Feature-level Knowledge Distillation Framework for Horizontal Dimensionality Reduction Process in Multitask Learning Using UNITI Strategy

Swapnali Sunil Gawali[1], Dhananjay B. Kshirsagar[2]

[1,2] Department of Computer Engineering, Sanjivani College of Engineering, Savitribai Phule Pune University, Kopargaon 423603, Maharashtra, India.

Corresponding Author: gswap13@gmail.com

HOW TO CITE:
Swapnali Sunil Gawali, Dhananjay B. Kshirsagar (2026). Feature-level Knowledge Distillation Framework for Horizontal Dimensionality Reduction Process in Multitask Learning Using UNITI Strategy. International Journal of Special Education, 41(11s), 1287-1305

ABSTRACT:
Horizontal data dimensionality reduction is a technique that simplifies data by reducing complexity along the task and output dimensions instead of the input dimension. This approach decreases redundancy between tasks, which enhances knowledge sharing and lowers computational demands. Multi-label learning and other machine learning approaches faces challenges from high-dimensional data. However, existing dimensionality reduction frameworks make it complex to apply semi-supervised methods, enforce instance correlation constraints, or implement feature selection and extraction mechanisms. A major challenge in multi-task learning is the loss of important discriminative information. To overcome this issue, the proposed model integrates dimensionality reduction in a multi-task learning approach. Initially, high-dimensional data are collected from the relevant databases and then applied to Iterative Linear Discriminant Analysis (LDA)-based dimensional reduction, which results in a discriminative and stable low-dimensional representation that maintains the essential features. Then, multitask learning is performed using Feature-level Knowledge Distillation-based Unifying Neural networks with Inter-dataset for Task Integration (FKD-UNITI) framework, which performs continuous refinement of shared features, boosts the tasks integration across datasets and achieves scalable learning over heterogeneous data sources. The performance of the suggested dimensional reduction model is evaluated among existing techniques to ensure its effectiveness using different performance metrics.

Keywords: Horizontal Dimensionality Reduction, Multitask Learning, Iterative Linear Discriminant Analysis, Feature Level Knowledge Distillation, Unifying Neural networks with Inter-dataset for Task Integration

1. Introduction

With rapid technological advancements, large-scale and higher-dimensional data are widely used in many real-world applications. The notable instance is DNA microarrays that provide large

quantity of details about gene sequences in a single experiment [6]. The higher-dimensional features can describe data samples more accurately but they also introduce various issues, such as overfitting [7], increased complexity [8], and longer training times [9]. Additionally, computational demands during model construction grow rapidly with dimensionality, which is a problem known as curse of dimensionality in data mining and machine learning [10]. As a result, there is a growing need for effective feature selection methods to manage high-dimensional data efficiently. In multi-label classification tasks, an object can belong to multiple categories simultaneously [11]. Due to its broad applicability, multi-label classification has become an active research area in machine learning. However, like other machine learning tasks, multi-label learning deals with high-dimensional feature spaces [12], which prolongs training time and reduces the efficiency of algorithms that rely on full-space distance calculations.

The goal of multi-label feature dimensionality reduction is to create a lower-dimensional representations that closely preserves original feature details. Recent multi-label dimensionality reduction methods are mainly based upon three strategies: Canonical Correlation Analysis (CCA), LDA, and HSIC approaches. The CCA [13] and Principal Component Analysis (PCA) are both linear dimensionality reduction methods. The key variation is that PCA considers only the interdependence within a single group of variables, while CCA examines the interdependence between two groups of variables. CCA identifies the optimal correlation subspace between feature set and the class labels by increasing covariance, which selects a set of representative variables to create a dimensionality reduction structure similar to PCA [14]. However, this method has limitations: it is inherently linear, the covariance maximization requires scale adjustments and additional constraints, and solution variables are projected directions, which makes it complex to incorporate extra learning constraints [15].

The major contribution of the paper is given below.

- In this work, an advanced multitask learning model is developed that utilizes horizontal dimensionality reduction to extract essential features from high-dimensional data across different datasets, which allows the model to share and integrate important information across tasks. This approach improves learning efficiency and minimizes the necessity for every task. It also provides a more scalable, flexible, and robust multitask learning system that generalizes well across diverse datasets and applications.
- To improve multitask learning across diverse datasets, FKD-UNITI is developed. This allows a single approach to learn multiple tasks simultaneously while effectively integrating knowledge from different sources. By using feature-level knowledge distillation, FKD-UNITI transfers important information from pre-trained teacher mechanism to the student mechanism, which ensures that task-specific details are preserved by promoting shared representations. This continuous refinement of shared features allows better task integration and generalization across heterogeneous datasets by making the approach robust for complex multitask learning scenarios.

Organization: The literature review of the prior methods is described in Section 2. Section 3 presents the architectural explanation of the developed model. The mathematical modeling of multi-task learning mechanism is described in Section 4. Section 5 provides the results and assessment of designed model. Section 6 provides the conclusion.

2. Literature Survey

A. Related Works

In 2025, Elbakary *et al.* [1] implemented a Low-Rank Adaptation (LoRA) to reduce the trainable parameters. The simulation outcomes showed that the method worked effectively when compared with prior models for fine-tuning of Large Language Models (LLMs). The results indicated that the proposed approach achieved lower local loss for every client by maintaining similar global

performance.

In 2024, Wang *et al.* [2] implemented an Evolutionary Computation (EC) model, which became popular because of its simplicity as well as wide applicability. However, the different model of EC showed varied abilities in handling different types of data, and they failed to share information effectively. To overcome these issues, the PSO-based Multi-task Evolutionary Learning (MEL) was introduced. This method used Multi-task Learning to enable information sharing among distinct feature selection applications, which improved learning capability.

In 2023, Luo *et al.* [3] employed a data reduction method based on PCA. This approach reduced the size of the data to make the use of subspace methods more efficient. At the same time, it preserved higher-dimensional resolution of recognized mode shapes, which allowed accurate and detailed modal analysis.

In 2024, Wang *et al.* [4] presented a one-shot TSOM along with a TSOM Multi-Task Learning approach. The method used a dispersion objective lens to scan the object along optical axis. The approach applied a Multi-task Learning to conduct reverse inversion. The results showed that the system significantly speed up TSOM image construction by using about 100 times fewer images and achieved faster computation. The proposed model also supported the use of TSOM for inline measurement in integrated circuits and extended to other areas of IC technology.

In 2024, Li *et al.* [5] designed a Semi-supervised Multi-label Dimensionality Reduction based on Minimizing Redundant Correlation (SMDR-MRC). First, a matrix factorization was used to transform the HSIC-based dimensionality reduction model into least squares task. To solve the problem, a new solution method known as S-FISTA was applied. Experimental results showed that the SMDR-MRC approach performed better than other approaches.

B. Problem statement

Data dimension in multi-task learning process helps to diminish the complexity of correlations among multiple tasks. By compressing the tasks, dimensional reduction process effectively transfers the tasks with limited computational complexity. Deep networks adapt well to large datasets, which is automatically suitable for different data types that makes the dimensionality reduction more flexible. Yet, these traditional methods produce several challenges that are given in the points below. The features and issues of classical dimensional reduction models are included in Table I.

- Traditional models limit their capability to capture intricate features present in high-dimensional data, which results in loss of discriminative information. This issue can be rectified by performing non-linear feature transformation before dimensionality reduction, thereby reducing information loss.
- Inter-task relationship is more difficult in traditional models that lead to unnecessary feature representation. This issue is overcome by integrating unified task-aware feature representation in the proposed model.
- Due to insufficient labeled data, traditional models undergo poor generalization during dimensionality reduction. Hence, the proposed model limited the labeled data by stabilizing the feature learning in inadequate data conditions.
- Most of the traditional models use a homogeneous data distribution process, which is not applied to multiple datasets. The proposed model overcomes this issue by performing robust dimensional reduction across heterogeneous domains.

TABLE I. FEATURES AND CHALLENGES OF CLASSICAL DIMENSIONAL REDUCTION METHODS

Author [citation]	Methodology	Features	Challenges
Elbakary *et al.* [1]	LoRA	• This method significantly reduces the memory and computational overhead, which supports multi-task learning by maintaining efficient storage capacity.	• This method faces difficulty in capturing deep inter-task features due to low-ranking module.
Wang *et al.* [2]	MEL	• This method boosts learning efficiency in large-scale datasets.	• This method incurs high computational cost with extensive features.
Luo *et al.* [3]	PCA	• This method reduces the redundancy and boosts the training efficiency.	• Subspace constraints reduce the performance during multi-task learning process.
Wang *et al.* [4]	TSOM MTL	• This method provides efficient data sharing suitable for multi-task learning, which reduces its data storage space.	• Insufficient feature compression negatively affects the performance of model.
Li *et al.* [5]	SMDR-MRC	• This method performs dimensionality reduction effectively, which preserves the discriminative features.	• This method is highly sensitive to noise and imbalanced distribution.

3. Structural illustration of developed multi-task learning framework based on dimensionality reduction

A. Detailed Elucidation of Proposed Horizontal Dimensionality Reduction Framework

Horizontal dimensionality reduction in multitask learning reduces the feature space by learning multiple related tasks simultaneously. It identifies a shared low-dimensional representation of features that can be used across tasks and enables the model to capture correlations among them. By projecting high-dimensional input features into a smaller latent space, horizontal reduction improves computational efficiency and reduces noise across tasks. However, conventional dimensionality reduction methods have several demerits when applied to multitask settings. They usually treat each task independently, ignoring inter-task relationships and shared information. Many traditional methods do not consider task labels while reducing dimensions, which leads to the loss of discriminative features important for prediction. Additionally, conventional approaches assume linear dependency in data, which limits their effectiveness for complex nonlinear patterns. They also struggle with scalability and adaptability when the number of tasks increases, resulting in reduced performance in large as well as real-world multitask learning scenarios. The designed approach overcomes the complexities of existing dimensionality reduction methods by integrating a structured dimensionality reduction and multitask learning framework. Initially, high-dimensional data is collected from databases that contain heterogeneous features and large volumes of information. To manage complexity, the data is processed using an Iterative LDA-based

dimensionality reduction approach. The iterative LDA repeatedly refines the projection of data into a low-dimensional space by considering class separability. This process preserves the most discriminative and informative features while eliminating redundant and noisy attributes that results in a stable and compact feature representation. The reduced feature set improves computational efficiency and enhances the quality of learning for subsequent stages. After obtaining the optimized low-dimensional representation, multitask learning is performed using the FKD-UNITI framework. In this stage, knowledge distillation is applied at the feature level to transfer useful representations across multiple related tasks and datasets. The unified neural network continuously refines shared feature representations while maintaining task-specific learning capabilities. Furthermore, the inter-dataset integration enables the model to combine information from different datasets that improves task correlation and knowledge sharing. Thus, the FKD-UNITI facilitates scalable multitask learning, enhances task integration, and achieves robust performance across heterogeneous data sources. Fig 1 shows the block diagram of the proposed approach for multi-task learning.

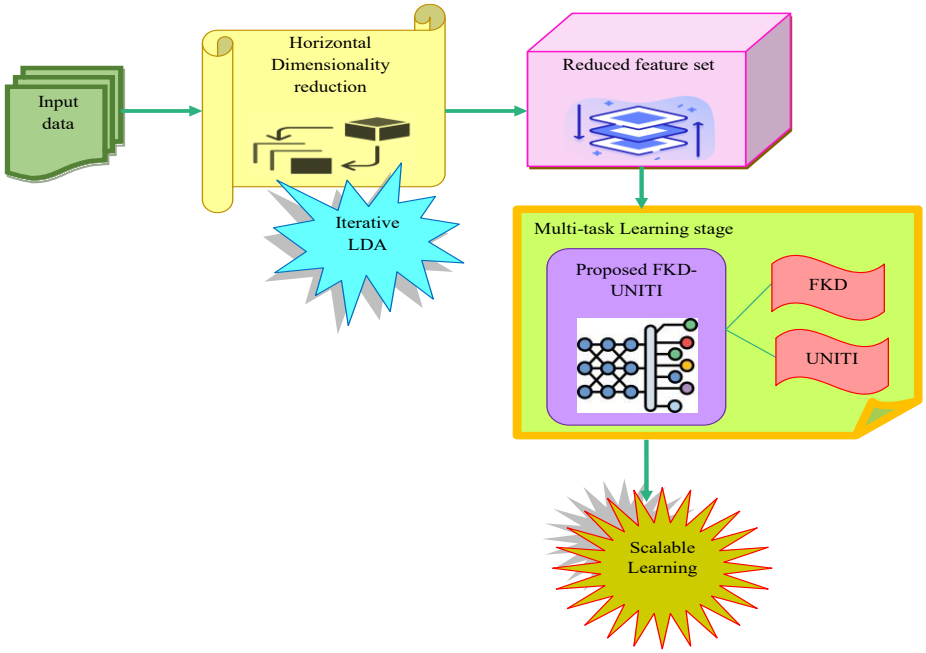


Fig. 1. Block Diagram of the Developed Method for Multi-task Learning

B. Dataset Description

Dataset 1: The Interactive Intro to Dimensionality Reduction dataset link is available at "https://www.kaggle.com/code/arthurtok/interactive-intro-to-dimensionality-reduction/input" accessed 2026-03-10. The dataset contains high-dimensional features that need to be reduced for easier analysis and better model performance. To handle this, three common dimensionality reduction techniques are considered: PCA, LDA, and t-Distributed Stochastic Neighbor Embedding (t-SNE). Interactive visualizations are generated using the Plotly library to provide the design of dataset after dimensionality reduction.

Dataset 2: The Dimensionality-Reduction-Techniques dataset link is "https://github.com/Pradnya1208/Dimensionality-Reduction-Techniques" accessed 2026-03-10. Machine learning and deep learning models can solve many tasks using powerful CPUs and GPUs, but having too many features in a dataset can slow down training and reduce accuracy. Dimensionality reduction is utilized to minimize the features while keeping the important information.

4. Detailed Description Of Dimensionality Reduction And Multi-Task Learning Using Unified Neural Networks

A. Iterative LDA-based Dimensional Reduction

LDA [16] is a statistical method for feature extraction and dimensionality reduction. It reduces higher-dimensional data to a low-dimensional space by increasing the separation among different classes. This separation is measured using the Fisher score (Fs), where a low score implies that the classes overlap or not well distinguished. LDA works by finding a linear combination of input variables, which separates two or more groups. For multiple classes, it generalizes to multi-class LDA and maximizes the ratio of variance between groups U_p to the variance within groups U_k in the projected space, which is defined in the expression below.

$$\max_k \left(\frac{k^T U_p k}{k^T U_k k} \right) \quad (1)$$

$$U_p = \sum_{h=1}^{N} (\alpha_h - \alpha)(\alpha_h - \alpha)^T \quad (2)$$

$$U_k = \sum_{h=1}^{N} \sum_{x=1}^{M_h} (s_{hx} - \alpha^h)(s_{hx} - \alpha^h)^T \quad (3)$$

Here, N implies total class, M_h denotes total samples, s_{hx} represents x^{th} sample of h^{th} class, α_h specifies mean vector, and α indicates mean vector of all samples. To determine maximum value of the ratio, the derivative is taken and set to zero, which is specified in Eq. (4).

$$\frac{d}{dk} \left(\frac{k^T U_p k}{k^T U_k k} \right) = 0 \quad (4)$$

The generalized eigenvector problem is solved using a matrix, where g represents the dimension of feature space. The projection space is then obtained based on Eq. (5).

$$z = k^T s \quad (5)$$

Here, specifies new feature in projection space, denotes input variable, and implies vector. The process of iterative LDA involves three main stages. In first stage, the dataset is divided into training and a testing set. In second stage, LDA performs repeatedly through an iterative procedure, and last stage applies the resulting reduced-dimensional features for analysis. The iteration begins by selecting features, where denotes total classes and applies LDA to produce new features. Each new feature is then combined with the next original feature as input for LDA and generates another set of reduced features. This process continues sequentially until all features are incorporated, which indicates the first epoch. To achieve optimal performance, this iterative procedure is repeated for multiple epochs and each epoch refines the feature representation until the results are considered optimal. The architecture of Iterative LDA model is depicted in Fig 2.

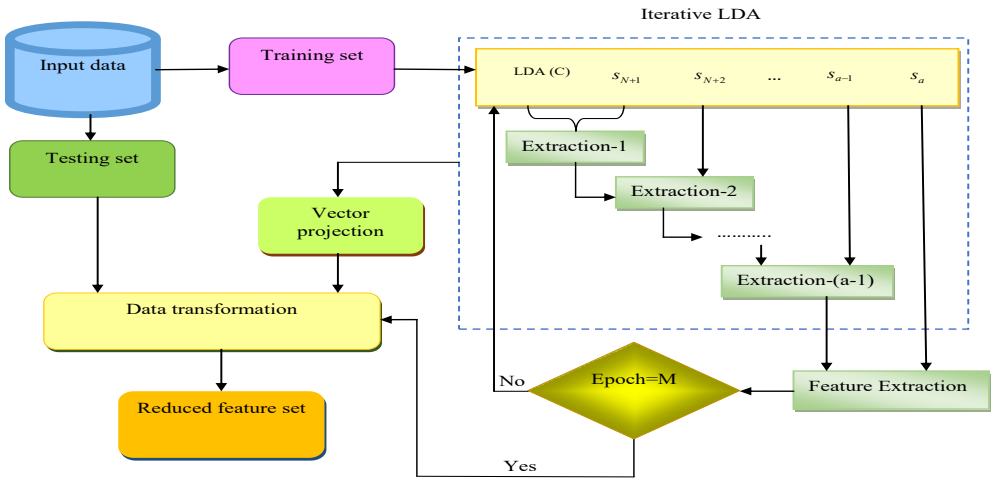


Fig 2. Architecture of Iterative LDA model

B. Feature-level Knowledge Distillation Module

Knowledge distillation [17] is used for model compression and knowledge transfer. It allows a simpler model to learn from a complex model, which enables the smaller model to achieve strong performance with lower computational cost. The main idea is to use the internal features of the larger model as guidance when training the smaller model. This approach reduces the complexity of the model by improving task performance. A knowledge distillation system typically consists of two key components.

Teacher network: This network is usually defined by its deep and complex architecture that gives strong feature extraction abilities. It captures detailed high-level semantic features and subtle patterns in the data, creating a rich source of knowledge that can be transferred to a smaller and simpler model.

Student network: This is a compact network created to mimic the outputs or the intermediate features of the larger model using a specially designed distillation loss. Through this process, it learns effective feature representations by keeping computational requirements low.

Feature-level knowledge distillation focuses on the representations in the intermediate layers of the larger model. By aligning spatial structures, channel relationships, and attention patterns, it helps the smaller model to capture the semantic meaning of deep features. This approach preserves high-level abstract information by enforcing consistency between the intermediate feature maps of both models. The difference between the teacher and student feature maps is measured using a feature loss function F_{Lfnc}, and it is described in Eq. (6).

$$F_{Lfnc} = \frac{1}{J}\sum_{b=1}^{C}\left\| y_b^T - y_b^R \right\|_2^2 \quad (6)$$

Here, C implies total feature maps, y_b^T and y_b^R indicates feature maps of teacher and student, and $\|\cdot\|_2^2$ specifies square of L_2 norm.

C. Multi-Task Learning through FKD-UNITI

FKD-UNITI is used for multi-task learning to effectively integrate knowledge from multiple tasks and datasets into a single unified model while preserving task-specific representations. However, conventional methods rely on simple joint training, which leads to limitations such as negative transfer, task interference, and imbalanced learning when datasets differ in size. These approaches also struggle when tasks are trained on separate datasets because standard multi-task frameworks typically require shared data and aligned labels. FKD-UNITI addresses these issues by applying feature-level knowledge distillation, where informative representations learned by task-specific teacher models are distilled into a unified student network. This allows the model to retain discriminative features for each task while learning shared representations.

In multi-task learning, FKD-UNITI enables efficient knowledge sharing by preserving task-specific representations. By distilling feature-level information from multiple teacher networks into a unified student model, FKD-UNITI allows the model for learning rich and discriminative representations from different tasks without requiring all tasks to be trained on the same dataset. This approach improves generalization and robustness by transferring complementary knowledge across tasks and datasets. Additionally, FKD-UNITI helps to reduce task interference and negative transfer, which are common issues in conventional multitask learning frameworks that rely on simple parameter sharing. The unified architecture leads to better computational efficiency because a single model performs multiple tasks instead of maintaining separate networks for each task.

The model collects high-dimensional data from relevant databases, which contain numerous features and heterogeneous information. Next, Iterative LDA is employed for dimensional reduction. In this stage, the method iteratively projects the original feature space into a lower-dimensional discriminative space by

maximizing the separability between classes while minimizing the variance within each class. This process produces a stable and compact feature representation that preserves the most important and informative characteristics of the data while removing redundant and noisy features. In addition to reducing the feature space, a horizontal dimensionality reduction process is considered in the multitask learning. The horizontal reduction focuses on selecting and aligning the most relevant feature subsets across multiple datasets, which ensures that common informative features are shared while maintaining task-specific variations. This helps to improve feature consistency and reduce unnecessary complexity when integrating multiple tasks. After obtaining the reduced and discriminative feature representations, the processed data are then used in the multitask learning stage. In this phase, the FKD-UNITI framework is applied. The UNITI framework [18] uses sequential dataset training, which enables a single model to learn from multiple heterogeneous datasets while reducing the problem of catastrophic forgetting. In this process, the encoder extracts shared feature representations from these datasets that capture important patterns across tasks. It also applies knowledge distillation to transfer knowledge from pre-trained teacher models that are trained on specific datasets, which allows the student model to retain important task-related information. The decoder then uses these encoded features to generate task-specific outputs. Reparameterization is employed to reduce the model's complexity. In addition, UNITI uses feature-level distillation to transfer internal feature representations instead of relying only on output-based distillation, which helps to overcome the issues with heterogeneous datasets. During training, the framework continuously refines these shared features, which improves their quality and enables effective integration of tasks across heterogeneous datasets. This approach ensures scalable learning that allows the model to generalize across multiple data sources while maintaining a unified architecture and high performance. Fig 3 shows the graphical representation of developed FKD-UNITI model for multi-task learning.

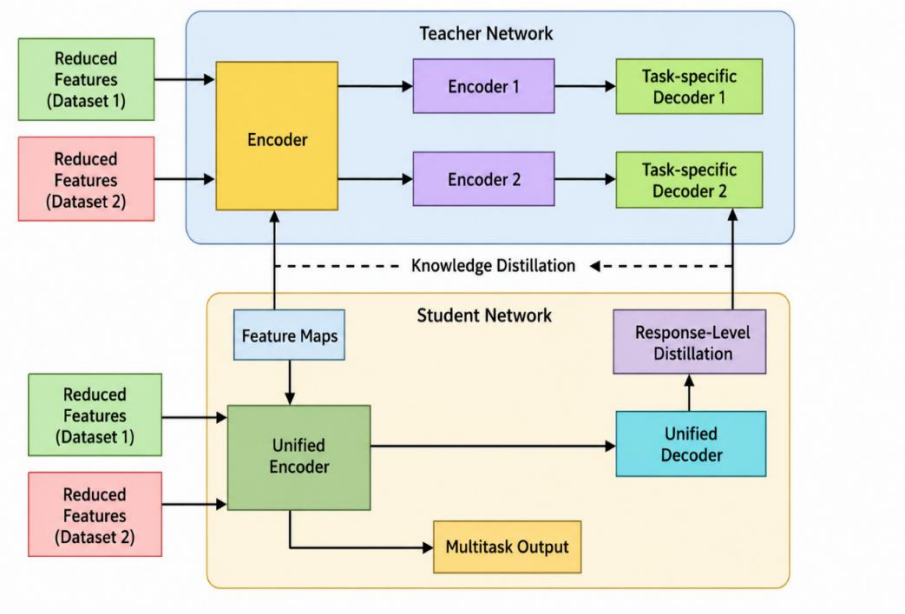


Fig 3. Graphical Representation of Developed FKD-UNITI Model for Multi-task Learning

5. RESULTS AND DISCUSSION

A. Simulation setup

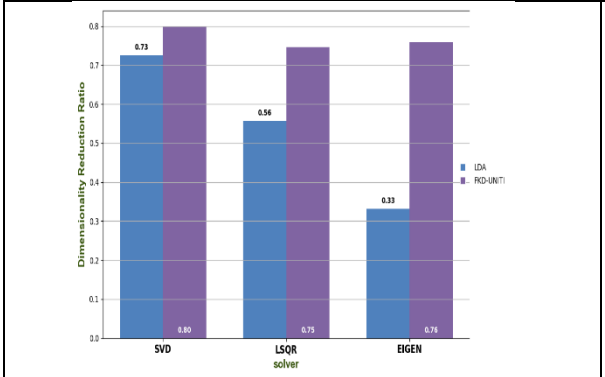
Python was used to perform experiments for the horizontal dimensionality reduction framework. The performance of the system was evaluated using standard evaluation metrics under different parameter settings. In addition, conventional dimensionality reduction methods such as LoRA [1], MEL [2], PCA [3], and LDA [19] were used

to compare the performance of the proposed FKD-UNITI model.

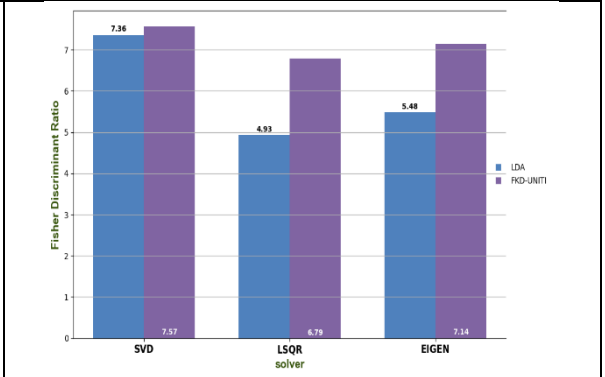
B. Performance analysis of proposed model using solver

Fig 4 illustrates the performance of the proposed FKD-UNITI model with LDA using different solvers such as SVD, LSQR, and EIGEN for Dataset 1 and Dataset 2. The evaluation is carried out using several performance metrics, including explained variance ratio, reconstruction error, information loss, Fisher discriminant ratio, silhouette score, dimension reduction ratio, computation time, and memory usage. These metrics help to measure the ability of each method to preserve important data characteristics and reduce the dimensionality of the dataset. The FKD-UNITI model provides better feature representation with higher explained variance and Fisher discriminant ratio, while achieving lower reconstruction error and reduced information loss compared to the LDA solvers. In addition, the proposed model demonstrates improved clustering quality through a higher silhouette score and maintains efficient dimensionality reduction with optimized computation time and lower memory consumption. Additionally, the FKD-UNITI framework provides more stable and efficient performance than the conventional LDA approaches with different solvers.

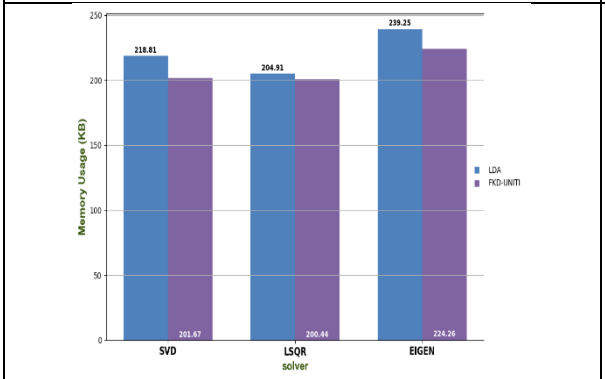
Dataset 1



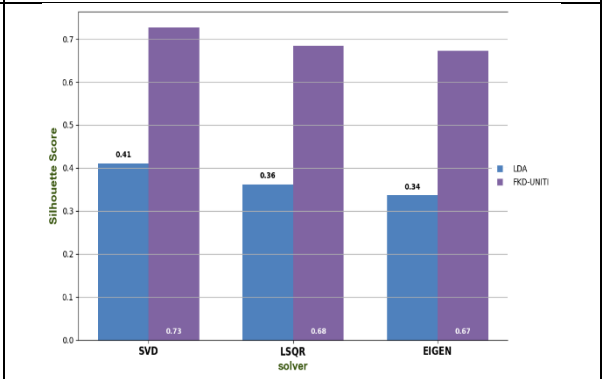
(a)



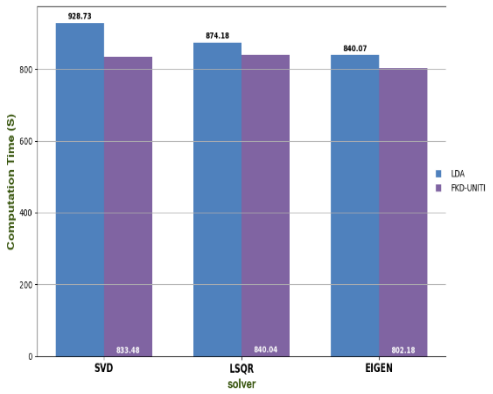
(b)



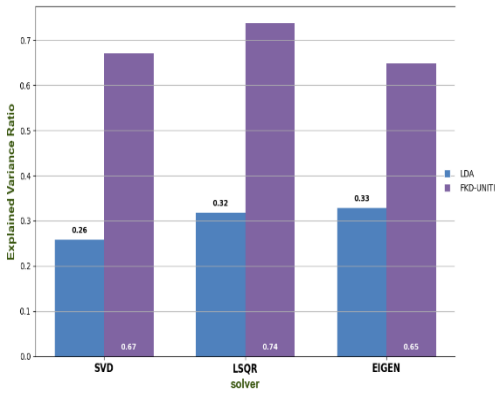
(c)



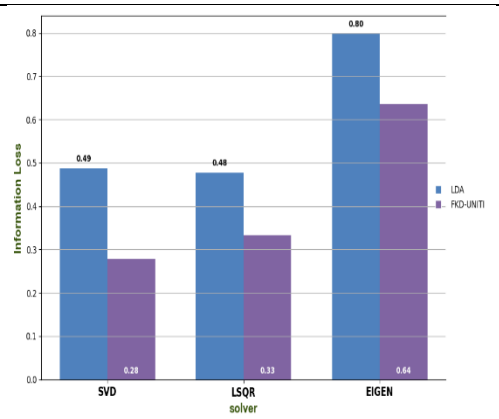
(d)



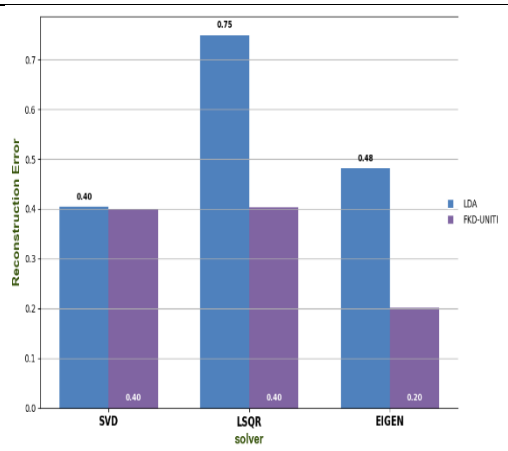
(e)



(f)

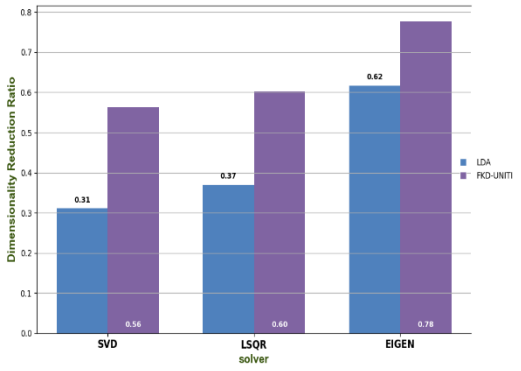


(g)

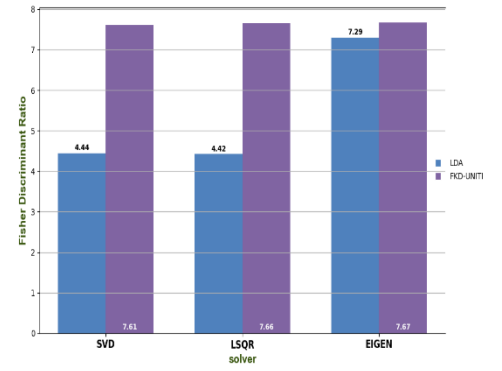


(h)

Dataset 2



(a)



(b)

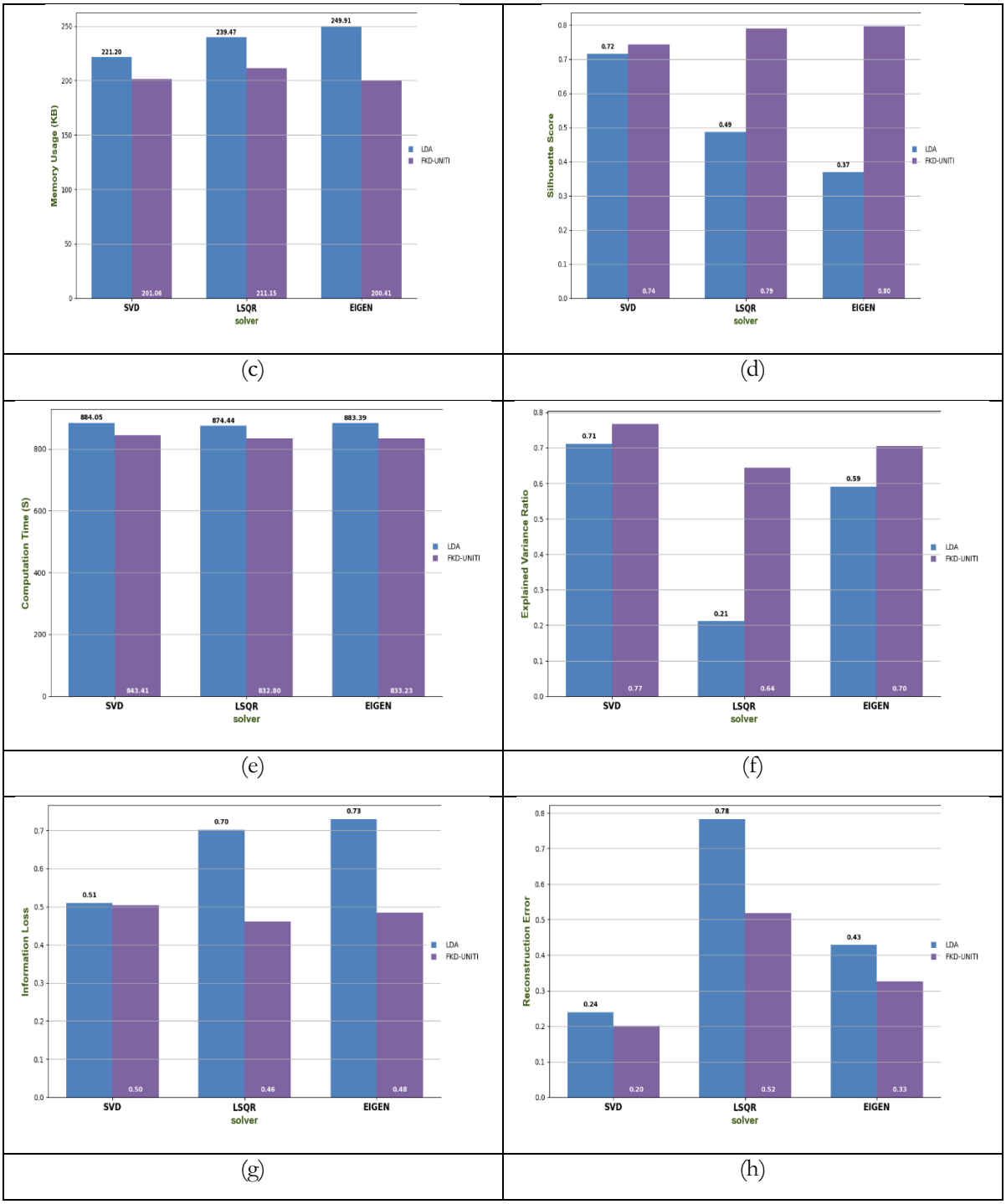


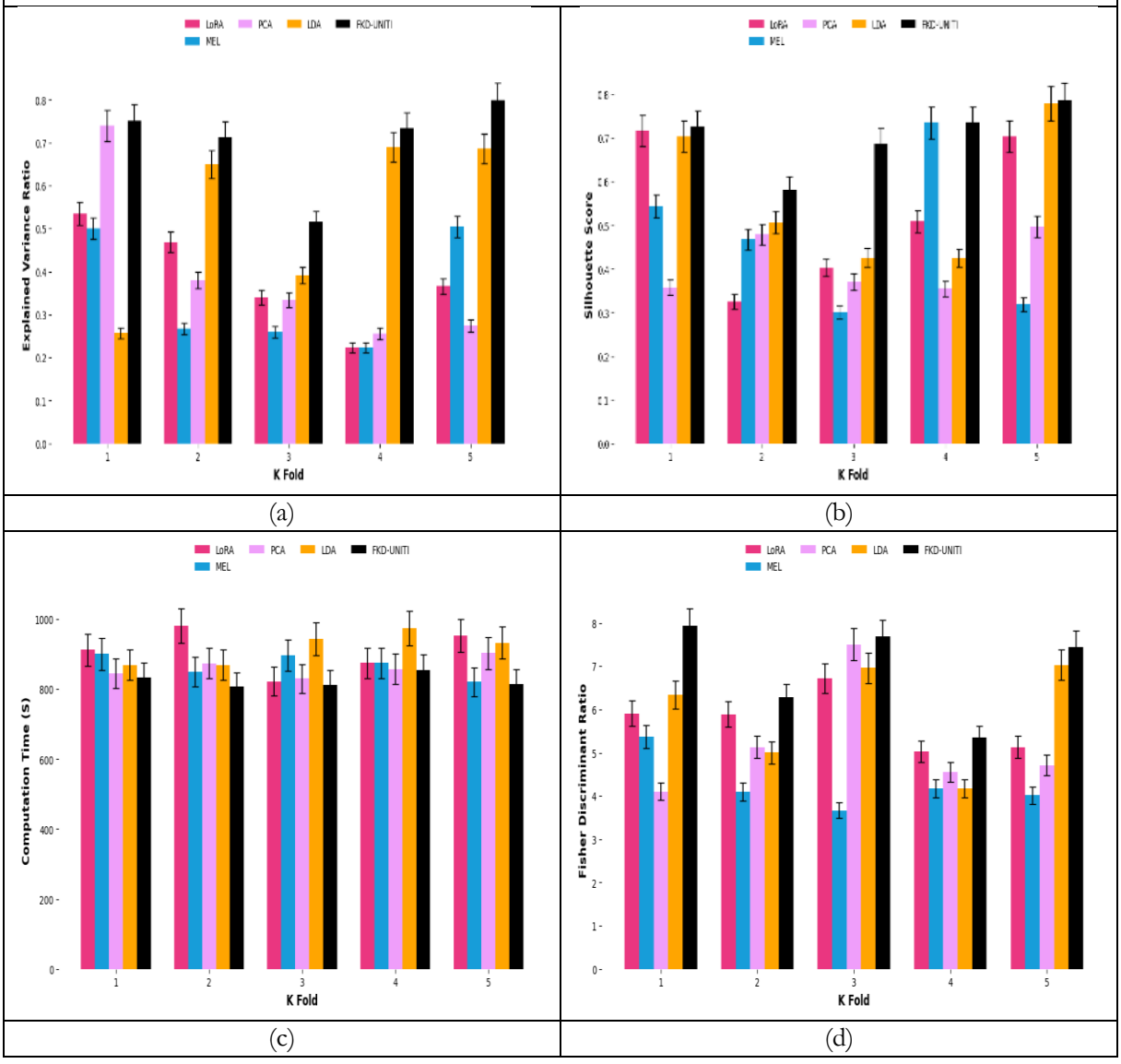
Fig 4. Performance Assessment of Proposed Model using Solver concerning (a) Dimensionality Reduction Ratio (b) Fisher Discriminant Ratio (c) Memory Usage (d) Silhouette Score (e) Computation Time (f) Explained Variance Ratio (g) Information Loss (h) Reconstruction Error

C. Kfold assessment of the proposed model

The Kfold analysis of proposed FKD-UNITI approach for Dataset 1 and Dataset 2 is displayed in Fig 5. The proposed model is compared with conventional approaches such as LoRA, MEL, PCA, and LDA using several evaluation metrics. The proposed model achieves lower reconstruction error, which indicates that the reduced feature representation preserves more important information. Metrics such as explained variance ratio, reconstruction error, and information loss are used to measure the ability of each method to preserve important data characteristics after dimensionality reduction.

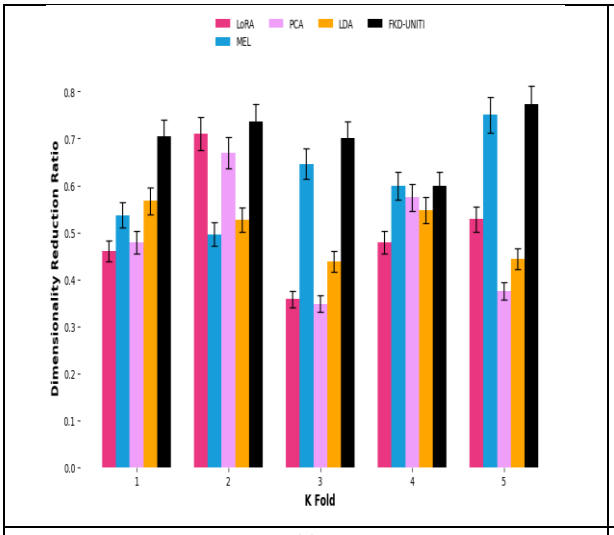
The Fisher discriminant ratio and silhouette score evaluate the quality of class separability and clustering structure in the reduced feature space. Additionally, dimension reduction ratio indicates the method's effectiveness to reduce the number of features while maintaining meaningful information. Computational efficiency is also assessed using computation time and memory usage measure the processing cost of each method. Through this analysis, the FKD-UNITI model demonstrates improved feature representation, better preservation of data information, enhanced class separability, and more efficient resource utilization compared to the conventional dimensionality reduction techniques. Table II depicts the statistical assessment of proposed model using classifier comparison. In dataset 1, the proposed FKD-UNITI model achieves better explained variance ratio of 11.19% than LoRA, 28.07% than MEL, 57.84% than PCA, and 22.06% than LDA at activation=Linear. Moreover, the proposed model achieves more reliable performance and better feature representation for complex data analysis tasks.

Dataset 1

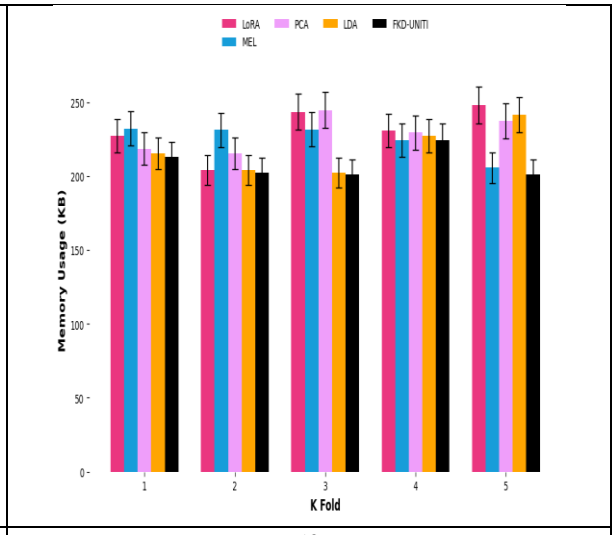


(a) (b)

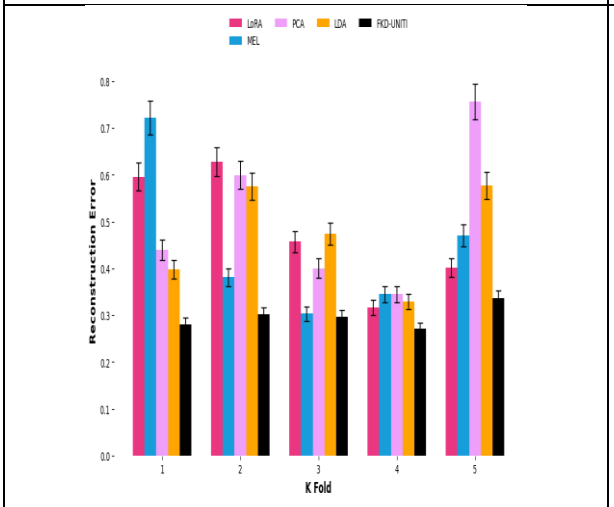
(c) (d)



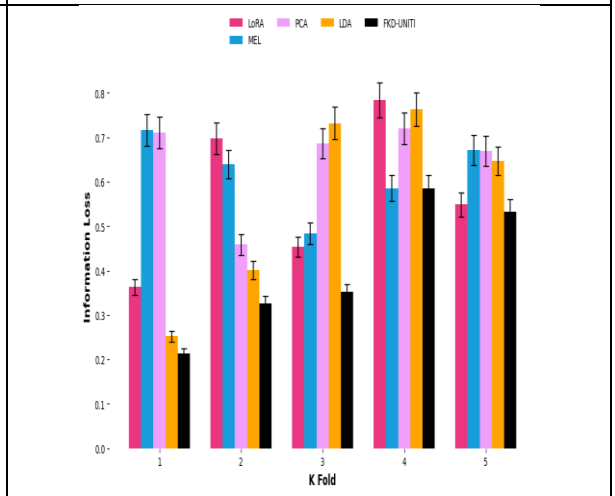
(e)



(f)

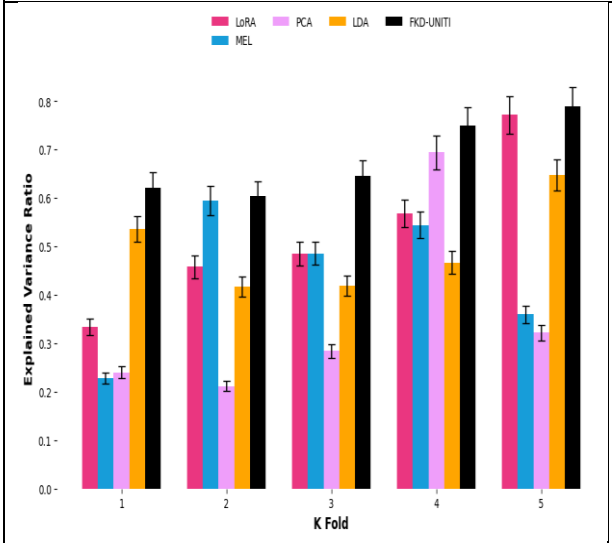


(g)

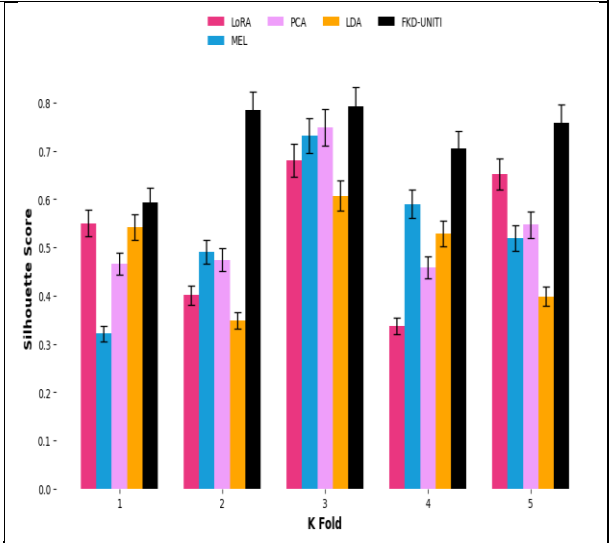


(h)

Dataset 2



(a)



(b)

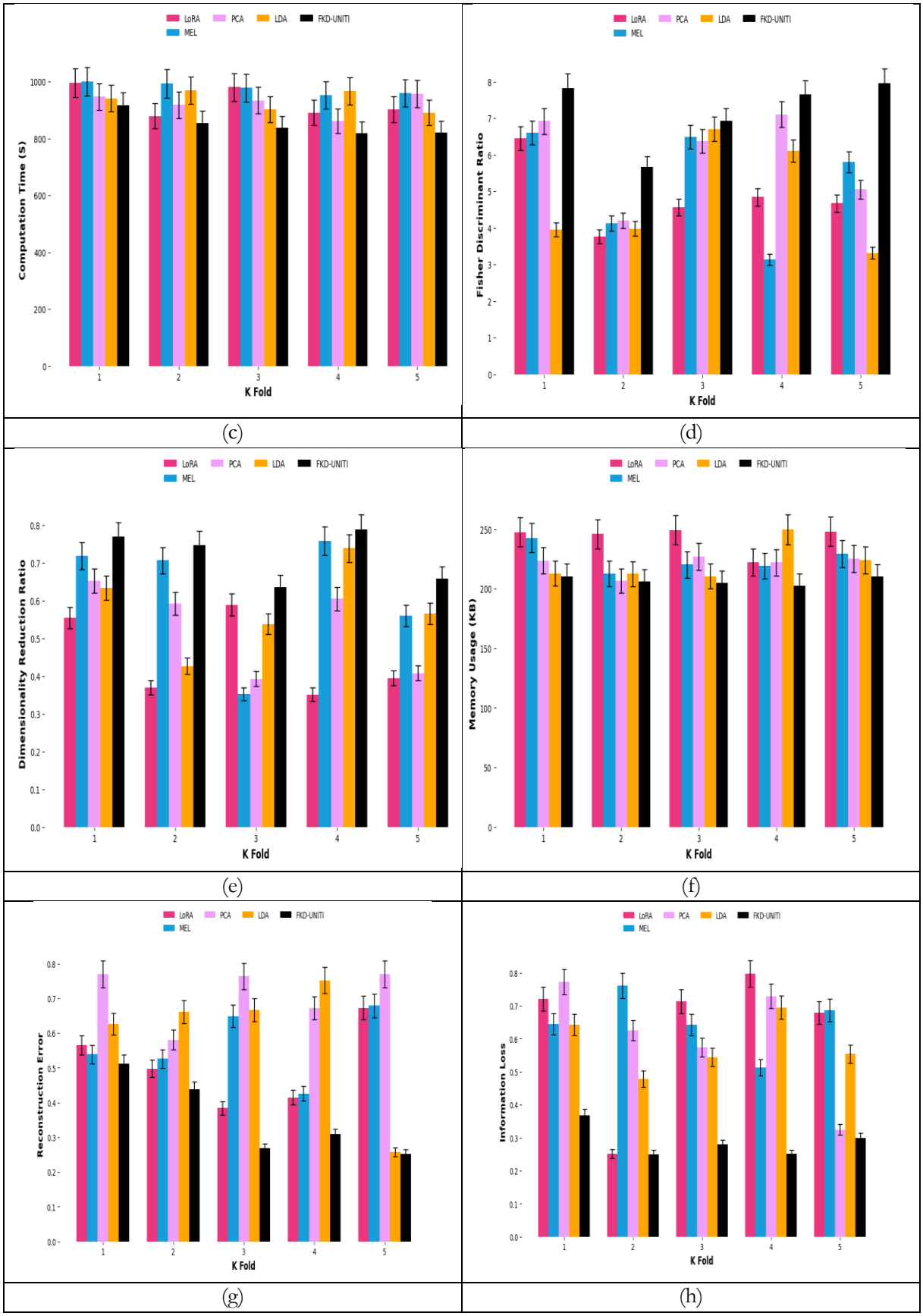


Fig 5. Kfold Assessment of Proposed Model using (a) Explained Variance Ratio (b) Silhouette Score (c) Computation Time (d) Fisher Discriminant Ratio (e) Dimensionality Reduction Ratio (f) Memory Usage (g) Reconstruction Error (h) Information Loss

Table II.. Statistical analysis of proposed model based on classifier comparison

Dataset 1	Activation	LoRA [1]	MEL [2]	PCA [3]	LDA [19]	FKD-UNITI
	Explained Variance Ratio					
	Linear	0.666202	0.539604	0.316265	0.584671	0.750133
	ReLU	0.676428	0.257372	0.220798	0.274163	0.751537
	Leaky ReLU	0.227873	0.547141	0.331146	0.633472	0.79398
	TanH	0.463683	0.331544	0.578443	0.257684	0.792192
	Sigmoid	0.295132	0.595333	0.486219	0.556828	0.64598
	Reconstruction Error					
	Linear	0.679515	0.361628	0.421091	0.763715	0.356128
	ReLU	0.330025	0.527579	0.772253	0.319531	0.23471
	Leaky ReLU	0.472573	0.644418	0.420498	0.607738	0.209385
	TanH	0.356184	0.655132	0.72889	0.523052	0.218487
	Sigmoid	0.481273	0.612864	0.287422	0.779127	0.231285
	Information Loss					
	Linear	0.626285	0.414166	0.324703	0.38009	0.322272
	ReLU	0.788233	0.516273	0.445143	0.443148	0.318104
	Leaky ReLU	0.337517	0.722382	0.509331	0.615416	0.332063
	TanH	0.573121	0.558385	0.639694	0.618602	0.293393
	Sigmoid	0.714335	0.563535	0.74749	0.717083	0.318413
	Fisher Discriminant Ratio					
	Linear	5.408633	5.924586	5.793207	6.902258	7.114442
	ReLU	5.045391	4.923821	7.266992	5.922003	7.936924
	Leaky ReLU	5.356255	4.90957	6.707408	5.352976	7.939757
	TanH	5.995685	3.835551	5.595085	4.261258	6.681362
	Sigmoid	4.908101	6.179015	4.728566	4.505514	6.658981
	Silhouette Score					
	Linear	0.398581	0.644467	0.550796	0.341624	0.652208
	ReLU	0.62375	0.491447	0.311444	0.643094	0.76207
	Leaky ReLU	0.487414	0.766197	0.646372	0.441868	0.77872
	TanH	0.524011	0.468559	0.330553	0.680084	0.765518
	Sigmoid	0.471687	0.344423	0.352254	0.441314	0.64748

	Dimension Reduction Ratio					
	Linear	0.440412	0.438205	0.396362	0.436474	0.542785
	ReLU	0.688891	0.730451	0.637807	0.54509	0.731404
	Leaky ReLU	0.329521	0.370466	0.532918	0.729166	0.784855
	TanH	0.471499	0.396106	0.326193	0.346384	0.747801
	Sigmoid	0.658545	0.444058	0.349057	0.664279	0.748037
	Computation Time					
	Linear	52.02449	55.19053	56.31454	53.38105	50.04258
	ReLU	58.75811	59.75871	56.36245	56.15426	51.13565
	Leaky ReLU	54.42976	52.61842	53.02794	58.43308	50.97287
	TanH	58.59563	58.16553	58.13676	58.73985	54.8764
	Sigmoid	53.28661	55.20681	56.05667	61.94476	50.24655
	Memory Usage					
	Linear	241.7442	237.9323	238.7238	244.9616	220.4701
	ReLU	238.6351	224.4507	231.9204	234.2054	206.2232
	Leaky ReLU	239.7964	243.6381	210.0028	230.2602	206.7785
	TanH	226.5649	218.7516	223.4598	200.4739	200.2964
	Sigmoid	248.6699	240.577	247.0065	227.4236	208.415
Dataset 2	Explained Variance Ratio					
	Linear	0.364296	0.208779	0.208779	0.356694	0.721371
	ReLU	0.481988	0.481988	0.520712	0.560385	0.662751
	Leaky ReLU	0.682169	0.580841	0.220009	0.245858	0.772365
	TanH	0.727793	0.474924	0.465682	0.532883	0.776676
	Sigmoid	0.410145	0.400397	0.590229	0.558537	0.712535
	Reconstruction Error					
	Linear	0.654418	0.264626	0.214542	0.417162	0.214542
	ReLU	0.300346	0.475918	0.300346	0.61878	0.215493
	Leaky ReLU	0.538566	0.436827	0.737178	0.727931	0.389976
	TanH	0.34715	0.505788	0.46627	0.510965	0.324247
	Sigmoid	0.712555	0.523128	0.395736	0.740085	0.273931
	Information Loss					
	Linear	0.524237	0.752317	0.214667	0.712966	0.214667
	ReLU	0.752637	0.767645	0.752637	0.695006	0.534549

	Leaky ReLU	0.741086	0.718132	0.765632	0.54346	0.470111
	TanH	0.462628	0.582484	0.797083	0.582634	0.221266
	Sigmoid	0.692814	0.421289	0.250866	0.779284	0.235124
	Fisher Discriminant Ratio					
	Linear	7.134307	4.851981	4.851981	4.582816	7.495537
	ReLU	3.66847	6.402574	5.648543	3.66847	7.828006
	Leaky ReLU	5.55926	6.191063	3.227388	3.006962	7.803238
	TanH	3.411785	4.369026	5.360647	4.370568	6.873998
	Sigmoid	3.736742	3.675905	3.734895	5.145512	5.556053
	Silhouette Score					
	Linear	0.482135	0.490126	0.715603	0.609225	0.715603
	ReLU	0.335731	0.335731	0.535051	0.413881	0.628098
	Leaky ReLU	0.513592	0.594677	0.69869	0.426855	0.730013
	TanH	0.721544	0.408676	0.346836	0.445999	0.74125
	Sigmoid	0.685488	0.332271	0.731853	0.728447	0.750274
	Dimension Reduction Ratio					
	Linear	0.505812	0.689505	0.505812	0.542774	0.791883
	ReLU	0.781212	0.435004	0.622524	0.67236	0.781212
	Leaky ReLU	0.570139	0.671739	0.568198	0.344474	0.74069
	TanH	0.658192	0.629839	0.312757	0.402773	0.696142
	Sigmoid	0.391373	0.402481	0.317521	0.400204	0.419159
	Computation Time					
	Linear	58.69655	56.48547	52.6389	52.6389	51.12442
	ReLU	53.27867	53.27867	54.6167	60.68743	50.95153
	Leaky ReLU	56.1713	59.30416	53.85137	60.94363	52.7324
	TanH	51.158	57.05661	59.5229	56.16401	50.30644
	Sigmoid	61.5688	58.43079	57.72031	60.84155	56.7921
	Memory Usage					
	Linear	201.1657	216.2691	202.6867	202.6867	200.5885
	ReLU	247.6779	247.6779	220.5861	242.7254	208.1709
	Leaky ReLU	229.0433	244.2427	239.4901	237.946	225.7691
	TanH	218.8135	219.8925	241.5145	212.6199	208.8069
	Sigmoid	235.1506	214.2822	245.3401	246.8487	207.6479

6. Conclusion

This work presents an advanced multitask learning framework that leverages horizontal dimensionality reduction and feature-level knowledge distillation to integrate and share important features across heterogeneous datasets. The proposed FKD-UNITI model demonstrates effectiveness in multitask learning while maintaining task-specific information. The framework improves learning efficiency by enabling the model to handle multiple tasks simultaneously. It enhances scalability and generalizes well across diverse datasets that reduces the need for multiple models by providing a unified architecture. Additionally, it improves task integration by refining shared features across datasets and preserves critical task-specific information through feature-level knowledge distillation. However, the approach also has complexities, such as increased computational complexity due to feature-level distillation and issues when handling extremely large and highly imbalanced datasets.

References

Elbakary, A., Ben Issaid, C., ElBatt, T., Seddik, K., & Bennis, M. (2025). Mira: A method of federated multi-task learning for large language models. *IEEE Networking Letters*.

Wang, X., Shangguan, H., Huang, F., Wu, S., & Jia, W. (2024). MEL: Efficient multi-task evolutionary learning for high-dimensional feature selection. *IEEE Transactions on Knowledge and Data Engineering, 36*(8), 4020–4033.

Luo, Z., Merainani, B., Döhler, M., Baltazart, V., & Zhang, Q. (2023). High dimensional data reduction in modal analysis with stochastic subspace identification. *IFAC-PapersOnLine, 56*(2), 10371–10376.

Wang, H., Hu, J., Wei, S., & Qu, Y. (2024). One-shot TSOM with a multi-task learning model for simultaneous dimension measurement and defect inspection. *Optics and Lasers in Engineering, 181*, 108345.

Li, R., Zhou, G., Li, X., Jia, L., & Shang, Z. (2024). Semi-supervised multi-label dimensionality reduction learning based on minimizing redundant correlation of specific and common features. *Knowledge-Based Systems, 294*, 111789.

Migenda, N., Möller, R., & Schenck, W. (2021). Adaptive dimensionality reduction for neural network-based online principal component analysis. *PLOS ONE, 16*(3), e0248896.

Mor, U., Cohen, Y., Valdés-Mas, R., Kviatcovsky, D., Elinav, E., & Avron, H. (2022). Dimensionality reduction of longitudinal 'omics data using modern tensor factorizations. *PLOS Computational Biology, 18*(7), e1010212.

Zhang, Y., Yuan, P., Jiang, L., & Ewe, H. T. (2022). Novel data-driven spatial-spectral correlated scheme for dimensionality reduction of hyperspectral images. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 15*, 3877–3890.

Syed, F. I., Muther, T., Kalantari Dahaghi, A., & Negahban, S. (2022). Low-rank tensors applications for dimensionality reduction of complex hydrocarbon reservoirs. *Energy, 244*, 122680.

Behrens, G., Beucler, T., Gentine, P., Iglesias-Suarez, F., Pritchard, M., & Eyring, V. (2022). Non-linear dimensionality reduction with a variational encoder decoder to understand convective processes in climate models. *Journal of Advances in Modeling Earth Systems, 14*(8).

Yang, H., Yang, Q., Mu, Z., Du, X., & Chen, L. (2022). Optimal design of three-dimensional circular-to-rectangular transition nozzle based on data dimensionality reduction. *Energies, 15*(24), 9316.

Gao, W., Ma, Z., Xiong, C., & Gao, T. (2023). Dimensionality reduction of SPD data based on Riemannian manifold tangent spaces and local affinity. *Applied Intelligence, 53*(2), 1887–1911.

Rodriguez-Lozano, F. J., Gámez-Granados, J. C., Palomares, J. M., & Olivares, J. (2023). Efficient data dimensionality reduction method for improving road crack classification algorithms. *Computer-Aided Civil and Infrastructure Engineering, 38*(16), 2339–2354.

Pan, S., Brunton, S. L., & Kutz, J. N. (2023). Neural implicit flow: A mesh-agnostic dimensionality reduction paradigm of spatio-temporal data. *Journal of Machine Learning Research, 24*(41), 1–60.

Kubo, H., Kimura, T., & Shiomi, K. (2023). Exploratory data analysis of earthquake moment tensor catalog in Japan using non-linear graph-based dimensionality reduction. *Pure and Applied Geophysics, 180*(7), 2689–2703.

Sudibyo, U., Rustad, S., Andono, P. N., Fanani, A. Z., & Supriyanto, C. (2024). iLDA: A new dimensional reduction method for non-Gaussian and small sample size datasets. *Egyptian Informatics Journal, 27*, 100533.

Zhao, E., Zhang, H., Qu, N., Wang, Y., & Zhao, Y. (2025). KDAD: Knowledge distillation-based anomaly detection for thermal infrared hyperspectral image. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Kim, S., Park, Y., & Lee, E. C. (2025). UNITI: Framework for multi-task learning across datasets to mitigate overfitting. *Expert Systems with Applications, 282*, 127653.

Zhou, H., Kang, Y., Liu, G., & You, G. (2025). An improved LDA dimension reduction algorithm for multivariate time series classification. *Journal of Applied Statistics*, 1–14.