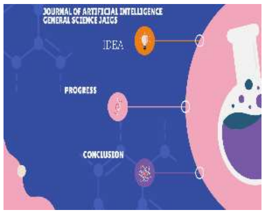
Vol.3,Issue 01, March 2024
Journal of Artificial Intelligence General Science JAIGS



Cultivating Privacy in Collaborative Data Sharing through Auto-encoder Latent Space Embeddings

Vinayak Raja[1], Bhuvi chopra[2]

[1]Software engineer, Meta

[2]Product manager, Google

DOI: https://doi.org/10.60087/jaigs.vol03.issue01.p283

ARTICLEINFO
Article History:
Received:
01.03.2024
Accepted:
15.03.2024
Online: 30.03.2024
Keyword:privacy-preserving machine learning, collaborative data sharing, confidentiality, sensitive information, machine learning process, model training

ABSTRACT

Ensuring privacy in machine learning through collaborative data sharing is imperative for organizations aiming to leverage collective data without compromising confidentiality. This becomes particularly crucial when sensitive information must be safeguarded throughout the entire machine learning process, spanning from model training to inference. This paper introduces a novel framework employing Representation Learning through autoencoders to produce privacy-preserving embedded data. Consequently, organizations can share these representations, fostering improved performance of machine learning models in scenarios involving multiple data sources for a unified predictive task downstream.

Introduction

Collaborative data sharing strategies within Artificial Intelligence (AI) frameworks have become commonplace among organizations aiming to enhance prediction model performance and fortify data reliability, thus gaining competitive advantages [1]. However, in real-world scenarios, data sharing processes may encounter obstacles due to privacy policies or intellectual property regulations, notwithstanding the safety of communication infrastructures between peers [2]. Consider, for instance, two companies each possessing distinct sets of variables pertaining to the same group of users. While these peers could potentially leverage complementary information from one another to predict a response variable and inform decision-making processes, the presence of sensitive user data on both ends often precludes data sharing, thereby hindering potential model performance improvements. Consequently, devising strategies to facilitate information sharing without compromising predictive capabilities is essential for ML model development within such organizations.

In response to this challenge, academia and private entities have devised various solutions and frameworks facilitating data sharing through technological and machine learning approaches. Many of these approaches rely on cryptographic techniques (e.g., homomorphic encryption [3, 4]), data perturbation methods (e.g., differential privacy [5], local differential privacy [6], dimensionality reduction [7]), and distributed architectures (e.g., federated learning) [8, 9]. Notably, these solutions primarily focus on preserving privacy during communication, altering individual observation patterns, and often entail high maintenance requirements. Consequently, our focus lies in constructing a privacy-preserving framework utilizing recent advancements in deep learning models to enable collaborative peers to share data without compromising the predictive power of the original features.

This paper introduces an innovative framework harnessing representation learning through auto-encoders to generate privacy-preserving embeddings of sensitive information, facilitating collaboration among multiple data sources in the development of trustworthy machine learning models. Additionally, we apply the proposed framework to three distinct scenarios to assess its practical applicability. The structure of this work can be summarized as follows: Firstly, we survey existing case studies on privacy-preserving machine learning to discern their limitations and identify avenues for improvement. Subsequently, we delve into the proposed methodology, outlining the key stages of the general process for method validation. Following this, we introduce and elaborate on the selected case studies, presenting their respective outcomes. Finally, we offer conclusions and insights for future research directions.

Background

Given our proposed method's novelty in privacy-preserving machine learning for collaborative model development utilizing deep-learning autoencoders, it's essential to review traditional privacy-preserving approaches and representation learning to ascertain their potential integration as a solution to our addressed problem.

Privacy-preserving Machine Learning

Privacy-preserving machine learning resides within the AI ecosystem and aims to reconcile data ownership rights with the advantages of employing machine learning models with said data. These models can safeguard either the data itself or the developed model [10]. As our strategy aims to facilitate secure data sharing among peers, we will delve into data-oriented privacy guarantee applications.

Three primary traditional approaches exist for addressing privacy-preserving ML. Firstly, Encryption-Based Privacy-Preserving methods transform the feature set into ciphertext, thwarting data leakage between peers [11]. Despite the security merits, like those provided by Homomorphic Encryption, these solutions encounter limitations in real-world scenarios due to technological requirements. Architecture-based approaches, such as Federated Learning, create decentralized model development pipelines with data distributed across multiple peers, suitable when contributors share common information but inadequate when peers possess disparate datasets [12].

The third traditional approach involves perturbing original features, with differential privacy being a prevalent strategy leveraging data distribution to obscure individual observation values [13]. However, this method may introduce substantial noise, diminishing data utility. Alternatively, dimensionality reduction techniques preserve variance while obfuscating original features.

Principal Component Analysis is one such technique that generates a representation vector of the data, which can then be utilized in downstream models of interest. Nonetheless, linear transformations for dimensionality reduction may overlook certain data relationships. Nguyen et al. [9] employed representation learning in their work "AutoGAN-based Dimension Reduction for Privacy Preservation" to encode image privacy and integrate the embeddings into anomaly detection.

Representation Learning

Representation Learning, a domain within Deep Learning, enables algorithms to automatically learn representations of input data. Widely applied in diverse data types like images, speech, or text, its uses encompass anomaly detection, pattern recognition, and dimensionality reduction. Autoencoders, specific neural networks, encode input data and reconstruct the original dataset with minimal error [14]. Consisting of encoder and decoder structures, they are linked via the latent space representation, an embedded vector of the original data [15].

Representation Learning serves as a principal dimensionality reduction strategy, structuring a supervised machine learning model that seeks optimal nonlinear feature combinations representing the original data [16]. Consequently, the latent space representation forms an abstract multidimensional space encoding the original feature set while preserving proximity between similar observations.

Privacy-preserving and representation learning intersect as complementary research areas. Representation learning offers a deep learning strategy for encoding data while retaining core information and observation representation. This combination enables the achievement of our primary objective: fostering trustful data sharing among collaborative peers for machine learning model development.

Privacy-Preserving Machine Learning for Collaborative Data Sharing via Auto-encoder Latent Space Embeddings

Our proposed method advocates for leveraging representation learning to embed data as a privacy-preserving strategy, enabling multiple peers to share data without compromising predictive power. Demonstrating the efficacy of this approach in facilitating trustworthy data sharing while maintaining predictive model performance expands the horizons of AI collaboration practices for organizations. Illustrated in Figure 1, our framework accommodates multiple data sources eager to contribute to one another by sharing data while upholding the privacy of sensitive information. Notably, collaboration revolves around enhancing the feature set of an observation identified by a standard ID.

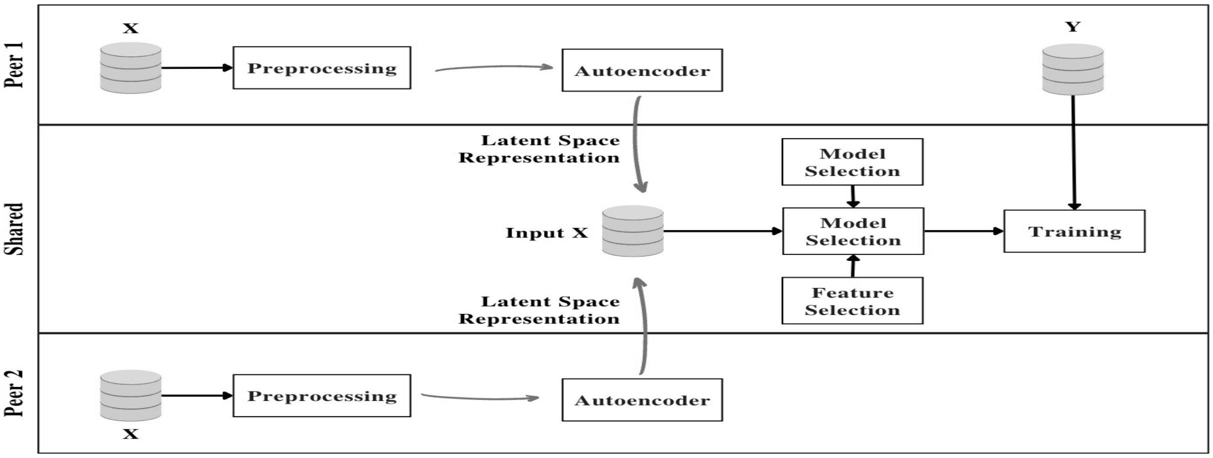
In conventional data-sharing pipelines for training collaborative machine learning models, both ends typically contribute by sharing raw datasets, which are then merged using a standard ID across all observations. Following appropriate data preprocessing, one or both peers may train a machine-learning model using the augmented feature set, thereby potentially enhancing predictive power over the target variable. In contrast to this conventional approach, our method introduces an additional step preceding data merging. Here, peers generate a latent space representation of their original data, effectively producing an obfuscated dataset primed for sharing. Consequently, peers integrate these data representations to jointly train a supervised downstream task aimed at predicting the same target variable without sacrificing predictive power, with the aim of improving overall performance through data sharing.

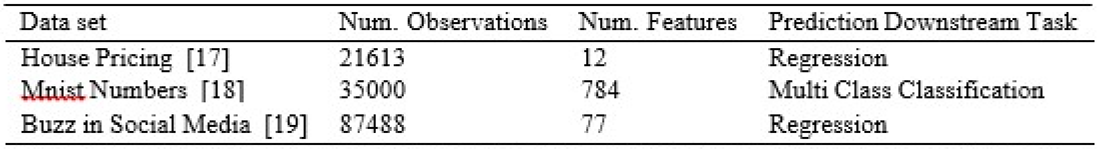
Initially scoped for two peers, our method holds potential for scalability to accommodate more collaborators. Furthermore, we assume that involved peers will share representations of the entire feature set. However, in real-world applications, both peers may not necessarily need to implement privacy-preserving strategies.

Evaluation

Data Sets

To assess the effectiveness of our proposed framework and to simulate real-life scenarios, we curated three public datasets: House Pricing [17], Mnist Numbers [18], and Buzz in Social Media [19]. These datasets were chosen to evaluate the framework's performance across various characteristics and to encompass potential scenarios encountered in practical applications, ensuring the generality of our framework. Consequently, we encompass both regression and classification prediction tasks, thus ensuring comprehensive testing. Furthermore, we deliberately considered variations in feature dimensions and types to encompass diverse downstream tasks, dimensions, and feature types, thereby evaluating the robustness and scalability of our solution.





Data set	Num. Observations	Num. Features	Prediction Downstream Task
House Pricing [17]	21613	12	Regression
Mnist Numbers [18]	35000	784	Multi Class Classification
Buzz in Social Media [19]	87488	77	Regression

Table 1: This table shows the widely-known benchmark data sets that we used to test our privacy-preserving framework. The number of total samples, number of features, and Machine Learning tasks performed over each dataset validates are correspondingly reported.

Experiments

We established a baseline model devoid of privacy-preserving strategies and four distinct privacy-preserving scenarios to ensure dependable and comparable results, as delineated in Section 3.

Scenario 0 | Baseline: This scenario entails training a predictive model for the downstream task using a single data source, denoted as the raw dataset herein. Employing a traditional supervised machine learning model, we incorporate randomized search as a hyperparameter tuning strategy. The performance of this baseline model serves as the benchmark against which subsequent scenarios are evaluated.

Scenario 1 | Representation Learning with a Single Shared Autoencoder: Here, we preprocess a unified dataset to derive a single representation vector, utilizing it to train a predictive model for the downstream task. This scenario evaluates the predictive performance facilitated by an accurate representation.

Scenario 2 | Representation Learning with Individual Autoencoders: This scenario simulates two peers by partitioning the initial dataset and individually preprocessing them to obtain a representation vector for each source. Subsequently, to train the predictive model for the downstream task, we combine these vectors using the observations' IDs.

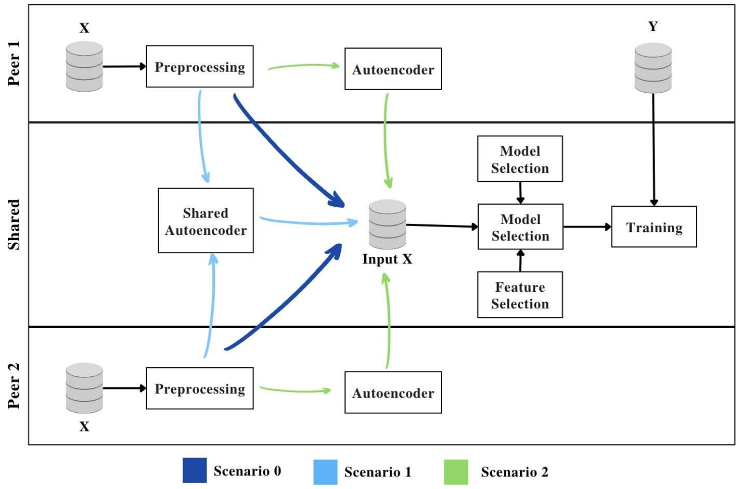


Figure 2: General Autoencoder Structure for Training Phase and Explored Scenarios

We devised the following two scenarios considering the hypothesis that the autoencoder could adopt a non-naive approach during latent space representation estimation, potentially enhancing downstream task metrics.

Scenario 3 | Representation Learning with Shared Autoencoder Non-naive Approach: This scenario assesses the downstream task's impact when the autoencoder model incorporates the predictive variable in the principal model. We modify the autoencoder, transforming it into a multitask neural network that predicts representation performance and the objective variable simultaneously. Given that both peers have access to the predictive variable, we replicate the second scenario but update the encoder stage accordingly.

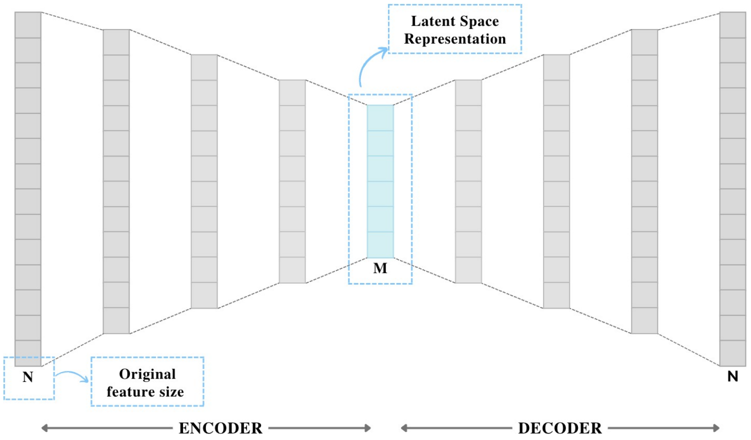
Scenario 4 | Representation Learning with Individual Autoencoder Non-naive Approach: Similarly, this scenario evaluates the downstream task's impact when the autoencoder model integrates the predictive variable in the principal model. We adapt the autoencoder into a multitask neural network predicting representation performance and the objective variable concurrently. Following the same rationale as Scenario 3, we update the encoder stage of the process as described above.

Experimental Setup

Autoencoder Setup: We employed the Tensorflow framework to construct the Autoencoder Neural Network. Maintaining consistency across all experiments, we standardized the network's structure to draw conclusions regarding the general framework rather than the network's complexity. The Autoencoder model comprises two main components: the encoder and decoder, each composed of four layers, with the latent space representation layer serving as the connection between them. Figure 3 depicts the overall Autoencoder structure, where N represents the original feature size, and M represents the embedding size.

Encoder: The encoder consists of four layers. The input layer accommodates neurons equivalent to the features of the original dataset (N). Subsequently, three hidden layers conduct nonlinear transformations with dimensionality reduction between each layer, culminating in the final layer—the latent space representation. We determined the dimensions as N for the input layer, [128, 64, 40] for the hidden layers, and M for the latent space representation size.

Decoder: Mirroring the encoder, the decoder initiates its structure from the latent space representation size (M). Subsequently, three layers replicate the encoder's design in reverse to attain the final layer, corresponding to the original feature size (N). We selected dimensions as M for the input layer, [128, 64, 40] for the hidden layers, and N for the final output size.



Additional Considerations: Due to the nature of the input data utilized in our framework, we employed ReLU as the activation function for every layer. Furthermore, considering the scaling of input data, we opted for Mean Absolute Error as the loss function for the Autoencoder. Lastly, we adopted an Adam Optimizer with a learning rate set to 0.0001.

Results

Encoding Performance with Shared Autoencoder

We trained the autoencoder model using complete datasets to evaluate various scenarios as previously discussed. Representation accuracy was assessed using the autoencoder model's loss function, supplemented by the metric of average correctly estimated observations per feature, defined as observations with less than 5% Mean Absolute Percentage Error (MAPE). For the House Pricing dataset, the representation error is 5%, with an average estimated observations per feature rate of 98%. In the case of Mnist, the representation error is 7%, with an average estimated observations per feature rate of 96%. Lastly, for Buzz in Social Media, the representation error is 6%, with an average estimated observations per feature rate of 97%.

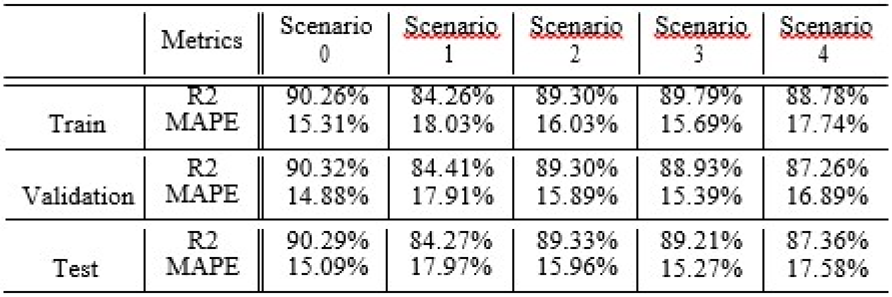
Individual Autoencoders

We trained a separate autoencoder model for each simulated data source to explore the aforementioned scenarios. Similar to the shared autoencoder approach, representation accuracy was assessed using the autoencoder model's loss function, alongside the metric of average correctly estimated observations per feature. For the House Pricing dataset, the average representation error is 11%, with an average estimated observations per feature rate of 86%. Regarding Mnist, the average representation error is 9%, with an average estimated observations per feature rate of 94%. For Buzz in Social Media, the representation error is 8%, with an average estimated observations per feature rate of 94%.

Representation Learning - House Pricing Framework

In this experiment, the downstream task involves estimating the price of a house in USD based on certain characteristics. To predict this task, we utilize an XGBoostRegressor model. Moreover, we incorporate hyperparameter tuning using Randomized Search Cross Validation, considering the following parameters: learning rate, max depth, min child weight, gamma, and colsamplebytree.

Table 3: House Pricing Scenarios Metrics



	Metrics	Scenario 0	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Train	R2 MAPE	90.26% 15.31%	84.26% 18.03%	89.30% 16.03%	89.79% 15.69%	88.78% 17.74%
Validation	R2 MAPE	90.32% 14.88%	84.41% 17.91%	89.30% 15.89%	88.93% 15.39%	87.26% 16.89%
Test	R2 MAPE	90.29% 15.09%	84.27% 17.97%	89.33% 15.96%	89.21% 15.27%	87.36% 17.58%

The findings indicate that despite the dimensionality augmentation necessitated by the dataset's limited feature count, the latent space representation exhibits minimal loss in predictive power. Furthermore, the downstream model retains its ability to accurately predict the objective variable. In scenarios where the principal dataset simulates two data sources, employing both latent space representations yields performance levels comparable to those observed in Scenario 1.

Mnist Numbers

In this experiment, the downstream task involves predicting which number, ranging from 0 to 9, corresponds to an image. To accomplish this task, we utilize a Multinomial Logistic Regression model. Notably, for this specific case, we preprocess the images to convert them into tabular data, facilitating their use in the model.

Conclusions & Future Work

In this paper, we introduce an alternative solution to traditional privacy-preserving approaches in machine learning, demonstrating that with an accurate representation learning model, peers can share an embedded dataset that maintains the patterns and behavior of the original observations. Transitioning from original features to a latent space representation does not significantly degrade the performance of downstream tasks. In our experiments, model results experienced a decrease of less than 10 percentage points, with representation errors ranging from 5% to 11%. Consequently, peers or organizations can collaborate without compromising organizational privacy policies or infringing upon potential clients' privacy concerns.

For future considerations, each data source should develop a customized autoencoder neural network implementation to enhance representation performance and ensure alignment with dataset requirements. Additionally, despite assuming that dimensionality reduction preserves data privacy, we aim to develop metrics for quantifying the privacy level of each dataset. These metrics will consider the complexity of the embedding and the difficulty for potential attackers to decode the original dataset. Finally, we intend to validate this framework using organizational data from various sources to draw conclusions regarding real-life scenarios.

References List

[1]. Shuford, J. (2023). Contribution of Artificial Intelligence in Improving Accessibility for Individuals with Disabilities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 421-433.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p433

[2]. Chentha, A. K., Sreeja, T. M., Hanno, R., Purushotham, S. M. A., &Gandrapu, B. B. (2013). A Review of the Association between Obesity and Depression. Int J Biol Med Res, 4(3), 3520-3522.

[3]. Gadde, S. S., &Kalli, V. D. R. (2020). Descriptive analysis of machine learning and its application in healthcare. *Int J Comp Sci Trends Technol*, *8*(2), 189-196.

[4]. Atacho, C. N. P. (2023). A Community-Based Approach to Flood Vulnerability Assessment: The Case of El Cardón Sector. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 434-482. DOI:https://doi.org/10.60087/jklst.vol2.n2.p482

[5]. jimmy, fnu. (2023). Understanding Ransomware Attacks: Trends and Prevention Strategies. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(1), 180-210. https://doi.org/10.60087/jklst.vol2.n1.p214

[6]. Bayani, S. V., Prakash, S., &Malaiyappan, J. N. A. (2023). Unifying Assurance A Framework for Ensuring Cloud Compliance in AIML Deployment. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(3), 457-472.DOI: https://doi.org/10.60087/jklst.vol2.n3.p472

[7]. Bayani, S. V., Prakash, S., &Shanmugam, L. (2023). Data Guardianship: Safeguarding Compliance in AI/ML Cloud Ecosystems. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(3), 436-456.

DOI: https://doi.org/10.60087/jklst.vol2.n3.p456

[8]. Karamthulla, M. J., Malaiyappan, J. N. A., & Prakash, S. (2023). AI-powered Self-healing Systems for Fault Tolerant Platform Engineering: Case Studies and Challenges. *Journal of*

Knowledge Learning and Science Technology ISSN: 2959-6386 (online), *2*(2), 327-338. DOI: https://doi.org/10.60087/jklst.vol2.n2.p338

[9]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2023). Unlocking Insights: AI/ML Applications in Regulatory Reporting for US Banks. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 177-184.DOI: https://doi.org/10.60087/jklst.vol1.n1.p184

[10]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2023). From Burden to Advantage: Leveraging AI/ML for Regulatory Reporting in US Banking. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 167-176. DOI: https://doi.org/10.60087/jklst.vol1.n1.p176

[11]. Prakash, S., Venkatasubbu, S., &Konidena, B. K. (2022). Streamlining Regulatory Reporting in US Banking: A Deep Dive into AI/ML Solutions. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *1*(1), 148-166.DOI: https://doi.org/10.60087/jklst.vol1.n1.p166

[12]. Tomar, M., &Jeyaraman, J. (2023). Reference Data Management: A Cornerstone of Financial Data Integrity. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(1), 137-144.DOI: https://doi.org/10.60087/jklst.vol2.n1.p144

[13]. Tomar, M., &Periyasamy, V. (2023). The Role of Reference Data in Financial Data Analysis: Challenges and Opportunities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 1(1), 90-99.

DOI: https://doi.org/10.60087/jklst.vol1.n1.p99

[14]. Tomar, M., &Periyasamy, V. (2023). Leveraging Advanced Analytics for Reference Data Analysis in Finance. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(1), 128-136.

DOI: https://doi.org/10.60087/jklst.vol2.n1.p136

[15]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). Unlocking Sales Potential: How AI Revolutionizes Marketing Strategies. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 231-250.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p250

[16]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). Optimizing Sales Funnel Efficiency: Deep Learning Techniques for Lead Scoring. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 261-274.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p274

[17]. Shanmugam, L., Tillu, R., &Tomar, M. (2023). Federated Learning Architecture: Design, Implementation, and Challenges in Distributed AI Systems. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 371-384.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p384

[18]. Sharma, K. K., Tomar, M., &Tadimarri, A. (2023). AI-driven Marketing: Transforming Sales Processes for Success in the Digital Age. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online)*, *2*(2), 250-260.

DOI: https://doi.org/10.60087/jklst.vol2.n2.p260

[19]. Gadde, S. S., &Kalli, V. D. (2021). The Resemblance of Library and Information Science with Medical Science. International Journal for Research in Applied Science & Engineering Technology, 11(9), 323-327.

[20]. Gadde, S. S., &Kalli, V. D. R. (2020). Technology Engineering for Medical Devices-A Lean Manufacturing Plant Viewpoint. Technology, 9(4).

[21]. Gadde, S. S., &Kalli, V. D. R. (2020). Medical Device Qualification Use. International Journal of Advanced Research in Computer and Communication Engineering, 9(4), 50-55.

[22]. Gadde, S. S., &Kalli, V. D. R. (2020). Artificial Intelligence To Detect Heart Rate Variability. *International Journal of Engineering Trends and Applications*, *7*(3), 6-10.

[23]. Chentha, A. K., Sreeja, T. M., Hanno, R., Purushotham, S. M. A., &Gandrapu, B. B. (2013). A Review of the Association between Obesity and Depression. *Int J Biol Med Res*, *4*(3), 3520-3522.

[24]. Tao, Y. (2022). Algorithm-architecture co-design for domain-specific accelerators in communication and artificial intelligence (Doctoral dissertation).

https://deepblue.lib.umich.edu/handle/2027.42/172593

[25]. Tao, Y., Cho, S. G., & Zhang, Z. (2020). A configurable successive-cancellation list polar decoder using split-tree architecture. IEEE Journal of Solid-State Circuits, 56(2), 612-623.

DOI: https://doi.org/10.1109/JSSC.2020.3005763

[26]. Tao, Y., & Choi, C. (2022, May). High-Throughput Split-Tree Architecture for Nonbinary SCL Polar Decoder. In 2022 IEEE International Symposium on Circuits and Systems (ISCAS) (pp. 2057-2061). IEEE.

DOI: https://doi.org/10.1109/ISCAS48785.2022.9937445

[27]. Tao, Y. (2022). Algorithm-architecture co-design for domain-specific accelerators in communication and artificial intelligence (Doctoral dissertation).

https://deepblue.lib.umich.edu/handle/2027.42/172593

[28]. Mahalingam, H., VelupillaiMeikandan, P., Thenmozhi, K., Moria, K. M., Lakshmi, C., Chidambaram, N., &Amirtharajan, R. (2023). Neural attractor-based adaptive key generator with DNA-coded security and privacy framework for multimedia data in cloud environments. Mathematics, 11(8), 1769.

https://doi.org/10.3390/math11081769

[29]. Padmapriya, V. M., Thenmozhi, K., Praveenkumar, P., &Amirtharajan, R. (2020). ECC joins first time with SC-FDMA for Mission “security”. Multimedia Tools and Applications, 79(25), 17945-17967.

DOI https://doi.org/10.1007/s11042-020-08610-5

[30]. Padmapriya, V. M. (2018). Image transmission in 4g lte using dwt based sc-fdma system. Biomedical & Pharmacology Journal, 11(3), 1633.

DOI :https://dx.doi.org/10.13005/bpj/1531

[31]. Padmapriya, V. M., Priyanka, M., Shruthy, K. S., Shanmukh, S., Thenmozhi, K., &Amirtharajan, R. (2019, March). Chaos aided audio secure communication over SC-FDMA system. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.

https://doi.org/10.1109/ViTECoN.2019.8899413

[31]. Padmapriya, V. M., Thenmozhi, K., Praveenkumar, P., &Amirtharajan, R. (2022). Misconstrued voice on SC-FDMA for secured comprehension-a cooperative influence of DWT and ECC. Multimedia Tools and Applications, 81(5), 7201-7217.

DOI https://doi.org/10.1007/s11042-022-11996-z

[32]. Padmapriya, V. M., Sowmya, B., Sumanjali, M., &Jayapalan, A. (2019, March). Chaotic Encryption based secure Transmission. In 2019 International Conference on Vision Towards Emerging Trends in Communication and Networking (ViTECoN) (pp. 1-5). IEEE.

DOI https://doi.org/10.1109/ViTECoN.2019.8899588

[33]. Sowmya, B., Padmapriya, V. M., Sivaraman, R., Rengarajan, A., Rajagopalan, S., &Upadhyay, H. N. (2021). Design and Implementation of Chao-Cryptic Architecture on FPGA for Secure Audio Communication. In Emerging Technologies in Data Mining and Information Security: Proceedings of IEMIS 2020, Volume 3 (pp. 135-144). Springer Singapore

https://link.springer.com/chapter/10.1007/978-981-15-9774-9_13

[34]. Padmapriya, V. M., Thenmozhi, K., Avila, J., Amirtharajan, R., &Praveenkumar, P. (2020). Real Time Authenticated Spectrum Access and Encrypted Image Transmission via Cloud Enabled Fusion centre. Wireless Personal Communications, 115, 2127-2148.

DOI https://doi.org/10.1007/s11277-020-07674-8

[35].Thakur, A., & Thakur, G. K. (2024). Developing GANs for Synthetic Medical Imaging Data: Enhancing Training and Research. Int. J. Adv. Multidiscip. Res, 11(1), 70-82.

DOI: http://dx.doi.org/10.22192/ijamr.2024.11.01.009

[36]. Shuford, J. (2023). Contribution of Artificial Intelligence in Improving Accessibility for Individuals with Disabilities. Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), 2(2), 421-433.DOI: https://doi.org/10.60087/jklst.vol2.n2.p433

[37]. Schwartz, E. A., Bravo, J. P., Ahsan, M., Macias, L. A., McCafferty, C. L., Dangerfield, T. L., ...& Taylor, D. W. (2024). RNA targeting and cleavage by the type III-Dv CRISPR effector complex. Nature Communications, 15(1), 3324.

https://www.nature.com/articles/s41467-024-47506-y#Abs1

[38]. Saha, A., Ahsan, M., Arantes, P. R., Schmitz, M., Chanez, C., Jinek, M., & Palermo, G. (2024). An alpha-helical lid guides the target DNA toward catalysis in CRISPR-Cas12a. Nature Communications, 15(1), 1473.https://www.nature.com/articles/s41467-024-45762-6

[39]. Nierzwicki, Ł., Ahsan, M., & Palermo, G. (2023). The electronic structure of genome editors from the first principles. Electronic Structure, 5(1), 014003.DOI https://doi.org/10.1088/2516-1075/acb410

[40]. Bali, S. D., Ahsan, M., &Revanasiddappa, P. D. (2023). Structural Insights into the Antiparallel G-Quadruplex in the Presence of K+ and Mg2+ Ions. The Journal of Physical Chemistry B, 127(7), 1499-1512.https://doi.org/10.1021/acs.jpcb.2c05128

[41]. Ahsan, M., Pindi, C., &Senapati, S. (2022). Mechanism of darunavir binding to monomeric HIV-1 protease: A step forward in the rational design of dimerization inhibitors. Physical Chemistry Chemical Physics, 24(11), 7107-7120.https://doi.org/10.1039/D2CP00024E

[42]. Ahsan, M., Pindi, C., &Senapati, S. (2021). Hydrogen bonding catalysis by water in epoxide ring opening reaction. Journal of Molecular Graphics and Modelling, 105, 107894.https://doi.org/10.1016/j.jmgm.2021.107894

[43]. Ahsan, M., Pindi, C., &Senapati, S. (2020). Electrostatics plays a crucial role in HIV-1 protease substrate binding, drugs fail to take advantage. Biochemistry, 59(36), 3316-3331.

https://doi.org/10.1021/acs.biochem.0c00341

[44]. Pindi, C., Chirasani, V. R., Rahman, M. H., Ahsan, M., Revanasiddappa, P. D., &Senapati, S. (2020). Molecular basis of differential stability and temperature sensitivity of ZIKA versus dengue virus protein shells. Scientific Reports, 10(1), 8411.https://doi.org/10.1038/s41598-020-65288-3

[45]. Ahsan, M., &Senapati, S. (2019). Water plays a cocatalytic role in epoxide ring opening reaction in aspartate proteases: a QM/MM study. *The Journal of Physical Chemistry B*, *123*(38), 7955-7964.

https://doi.org/10.1021/acs.jpcb.9b04575

[46]. Dixit, S. M., Ahsan, M., &Senapati, S. (2019). Steering the lipid transfer to unravel the mechanism of cholesteryl ester transfer protein inhibition. Biochemistry, 58(36), 3789-3801.

https://doi.org/10.1021/acs.biochem.9b00301

[47]. Kommaraju, V., Gunasekaran, K., Li, K., Bansal, T., McCallum, A., Williams, I., & Istrate, A. M. (2020). Unsupervised pre-training for biomedical question answering. arXiv preprint arXiv:2009.12952.

[48]. Bansal, T., Gunasekaran, K., Wang, T., Munkhdalai, T., & McCallum, A. (2021). Diverse distributions of self-supervised tasks for meta-learning in NLP. arXiv preprint arXiv:2111.01322.

[49]. Gunasekaran, K., Tiwari, K., & Acharya, R. (2023, June). Utilizing deep learning for automated tuning of database management systems. In 2023 International Conference on Communications, Computing and Artificial Intelligence (CCCAI) (pp. 75-81). IEEE.

[50]. Gunasekaran, K. P. (2023, May). Ultra sharp: Study of single image super resolution using residual dense network. In 2023 IEEE 3rd International Conference on Computer Communication and Artificial Intelligence (CCAI) (pp. 261-266). IEEE.

[51]. Gillespie, A., Yirsaw, A., Gunasekaran, K. P., Smith, T. P., Bickhart, D. M., Turley, M., ... & Baldwin, C. L. (2021). Characterization of the domestic goat γδ T cell receptor gene loci and gene usage. Immunogenetics, 73, 187-201.

[52]. Yirsaw, A. W., Gillespie, A., Zhang, F., Smith, T. P., Bickhart, D. M., Gunasekaran, K. P., ... & Baldwin, C. L. (2022). Defining the caprine γδ T cell WC1 multigenic array and evaluation of its expressed sequences and gene structure conservation among goat breeds and relative to cattle. Immunogenetics, 74(3), 347-365.

[53]. Gunasekaran, K. P., Babrich, B. C., Shirodkar, S., & Hwang, H. (2023, August). Text2Time: Transformer-based Article Time Period Prediction. In 2023 IEEE 6th International Conference on Pattern Recognition and Artificial Intelligence (PRAI) (pp. 449-455). IEEE.

[54]. Gunasekaran, K., & Jaiman, N. (2023, August). Now you see me: Robust approach to partial occlusions. In 2023 IEEE 4th International Conference on Pattern Recognition and Machine Learning (PRML) (pp. 168-175). IEEE.

[55]. Gillespie, A., Yirsaw, A., Kim, S., Wilson, K., McLaughlin, J., Madigan, M., ... & Baldwin, C. L. (2021). Gene characterization and expression of the γδ T cell co-receptor WC1 in sheep. Developmental & Comparative Immunology, 116, 103911.

[56]. Gunasekaran, K. P. (2023). Leveraging object detection for the identification of lung cancer. *arXiv preprint arXiv:2305.15813*.

[57]. Gunasekaran, K. P. (2023). Exploring sentiment analysis techniques in natural language processing: A Comprehensive Review. arXiv preprint arXiv:2305.14842.

[58]. Hasan, M. R., Gazi, M. S., & Gurung, N. (2024). Explainable AI in Credit Card Fraud Detection: Interpretable Models and Transparent Decision-making for Enhanced Trust and Compliance in the USA. *Journal of Computer Science and Technology Studies*, *6*(2), 01-12.

[58]. Gazi, M. S., Hasan, M. R., Gurung, N., & Mitra, A. (2024). Ethical Considerations in AI-driven Dynamic Pricing in the USA: Balancing Profit Maximization with Consumer Fairness and Transparency. *Journal of Economics, Finance and Accounting Studies*, *6*(2), 100-111.

[59]. Sarkar, M., Puja, A. R., & Chowdhury, F. R. (2024). Optimizing Marketing Strategies with RFM Method and K-Means Clustering-Based AI Customer Segmentation Analysis. *Journal of Business and Management Studies*, *6*(2), 54-60.