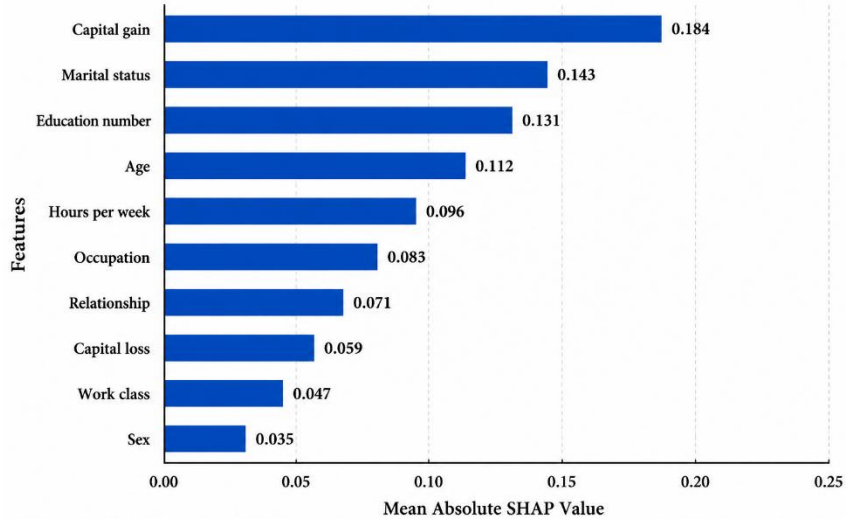

Figure 3. Global SHAP feature importance of the proposed model

Figure 3 shows the domination of economically meaningful attributes used in the prediction. Local SHAP attributions showed that the proposed approach could easily explain each prediction made by the model.

4.4 Privacy and Communication Efficiency Analysis

Effect of Differential Privacy on prediction accuracy was assessed with varying levels of privacy budget (ε). This is shown in the following table, Table 5.

Table 5. Effect of privacy budget on predictive performance

Privacy Budget (\varepsilon)	Accuracy	F1-score	ROC-AUC	Relative Privacy Level
0.5	0.817	0.672	0.852	Very high
1.0	0.838	0.704	0.875	High
2.0	0.854	0.730	0.891	Moderate to high
4.0	0.857	0.735	0.895	Moderate
8.0	0.858	0.737	0.896	Relatively low

Table 5 illustrates that increased privacy budget led to decreased predictive accuracy since noise was added during the training process. The most efficient setting for the privacy budget ε was equal to 2.0, where privacy was balanced with the accuracy of 85.4%.

The correlation between the privacy budget and accuracy can be viewed in Figure 4.

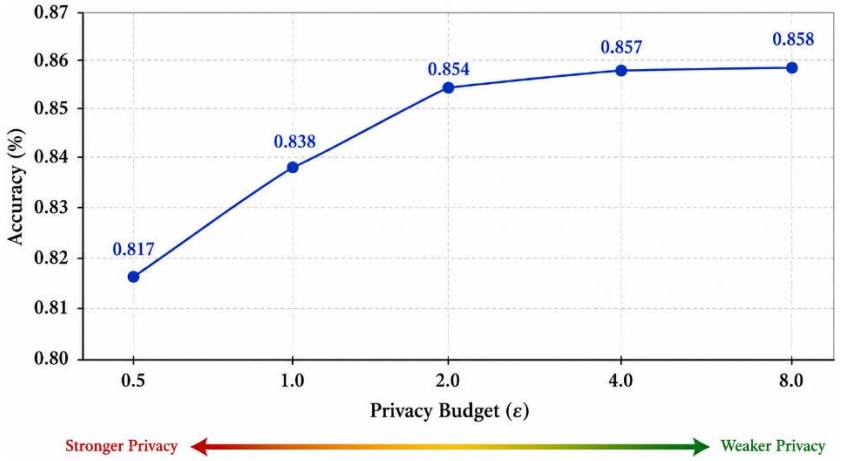

Figure 4. Accuracy under different Differential Privacy budgets

Figure 4, model performance had a sharp improvement until ε = 2.0, beyond which there was almost no improvement. Hence, ε = 2.0 was determined as the best operating point for the proposed framework.

Furthermore, the proposed framework also decreased the membership inference success ratio to 0.51 from 0.64 and the update reconstruction similarity to 0.17 from 0.69, making the framework more robust against privacy threats. Though Secure Aggregation increased the communication overhead and training time, the extra overhead was negligible.

4.5 Comparative and Statistical Discussion

The experimental results show that the integration of explainability with federated learning is possible while keeping up with predictive performance in the same class. The framework under discussion consists of Federated Learning, Differential Privacy, Secure Aggregation, and SHAP algorithms in a single structure.

The statistical significance was tested based on Friedman testing with $\chi^2 = 14.52$ ($p = 0.0023$). Thus, there is a significant difference in predictive performance among compared models. Using post-hoc Wilcoxon signed-rank test with Holm correction, it was established that:

- The framework suggested significantly outperformed the traditional DP-FL method.
- The difference between the framework suggested and Standard FL was insignificant at $p < 0.05$.

4.6 Practical Implications

Proposed solution provides an approach to perform collaborative predictive analytics while avoiding data exposure, thus being applicable for healthcare, financial industry, cyber security, industrial IoT, and many other cloud computing use cases.

Main advantages of the proposed solution are:

- Data privacy protection using Differential Privacy and Secure Aggregation.
- Increased model transparency by using SHAP explanation method.
- Maintaining data locality and supporting regulatory compliance.

- Secure collaboration between multiple parties.

In addition, practical implementation will require secure communication protocol, client authentication and model fairness monitoring.

4.7 Limitations and Future Directions

The suggested architecture was tested based on one benchmarking dataset with federated clients simulated, which might not be sufficient to reflect large-scale environments. Moreover, security assessment is concerned only about Differential Privacy and Secure Aggregation but does not take into account the model poisoning or Byzantine failures attacks.

Future work should address:

- Large-scale real-world datasets and heterogeneous cloud-edge environments.
- Adaptive Differential Privacy, homomorphic encryption, and resilient aggregation.
- Fairness-aware federated learning and efficient communication optimization.

4.8 Overall Discussion

The proposed Explainable Federated Learning system strikes an effective balance between predictive ability, privacy protection, efficient communication, and model interpretability. Its accuracy was 85.4%, F1-score 0.730, and ROC-AUC 0.891 with performance levels that were comparable with Standard Federated Learning while offering better privacy protection.

The use of Differential Privacy, Secure Aggregation, and SHAP increased data privacy and model interpretability with just a minor increase in communication costs. In summary, the results confirm the viability of the proposed approach as an efficient privacy-preserving method for predictive modeling.

5. Conclusion

The current research introduced an explainable federated machine learning approach for privacy-preserving predictive analytics in cloud-based big data settings through a combination of Federated Learning, Differential Privacy, Secure Aggregation, and explainability based on SHAP method. The framework allows for joint model training while ensuring data privacy and boosting model transparency. The experimental evaluation on Adult Income data set resulted in accuracy 85.4%, F1-Score 0.730, and ROC-AUC 0.891 without sacrificing the quality compared to Standard Federated Learning. Moreover, the framework was shown to have stability of convergence under different distribution of data sets, as well as SHAP allowed providing meaningful explanations on both global and local levels. Thus, the research outcomes support the possibility of effective inclusion of the concepts of privacy preservation and explainability in federated learning while maintaining the predictive accuracy of the model. The proposed framework is a viable solution for the application of trustworthy predictive analytics in cloud-based healthcare, finance, cybersecurity, IoT settings, and other fields. Further work will concentrate on the use of large-scale real-world data sets, as well as incorporating adaptive privacy, advanced security techniques, and communication-efficient federated learning approaches.

References

1. Abadi, M., Chu, A., Goodfellow, I., McMahan, H. B., Mironov, I., Talwar, K., & Zhang, L. (2016, October). Deep learning with differential privacy. In *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security* (pp. 308-318).
2. Alkadi, O., Moustafa, N., Turnbull, B., & Choo, K. K. R. (2020). A deep blockchain framework-enabled collaborative intrusion detection for protecting IoT and cloud networks. *IEEE Internet of Things Journal*, *8*(12), 9463-9472.
3. Antunes, N., Balby, L., Figueiredo, F., Lourenco, N., Meira, W., & Santos, W. (2018, June). Fairness and transparency of machine learning for trustworthy cloud services. In *2018 48th annual IEEE/IFIP international conference on dependable systems and networks workshops (DSN-W)* (pp. 188-193). IEEE.
4. Atalor, S. I. (2019). Federated learning architectures for predicting adverse drug events in oncology without compromising patient privacy. *IRE Journals*, *2*(12), 1-15.
5. Basilakis, J. (2020). Cloud-based homomorphic encryption for privacy-preserving machine learning in clinical decision support. *Western Sydney University: Sydney, Australia*.
6. Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., ... & Seth, K. (2017, October). Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (pp. 1175-1191).
7. Hartley, A. J., McKenzie, D. R., & Narayanan, P. K. (2020). Privacy-Preserving Machine Learning and Secure Data Sharing for Big Data Analytics. *The Artificial Intelligence Journal*, *1*(3).
8. Hashem, I. A. T., Yaqoob, I., Anuar, N. B., Mokhtar, S., Gani, A., & Khan, S. U. (2015). The rise of "big data" on cloud computing: Review and open research issues. *Information systems*, *47*, 98-115.
9. Jani, P. (2020). Privacy-Preserving AI in Provider Portals: Leveraging Federated Learning in Compliance with HIPAA. *The Distributed Learning and Broad Applications in Scientific Research*, *6*, 1116-1145.
10. Oloke, K. (2019). Architecting autonomous financial decision engines through federated learning and hybrid cloud frameworks. *Int J Appl Res*, *5*(6), 500-510.
11. Polamarasetti, A. (2019). Advanced Machine Learning Models for Predictive Analytics in Cloud-Based Systems. *International Journal of Advanced Engineering Technologies and Innovations*, *1*(3), 73-102.
12. Ramanathan, A. (2018). The Impact Of Federated Learning On Preserving Data Privacy In Cloud-Based AI Models. *International Journal of Scientific Research & Engineering Trends*, *4*(2).
13. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).
14. Shah, H. (2017). Deep learning in cloud environments: innovations in AI and cybersecurity challenges. *Revista Espanola de Documentacion Cientifica*, *11*(1), 146-160.
15. Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated machine learning: Concept and applications. *ACM Transactions on Intelligent Systems and Technology (TIST)*, *10*(2), 1-19.