An Interpretable Fuzzy Graph Learning for Label Propagation Assisting Data Classification

Cherukula Madhu, *Member, IEEE,* Sudhakar M.S.

***Abstract*—Graph-based Semi-Supervised Learning (GSSL) has become an indispensable tool for data classification recently, owing to its innate capability of efficient data structuring and representation. However, their reliance on predefined graphs constrains the efficacy of Label Propagation (LP) and interpretability in predictions, especially in high-dimensional feature spaces with limited information. Addressing these challenges, this paper employs a Fuzzy Graph-based Label Propagation (FGLP) model, which is inherently interpretable in exploring the similarities of the normalized histogram envelope-based scaled features assisting data categorization. FGLP initiates the structuring of an undirected fuzzy-weighted graph using the novel fuzzy distance matrix by exploiting the local data affinity with reduced influence of outliers. The learned information is then optimized using two distinct SoftMax-constrained objective functions coupled with cross-entropy and lasso regularization to construct the similarity and projection matrices in tandem assisting LP and Feature Selection in scarcely labeled high-dimensional feature space. Performance validation on heterogeneous datasets showcases FGLP's superiority, achieving over 88% accuracy with just 10% labeled data, surpassing prior methods by an average enhancement of 18.29%.**

***Index Terms*—Fuzzy Distance Matrix (FDM), Fuzzy Graph (FG), Fuzzy Graph-based Label Propagation (FGLP), Feature Selection (FS), Graph-based Semi-supervised learning (GSSL), Normalized Histogram Envelope (NHE), Similarity matrix.**

I. Introduction

THE daily expansion and exponential spread of social media networks across the globe generate tons of data, of which only a tiny percentage are labeled, as labeling is cost-intensive and incur more time with human labor [1]–[3]. Consequently, posing a significant challenge to the machine learning society in performing effective data exploration and structuring supplementing classification [2]. Over the years, Semi-Supervised Learning (SSL) faired handy owing to its simplicity with sophistication and has received significant attention in the area of label propagation (LP). Majorly, the problem with LP is the assignment of labels for the entire set of feature vectors (FV) with relatively lesser available labels in the FV subset [4]–[6]. This elucidates SSL's potency in addressing a wider variety of issues such as poor model generalization and low classification accuracy concerned with supervised and unsupervised learning [2].

Among the diverse SSL variants, particularly, GSSL benefits data scientists in managing and analyzing exponentially growing unstructured data due to their inherent convexity, scalability, and unique suitability in detecting inherent connections among data points [7], [8]. Graphs serve as the building blocks in GSSL [9]–[12] covering subspace and manifold learning [13], [14] wherein, the affinity between FVs is modeled as pairwise edges [15], [16] assisting LP [8]. Generally, GSSL graphs a dataset with each sample denoted as a vertex and their relation defined by an edge [17] with its nodal affinity defined by the degree of closeness determined using the distance measure.

Despite the numerous advantages and simplicity of GSSL, their reliability depends on the constructed similarity matrix [18] weighted by the Gaussian ($S_{ij}^{x} = e^{-\sigma(d)^2}$; $d-$ is the Euclidean distance measure on $\mathcal{X} = \{x_i\}_{i\in\mathcal{Z}}$; $\sigma-$ scaling parameter) which in turn relies on σ [12]. GSSL further insists that nearby nodes are strongly related and are presumed to have comparable labels. Subsequently, GSSL creates several manifolds that intersect or partially overlap with each other [11], when dealing with the distribution of items in diverse datasets. As the labels are allocated based on their proximity, the class information on the graph gets smoothened by neglecting the dissimilar edges [12]. However, GSSLs fail miserably, when the labeled data is highly limited [19], [20] and increases the processing cost while handling large-scale data. Moreover, the graph connectivity built by Euclidean among samples during propagation is unreliable [21]. In addendum, the accuracy of data categorization depends critically on the quality of the initialized graph and is questionable, as GSSLs tend to classify the data on a predetermined or fixed graph with subjective label information.

Inferences from the erstwhile literature on GSSL applied to LP reveal the following challenges: 1. graph structuring, 2. Reliability, and 3. LP efficacy, demanding exhaustive investigations to deal with. Especially, these challenges are notable when dealing with high-dimensional feature spaces limited by the label information. Accordingly, to address the aforementioned issues, this work proposes the FGLP by replacing the conventional crisp graph (CG) with FG for data representation offering interpretability, transparency, dynamicity, and adaptivity [22], [23]. To begin with, FGLP initiates graph structuring using a novel fuzzy distance matrix (FDM) instead of the conventional Euclidean easing computational operations in high-dimensional data spaces with simultaneous reduction of outliers [21], [24]. The fuzzy-natured FDM realizes edges by capturing the likelihood of two nodes sharing the same label which is in contrast to the CG, relying heavily on the subjective label information prone to misclassification [25], [26]. Further, the parameter-free FGLP model ensures reliability by adaptive determination of cross-entropy and lasso regularizers penalty parameters through recursive tuning of the Lagrangian operator that mitigates the biases in label prediction. Additionally, FGLP favors the downing of computational complexity with enhanced interpretability using the introduced NHE-based fuzzy Feature extraction and FS warranting reduced FV dimensions. Moreover, the entire operations in fuzzy space enable reliable prediction of data connections without solely relying on the subjective label information, greatly enhancing the LP efficiency, even under constrained circumstances with greater interpretability. To elevate the appropriateness and help understand the significance of the proposal, the issues in GSSL are orderly emphasized based on the recent contributions in Section I-A.

A. Related Work

The quest for efficient data storage and management systems is on the rise owing to the humongous volume of unstructured data delivered across multiple domains catering to several real-world applications. Scores of trending supervised and unsupervised models necessitate higher computational demands with biased outcomes, and false classification tarnishing data categorization. Therefore, a suitable learning alternative in the form of inductive or transductive GSSL encompassing convexity with connectivity is adopted to improve data

Cherukula Madhu, Sudhakar M.S. (Corresponding author) are with the School of Electronics Engineering (SENSE), VIT, Vellore, 632014,Tamil Nadu, India. (e-mail: cherukulamadhu@gmail.com, sudhakar.ms@vit.ac.in)

classification. Induction learns a global representation of the data space from the "unknown" data, while transduction propagates labels by learning a local representation utilizing labeled and unlabeled data thereby making it progressive and hence adopted in this approach. Consequently, the below section outlines the issues related to the recent literature on transductive learning-based GSSLs employed for LP and embedding in the context of Dimensionality Reduction (DR) and graph structuring.

At the onset, GSSLs such as the local-global consistency (LGC) [27] and Gaussian field-based harmonic functions (GFHF) [28] reduced FVs dimensions by smoothing the graph's manifold that closed the gap between the predicted labels. Despite being straightforward and effective, these models gradually lost relevance with the increase in dataset size, which was addressed by Marginal Fisher Analysis (MFA) [29] by compacting the interclass deviations using an intrinsic graph. However, graph construction and projection learning demanded two different procedures that increased the computational complexity. Likewise, the semi-supervised low-rank representation [1] significantly reduced DR by maintaining a trade-off between accuracy and representational complexity. Alternatively [11] tuned the similarity matrix with fixed and adaptive FV weighting followed by LP which was quite expensive and the method failed with lesser labeled samples. The variant of [11] termed Semi-supervised Projection with Graph Optimization (SPGO) [30] learned the graph similarity in low dimensional space by neglecting the labeled FVs that degraded feature projection. Consequently, the multi-category classification [31], [32] employed a pair of linearly modeled unconstrained minimization criteria for learning the graph structure and label prediction. The engaged criteria increased the model's complexity proportional to the FVs size and datasets. The unification of LP with Structured Graph Learning (LPSGL) [33] reduced dataset dimensions using an iterative optimizer that demanded extensive computations and failed to perform well if the labeled samples are scarce. The probability-tuned similarity graph [2] performed LP using the Euclidean distance measure using iterative optimization that frequently altered the learned labels thereby, unsettling. The above discussions outline the need for an effective and adaptive LP mechanism functioning using a minimal number of computations.

Similarly, recent and prevailing graph embedding models are surveyed and chronologically enunciated hereunder, to showcase the downfalls that are addressed by FGLP. The widely popular semi-supervised discriminant analysis (SDA) [34] optimized location preservation and discrimination using both labeled and unlabeled samples. However, in cases of higher levels of data non-linearity, its inherent geometric structure was ignored, while flexible manifold embedding (FME) [35] smoothened the former's manifold for DR. Likewise, L1-Semi [36] constructed graphs by replacing the Gaussian weighting with for spectral embedding as the innate label details were neglected during propagation. Similarly, the margin-based discriminant embedding [37] constructed the non-negative sparse graph (NNSG) using Euclidean-tuned Laplacian that achieved discriminative non-linear data projection. NNSGs increased constructional complexity coupled with local information negligence led to the model failure. Few other embedding models aligning with NNSG such as FDEFS [38], DLA [39], GCSE [40], and DFEFP [41] efficiently projected FVs in low dimensional space for data classification. The strength of the available labeled information determined these models' efficacy while their performance deteriorated when dealing with large datasets. Alternatively, [42] constructed graphs using an energy-based distance metric that neglected the data distribution. However, the highly time-consuming likelihood-learning failed to surpass the accuracies registered by recent GSSLs.

As evident from the above discussions, the labeling accuracy of contemporary models highly relies on the availability of labeled data, and the distance measure utilized for graph structuring. Moreover, the increased propagation complexity is directly related to the dataset size and complex learning modules incorporated into their structurization. Furthermore, their operation in crisp space lessens flexibility and adaptability thereby downing accuracies, rather, these characteristics are deemed essential. Thus, from the literature reviewed above, the following objectives are formulated to fill the gaps in the usage of GSSL for LP and FS aiding data classification.

- Design a parameter-free dynamic graph structuring model replacing the widely popular and parametric Gaussian, KNN graphs. Also, the intended graph model should be fair, trustworthy, and interpretable which is lacking in the conventional CG.
- Develop a new distance measure demonstrating resilience toward noise and outliers experienced by the traditional Euclidean metric [21], [24].
- Realize a unique graph regularization model to optimize the similarity matrix easing LP.
- Implement a simple LP algorithm performing effective information propagation with limited supervised labels.
- Adopt an optimal DR technique that downs the model's computational burden without compromising its performance.

To meet these research objectives the following contributions are made in this proposal:

- Introduced a parameter-free FG approach combining scalability, dynamicity, and interpretability into the graph structure that overcomes the limitations of predefined CGs [22], [23].
- Developed a novel fuzzy distance measure similar to geodesic distance, enhancing interclass separability with increased intraclass affinity.
- Improved FG learning using a SoftMax-tuned cross-entropy regularizer bonding intraclass FVs with heightened LP performance.
- Decomposition of optimal similarity matrix favoring efficient prediction of unknown labels followed by label learning using linguistically advanced soft computing technique offering improved classification even with scarcely available labeled FVs.
- Reduce FV dimensions with realization complexity performed in two stages:
  - Extraction of scaled NHE-based fuzzy FVs using a probability-tuned Membership Function retaining desired data variations;
  - Projection matrix learning using a Lasso regularized fuzzy entropy objective function addressing overfitting with improved interpretability.

To realize the aforementioned contributions mathematically the related notations utilized in their formulation and along with the characteristics are dealt below.

B. Mathematical Notations

A summary of mathematical notations denoting the diverse variables used for realizing the proposed model is presented in Table I.

The rest of this paper is organized as follows: Fuzzy fundamentals in the context of set theory, feature scaling, and FG learning are dealt with in Section II. In Section III, learning validity on real-world data sets is relatively analyzed in detail followed by complexity and interpretability investigations in Section IV. Finally, Section V, summarizes the novelty and efficacy of the proposed work.

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2023.3323093

TABLE I
MATHEMATICAL NOTATIONS WITH THEIR DESCRIPTIONS

Notations	Description
$\mathcal{X}$	Dataset
n	Number of FVs in $\mathcal{X}$
m	Length of the FV
f	Arbitrary FV
L	Number of levels in the FV f
$\|g(x)\|$	Magnitude of an arbitrary function g
$\|\|g(x)\|\|_2$	l_2- norm of an arbitrary function g
N_m	Number of pixels with a particular intensity level (histogram counts)
$H(f)$	Normalized Histogram Envelope (NHE)
$\mu_h(x)$	NHE in fuzzy space
$\|$	Set notation for 'such that'
$\mathcal{X}_l$	labeled data point set
$\mathcal{X}_u l$	unlabeled data point set
y_l	Vector with class label information for first l FVs
d_{ij}^x	Distance matrix calculated on $\mathcal{X}$ for FVs x_i & x_j
σ_{x_i} & σ_{x_j}	Standard deviation of i_{th} and j_{th} FV
$\mathcal{G}$	Fuzzy Graph (FG)
$\mathcal{V}$	Vertices of FG
μ_e & μ_v	Membership Functions edges and vertices in FG
S_{ij}	Similarity Matrix
w_{ij}	Projection Matrix
$\mathcal{L}$	Lagrangian Function
S_{ii}	Diagonal elements in the similarity matrix
$\gamma,\ \lambda$	Regularization parameter
$\eta,\ \eta_1,\ \beta$	Lagrangian Multipliers

II. METHODOLOGY

To address the research objectives formulated from the shortcomings of standard GSSL elucidated in the literature review, a novel FGLP is introduced in this section which is expected to be interpretable for enabling downstream applications [42]. Existing research mainly focuses on the inferred graph by circumventing each node's vicinity based on existing metrics [44]. In contrast, the proposed model constructs a graph by defining edges based on membership values generated by the distance matrix in the fuzzy space. As the degree of membership lies in the range $[0\ 1]$, the edge formation is completely based on the probability of closeness, thereby, making the graph highly interpretable. The graphical representation of the proposed LP is shown in Fig. 1.

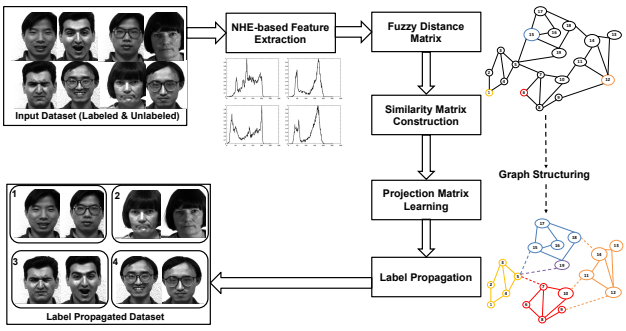


Fig. 1. A Fuzzy Graph-based Label Propagation Model

At the onset, the scaled fuzzy FVs are extracted by the Normalized Histogram Envelope (NHE) by exploiting the inherent flexibility and adaptability qualities of fuzzy logic. FG is then structured using the novel FDM to ease data categorization followed by the construction of the optimal similarity matrix on the resultant data by fitting a simple cost function. Later the higher dimensional FVs are projected to lower dimensional space using Lasso Regularization (LR)-based minimization tuned by fuzzy entropy for optimal FS. Finally, a unified framework is realized addressing similarity learning, and FS followed by LP. To better understand the fuzzy mathematics involved in developing different modules of the FGLP framework, firstly, a brief introduction to fuzzy sets is provided followed by discussions on other modules.

A. Fuzzy Sets

Fuzzy rule-based system modeling bestows easy interpretability [45], rather their accuracies diminish on par with their peers. Therefore, fuzzy variables with minimal rules are desired for making a trade-off between interpretability and accuracy [46]. Generally, these systems are equivalent to FG defined as granular representations of functional dependencies and relations [47]. FG with lesser nodes represents a fuzzy system with a minimalistic rule set offering improved interpretability while maintaining consistent accuracy. Additionally, the highly flexible nature of fuzzy-based learning supplements labeling with reduced computations when compared to its peers [48]–[50] thereby, motivating FG's engagement in this learning framework. Accordingly, the following fuzzy definitions are adopted to understand the presented learning model:

Definition 1: A fuzzy set A in universal set X is mathematically represented as a set of ordered pairs [47], [48] defined in (1)

$$A = \{(x, \mu_A(x)) \,|x \in X\} \,|\forall x \in \mathbb{R},\ \ \mu_A(x) \in [0\ \ 1] \qquad (1)$$

$\mu_A(x) \rightarrow$ degrees of membership of x concerning A; $\mu_A^C(x) = 1 - \mu_A(x) \rightarrow$ Fuzzy complement of $\mu_A(x)$ representing the non-membership values of x.

Definition 2: A fuzzy set A is said to be convex if it satisfies the condition in (2)

$$A(\alpha x + (1-\alpha) y) \geq \min\{\mu_A(x), \mu_A(y)\};\ \ \forall \alpha \in [0\ \ 1] \qquad (2)$$

$\alpha \rightarrow$ constant that cuts the membership function and is hence termed as $\alpha-$cut which is necessary for determining the convexity of the membership function.

Definition 3: If A and B are the fuzzy sets from universes X and Y, then the relation R between the sets A and B is a cartesian product represented as $R \rightarrow A \times B$, and the membership function of the relation R is given in (3)

$$\mu_R(x, y) = \mu_{A \times B}(x, y) = \min(\mu_A(x), \mu_B(y)) \qquad (3)$$

B. NHE-based Fuzzy Feature Extraction with Scaling

Large-sized FVs are a serious problem in computer vision, data mining, machine learning, and pattern recognition and require greater computational effort and storage space [51], [52]. Therefore, to reduce their space and process complexity, this work prescribes a straightforward approach to capture the highly prioritized local variations. Accordingly, each FV denoted by f of size $1 \times m$ with levels f_m lying in the range $[0\ L)\,|L \ll m$, are fabricated into histograms. These histograms represent the probability distribution of levels in terms of N_m defined by the count of each level in f to extract its envelope termed normalized histogram envelope (NHE) ($H(f)$) given in (4)

$$H(f) = \frac{N_m}{m} \qquad (4)$$

$H(f)$ obtained from (4) formulates membership function ($\mu_h(x)$) as in [43] for converting the crisp FVs $\{f = f_m | 0 \leq f_m < L\}$ to

fuzzy FV with its elements representing the degrees of membership [48]–[50]. The extracted FVs are constrained by the number of elements in $\mu_h(x)$ that is very much less than the number of elements in f, thereby reducing the feature dimensions and the computational cost. Accordingly, the transformed $H(f)$ is given in (5)

$$\mu_h(x) = \mu(H(f)) \,|\mu_h(x) \in [0\ \ 1] \tag{5}$$

$\mu_h(x)$ constitutes a dataset $\mathcal{X} = \{x_1, x_2, x_3, ..., x_l, x_{l+1}, ..., x_n\}$, and x_i represents a fuzzy FV. Among n data points only the first l samples are labeled as $y_l = \{1, 2, 3, ..., c\}$ with the remaining considered as unlabeled ul bounded between $l \ll ul$, and $n = l + ul$ as defined in (6) and (7)

$$\mathcal{X}_l = \{(x_1, y_1), (x_2, y_2), ..., (x_l, y_l)\} \tag{6}$$

$$\mathcal{X}_u l = \{x_{l+1}, ..., x_{l+u}\} \tag{7}$$

This arrangement facilitates the construction of a $n \times n$ matrix defining the affinity between nodes in the graph $\mathcal{G}$. Node affinity is quantified using the initial similarity matrix $\{S = s_{ij} | 0 \leq s_{ij} \leq 1\}$ using an appropriate distance measure d on $\mathcal{X}$. According to GSSL if $d(x_i, x_j)$ is very small then the labels $y_i \approx y_j$ and this assumption holds for both labeled and unlabeled data as defined in (8).

$$\mathcal{X} = \{\mathcal{X}_l, \mathcal{X}_{ul}\} = \{x_i\}; i = 1, 2, ..., n \tag{8}$$

C. Fuzzy Graph (FG) Structuring

The relation between fuzzy FVs $\mu_h(x)$ in the dataset $\mathcal{X}$ play an important role in LP, and to determine these relations $\mathcal{G}$ is constructed with each node representing a fuzzy FV or a fuzzy set $\mu_h(x) | x \in X$. FG is formally given in Definition 4.

Definition 4: The FG [53] $\mathcal{G} = (V, \alpha, \mu)$, is a non-empty vertex set $\mathcal{V}$ defined by the membership values of vertices $\{\alpha | \mathcal{V} \in [0\ 1]\}$ and edges $\{\mu | \mathcal{V} \times \mathcal{V} \in [0\ 1]\}$, $\forall (\mathrm{p}, \mathrm{q}) \in \mathcal{V}$ satisfying the condition given in (9)

$$\mu(p, q) \leq \min\{\alpha(p), \alpha(q)\} \tag{9}$$

where $\alpha(p)$, $\mu(p, q)$ are the membership values of the vertex p and edge pq in $\mathcal{G}$ respectively as shown in Fig. 2.

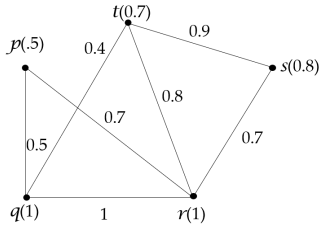


Fig. 2. An example of a Fuzzy Graph

Definition 5: A CG $\mathcal{G}^*$ is a special case of an FG with vertices $\{\alpha | V = 1\}$ and edges $\{\mu | V \times V = 1\}$.

Definition 6: An FG $G = (V, \alpha, \mu)$ is said to be a strong graph if if it satisfies (10)

$$\mu(p, q) = \min\{\alpha(p), \alpha(q)\} \tag{10}$$

FG adopted in this work corresponds to fuzzy relations with induced interpretability. Also, FG structuring is simpler when compared to the conventional CG as the data relations are determined using the probability-based membership values lying in the range $[0\ 1]$. To better understand the significance of swapping CG with FG, a few distinguishing points are given below:

- Conventional GSSLs consider graphs as stationary observations [25], [26]. When the graph is assumed to be random, the additional correlations between data characteristics, labels, and structure may be extracted from the joint distribution. Accordingly, the proposed FGLP considers various forms of uncertainty connected to the data and class/label information. This quality induces dynamicity into the graph for learning the relation between FVs.
- Likewise, Graph regularization using GSSLs demands detailed investigations particularly in the scaling aspect when the input size is large. Especially, many such models available in the literature either concentrate on reducing the computational complexity or improving the propagation accuracy. Rather in the proposed FGLP, almost half of the work is done by graph structuring as the node affinities are determined by considering the local variations in the data on a probability basis.
- Also, existing GSSLs completely trust the available subjective label information which is highly prone to misclassification [25], [26]. Instead, FGLP predicts the probability of two nodes sharing the same label to minimize misclassification.

The aforementioned advantages motivate the replacement of CG with FG and are adopted in this work for label learning. Later, the node affinity of $\mathcal{G}$ is determined by constructing the FDM of $n \times n$ dimensions from the dataset $\mathcal{X}$ as discussed below.

1) Fuzzy Distance Matrix (FDM): GSSLs accuracy is highly influenced by the local connectivity or affinity between FVs [11]. If there is no local linearity in a neighborhood then the Euclidean distance-based affinity fails to accurately represent the relationship between data points. Furthermore, the affinity between data points is impacted by the data distribution, which is mandatory while constructing the affinity matrix [21]. To address this issue, a new FDM (d) is proposed fulfilling the properties of conventional geodesic distance measures. Initially for FDM construction, the Euclidean metric evaluating the node similarity is used by rationalizing the samples using the standard deviation (σ) given in (11) to localize their variations, as they remain highly distinctive.

$$d_{ij}^{x} = \frac{1}{n}\left\{\left|\sum_{i,j=1}^{n}\left(\frac{x_i}{\sigma_{x_i}} - \frac{x_j}{\sigma_{x_j}}\right)^2 - \max\left(\frac{x_i}{\sigma_{x_i}} + \frac{x_j}{\sigma_{x_j}}\right)\right|\right\}^{\frac{1}{2}} \tag{11}$$

Equation (11) calculates the distance between the FVs in dataset $\mathcal{X}$. The first term in (11) represents the Euclidean distances between the FVs normalized by their respective standard deviations $\sigma_{x_i}, \sigma_{x_j}$.This normalization aids in reducing the effect of outliers [54], [55]. While the second term minimizes the intraclass heterogeneity and increases the interclass FVs deviations in the dataset $\mathcal{X}$. To constrain the values of fuzzy geodesic distance obtained in (11) the resultant is then normalized by adopting the fuzzification procedure [56]–[58]. A numerical instance of the fuzzification in attaining FDM is demonstrated in Example 1 of the Appendix by considering the fuzzy features defined in (12), with each row corresponding to FV elements packed in the interval $[0\ 1]$.

$$\mathcal{X} = \begin{bmatrix} 0.1 & 0 & 0.5 & 0.8 & 0.3 \\ 0.5 & 0.8 & 0 & 0.3 & 0.1 \\ 0.1 & 0 & 0 & 0.5 & 0.9 \end{bmatrix} \tag{12}$$

The FDM constructed from the FVs $\mathcal{X}$ in (12) using (11) is presented in (13)

$$d_{ij}^{x} = \begin{bmatrix} 0 & 1 & 0.4 \\ 1 & 0 & 0.7 \\ 0.4 & 0.7 & 0 \end{bmatrix} \tag{13}$$

Equations (11) and (13) outline a few mathematical facts related to the FDM with its corollaries presented below.

Corollary 1: The fuzzy geodesic distance $d(a,b)$ between two non-empty vectors a, b is always positive for all $0 \leq a, b \leq 1$

$$0 \leq d(a,b) \leq 1;\ \forall\ (a,b) \in [0\ 1] \tag{14}$$

Corollary 2: The developed FDM $d(a,b)$ satisfies the separability property of cartesian geodesic distances

$$d(a,b) = 0\ iff\ \ a = b;\ \ \forall\ (a,b) \in [0\ 1] \tag{15}$$

Corollary 3: The FDM in (11) satisfies the symmetry property of conventional geodesic distance

$$d(a,b) = d(b,a);\ \ \forall\ (a,b) \in [0\ 1] \tag{16}$$

Corollary 4: For any fuzzy FVs $a,\ b,\ c \mid a \subseteq b \subseteq c$, then the developed FDM satisfies triangular inequality.

$$d(a,b) \leq \max\{d(a,c), d(b,c)\} \tag{17}$$

Corollary 5: The proposed distance metric is equal to classical geodesic distance if the vectors a and b are in crisp form.

Corollaries 1 to 5 assist in understanding the data affinity captured by FDM from which FG is constructed. FGs' choice is mainly attributed to their ability in dealing with uncertain and ambiguous data existing in many real-world phenomena [59], whilst, the crisp counterparts obscure the ambiguity, interpretability, and validity [25]. The above-demonstrated FG construction process is extended to FVs in the dataset $\mathcal{X}$, wherein each FV corresponds to a vertex with d_{ij}^x representing the edges.

2) Similarity Learning: The similarity matrix (S_{ij}) constructed using the relation (18) should satisfy the positivity, symmetry, and separability properties.

$$S_{ij} = 1 - d_{ij}^x \tag{18}$$

According to GSSL [2], [11], [60], if the distance between the FVs x_i and x_j quantified in d_{ij}^x is large then their similarities in S_{ij} is reduced. Also, the diagonal elements of S_{ij} are zeroed for eliminating the self-loops in FG ($S_{ii} = 0$). Lastly, S_{ij} is normalized using the SoftMax to constrain the row-wise sum to 1 $\left(\sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} = 1\right)$ with the resultant equivalently interpreted as a probability distribution matrix. Application of the SoftMax function improvises the intraclass firmness and separability in the features of interclass data points [61]. To optimize S_{ij} , the formulated FG and the realized d_{ij}^x are combined by tuning the overall cost function (19) covering the aforementioned constraints.

$$\begin{aligned} J &= \min_{S_i} \sum_{i,j=1}^{n} S_{ij}.d_{ij}^x + \gamma \sum_{i,j=1}^{n} S_{ij} \cdot \log \frac{S_{ij}}{d_{ij}^x} \\ &\forall i, S_i \geq 0, S_{ii} = 0\ \&\ \sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} = 1 \end{aligned} \tag{19}$$

$\gamma-$ Regularization parameter.

The first term in (19) represents the loss function responsible for determining similarities between the samples, while, the second corresponds to the cross-entropy determining the dissimilarity measure in data points. In combination with SoftMax, this cross-entropy term achieves better interclass separability [62], [63]. Minimizing (19) produces S_{ij} in terms of d_{ij}^x defined in (20)

$$\frac{dJ}{dS_{ij}} = 0 \Rightarrow d_{ij}^x + \gamma + \gamma \log \frac{S_{ij}}{d_{ij}^x} = 0$$

$$S_{ij} = d_{ij}^x.e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)} \tag{20}$$

Reorganizing (19) in vector form using (20) given in [33] produces (21)

$$\begin{aligned} J &= \min_{S_i} \left\| S_{ij} - d_{ij}^x.e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)} \right\|_2^2 \\ &\forall i,\ S_i \geq 0,\ S_{ii} = 0\ \&\ \sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} = 1 \end{aligned} \tag{21}$$

The minimization problem in (21) is solved using the Lagrangian function given in (22)

$$\begin{aligned} \mathcal{L}(S_{ij}, \eta, \beta_i) =& \frac{1}{2}\left(S_{ij} - d_{ij}^x \cdot e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)}\right)^2 \\ &- \eta\left(\sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} - 1\right) - \beta_i S_i \end{aligned} \tag{22}$$

$\eta,\ \beta\ \geq\ 0$ are the Lagrangian multipliers. Solving (22) for $S_{ij},\ \eta,\ and\, \beta$ yields (23), (24), and (25).

$$\begin{aligned} \frac{d\mathcal{L}}{dS_{ij}} = 0 \Rightarrow &\left(S_{ij} - d_{ij}^x \cdot e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)}\right. \\ &\left. - \eta \sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} \left(1 - \sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}}\right) - \beta_i\right. \end{aligned} \tag{23}$$

$$\frac{d\mathcal{L}}{d\eta} = 0 \Rightarrow \sum_{j=1}^{n} \frac{e^{S_{ij}}}{\sum_{j=1}^{n} e^{S_{ij}}} = 1 \tag{24}$$

$$\frac{d\mathcal{L}}{d\eta} = 0 \Rightarrow S_i = 0 \tag{25}$$

The optimal S_{ij} satisfying Karush–Kuhn–Tucker (KKT) conditions is determined using [11], [16], [33] constrained by the parameters $\eta,\ \gamma$ is determined using (26)

$$S_{ij} = \left(-d_{ij}^x e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)} + \eta\left(\frac{n-1}{n^2}\right)\right)_+ \tag{26}$$

To determine $\eta,\ \gamma$, a sparse S_{ij} is constructed using (23) – (26) with the consideration that u non-zero elements (closest neighbors) of FG are selected adaptively based on known labels in the dataset represented by $S_{iu} > 0$ and the remaining $n - u$ elements are denoted as $S_{iu+1} \leq 0$. Later, S_{ij} is sorted in the ascending order to select the first u elements, and the rest are zeroed. This process is mathematically presented in (27), (28)

$$S_{iu} = d_{iu}^x \cdot e^{-\left(1+\frac{d_{iu}^x}{\gamma}\right)} + \eta\left(\frac{n-1}{n^2}\right) > 0 \tag{27}$$

$$S_{iu+1} = d_{iu+1}^x \cdot e^{-\left(1+\frac{d_{iu+1}^x}{\gamma}\right)} + \eta\left(\frac{n-1}{n^2}\right) \leq 0 \tag{28}$$

Also from (26), the value of the Lagrangian multiplier η is obtained by exploiting the constraint that the row-wise sum of S_{ij} is unity.

$$\sum_{j=1}^{u} d_{ij}^x \cdot e^{-\left(1+\frac{d_{ij}^x}{\gamma}\right)} + \eta\left(\frac{n-1}{n^2}\right) = 1 \tag{29}$$

Upon simplifying (29) the Lagrangian multiplier η for u closest neighbors is determined using (30)

$$\eta = \left[\sum_{j=1}^{u} d_{ij}^{x} \cdot e^{-\left(1+\frac{d_{ij}^{x})}{\gamma}\right)} - 1 \right] \frac{n^2}{u(n-1)} \tag{30}$$

Substituting (30) in (25) produces (31) from which the regularization parameter γ is calculated as in (32)

$$d_{i,u+1}^{x} \cdot e^{-\left(1+\frac{d_{i,u+1}^{x}}{\gamma}\right)} + \eta \left(\frac{n-1}{n^2} \right) = 0 \tag{31}$$

$$\gamma = \frac{\sum_{j=1}^{u} d_{ij}^{x} - d_{iu}^{x}}{1 + u - \sum_{j=1}^{u} \log\left(d_{ij}^{x}\right)} \tag{32}$$

Finally, by substituting (30), (32) in (22), the S_{ij} is obtained in (33)

$$S_{ij} = \begin{cases} d_{ij}^{x} \cdot s_1 + \frac{1}{u} - \frac{1}{u} \sum_{j=1}^{u} d_{ij}^{x} s_1, & \text{for} \quad j \leq u \\ 0, & \text{for} \quad j > u \end{cases} \tag{33}$$

where $S_1 = e^{\left(1+\frac{d_{ij}^{x}\left(1+u-\sum_{j=1}^{u}\log\left(d_{ij}^{x}\right)\right)}{\sum_{j=1}^{u} d_{ij}^{x} - d_{ij}^{x}} \right)}$.

D. Feature Selection (FS) and Projection Matrix Learning

In the case of high dimensional FVs, the computational complexity in learning the graph similarity is large, to compensate for this, FS is incorporated into the framework, which additionally improves its interpretability. Also, FS plays an important role in retaining or dropping FV elements based on their significance in the learning perspective. Accordingly, in this contribution, it is highly essential to assess the importance of data attributes extracted from the heterogeneous datasets and establish FGLPs' uniformity in dealing with diverse applications.

Accordingly, FS is performed using a novel fuzzy entropy-based optimization function tuned by LR for constructing the projection matrix w_{ij} of dimensions $m \times \omega$. The rationale is to exploit DR through FS which is accomplished by the LR's penalty term (responsible for the selection of relevant features) with simultaneous reduction of FV dataset dimensions from $n \times m$ to $n \times \omega$, supplemented by the introduced importance score, thereby, emphasizing the role of data attributes in classification.

Accordingly, the projection matrix w_{ij} for minimizing FV size is recursively tuned using (34)

$$\begin{aligned} &\min_{w_{ij}} \sum_{j=1}^{n} v_j \left\| x_j w_{ij} - p_j \right\|_2^2 + \lambda \sum_{j=1}^{n} \left\| w_{ij} \right\| \| \\ &w_{ij} \geq 0; \sum_{j=1}^{n} \frac{e^{w_{ij}}}{\sum_{j=1}^{n} e^{w_{ij}}} = 1 \end{aligned} \tag{34}$$

$x_j \rightarrow j^{th}$ FV with importance score v_j; $\lambda \rightarrow$ penalty parameter; $p_j \rightarrow$ the predicted label of x_j, determined from the pseudo labels matrix comprising of available labels and predicted labels represented as $P = \begin{pmatrix} P_l \\ P_{ul} \end{pmatrix}$. To ensure error-free w_{ij} the predicted labels are initially assumed zero ($P_{ul} = 0$).

The first term in (34) is the cost function approximating w_{ij} while the second corresponds to the penalty responsible for FS. The first term adjusted by the importance score v_j is formulated using the fuzzy entropy to capture the FVs uncertainty and potentially enhances FS by up-sizing or downsizing the FV length during learning. Also, it effectively distinguishes the supervised and unsupervised samples thereby, offering accurate label prediction. The second term in (34) biases optimization by effectively removing less informative or redundant features favoring representational sparsity [64]. λ's magnitude directly influences FS, if high very less features are selected while lower values result in the selection of more features thereby leading to underfitting or overfitting. Therefore, striking a balance in adaptively fitting λ is determined using the Lagrangian multiplier that overcomes the aforesaid issues with simultaneous reduction of FV dimensions. Accordingly, (34) is minimized to produce w_{ij} in vector form as in (35)

$$w_{ij} = \frac{2 v_j p_j - \lambda}{2 v_j x_j} \tag{35}$$

Lagrangian representation of the lasso problem in (34) is simplified in (36) using (35)

$$\mathcal{L}\left(w_{ij}, \eta_1\right) = \frac{1}{2} \left(w_{ij} - \frac{2 v_j p_j - \lambda}{2 v_j x_j^2} \right)^2 - \eta_1 \left(\sum_{j=1}^{n} \frac{e^{w_{ij}}}{\sum_{j=1}^{n} e^{w_{ij}}} - 1 \right) \tag{36}$$

The positive definite Lagrangian multiplier $\eta_1 > 0$ is determined by following the procedure adopted in evaluating S_{ij}followed by the construction of w_{ij} meeting KKT conditions and stated in (37)

$$w_{ij} = \left(-\frac{2 v_j p_j - \lambda}{2 v_j x_j^2} - \eta_1 \left(\frac{n-1}{n^2} \right) \right)_+ \tag{37}$$

η_1 in (37) is determined utilizing the constraint $\sum_{j=1}^{n} \frac{e^{w_{ij}}}{\sum_{j=1}^{n} e^{w_{ij}}} = 1$, for u closest neighbors and presented in (38)

$$\eta_1 = \left[\sum_{j=1}^{u} \frac{2 v_j p_j - \lambda}{2 v_j x_j} - 1 \right] \frac{n^2}{u\left(n-1\right)} \tag{38}$$

Finally, the penalty parameter λ is determined upon substituting (38) in (37) with the consideration that u non-zero elements (closest neighbors) of FG are selected adaptively based on unknown labels in the dataset and stated in (39)

$$\lambda = -\left[\frac{u - u \sum_{j=1}^{n} \left(\frac{p_j}{x_j} \right) - \frac{p_j}{x_j}}{\frac{1}{2 v_j x_j} - \frac{u}{\sum_{j=1}^{n} v_j x_j}} \right] \tag{39}$$

Substituting λ, η_1 in (35) produces the optimal projection matrix that helps in addressing the dimensionality issues in GSSL. The aforesaid mathematical process is outlined in Algorithm 1.

Algorithm 1 Algorithm for Feature Selection

Input:

1) Scaled and Fuzzified Dataset $\mu_h\left(x\right)$ comprising of l labeled FVs, and ul unlabeled FVs constrained as $l \ll ul$.
2) Pseudo Label Matrix $P = \begin{pmatrix} P_l \\ P_{ul} \end{pmatrix}$; $P_l = y_l$ is the known label matrix and $P_{ul} = 0$ are the pseudo predicted labels.

Output: Projection Matrix w_{ij} with reduced dimensions $m \times \omega$

Step 1: Compute the fuzzy entropy (Importance Score v_j)
Step 2: Determine the Lagrangian and LR parameters λ, η_1 meeting KKT conditions
FGLP framework for LP
for $\forall$ FVs in $\mathcal{X}$ **do**
Step 3: Update w_{ij} using (35)
end for

FS algorithm introduced above reduces the dimensions of the dataset $\mathcal{X}$ and S_{ij}, is utilized for label learning as discussed below.

E. Label Learning

Labeling unlabeled samples present in dataset $\mathcal{X}$ depends on the constructed FG, the information of labeled nodes, and the membership values of graph edges capitulated in S_{ij}. The optimized S_{ij} upon FS strictly confines the learning process to u nearest neighbors by eliminating outliers to acquire highly correlated FVs, thereby improving the classification accuracy. Later, to propagate the label x_l to x_{ul} in $\mathcal{X}$, a simple prediction matrix $P = \begin{pmatrix} P_l \\ P_{ul} \end{pmatrix}$ is formulated by decomposing S_{ij} as in (40)

$$\begin{pmatrix} P_l \\ P_{ul} \end{pmatrix} = \begin{pmatrix} s_{l,l} & s_{l,ul} \\ s_{ul,l} & s_{ul,ul} \end{pmatrix} \begin{pmatrix} y_l \\ 0 \end{pmatrix} \tag{40}$$

where $s_{l,l}$, $s_{l,ul}$, $s_{ul,l}$, $s_{ul,ul}$ are the local similarity features of labeled–labeled, labeled–unlabeled, unlabeled–labeled, and unlabeled–unlabeled FVs respectively. Using (40), the predicted labels for unlabeled FVs are determined using (41)

$$P_{ul} = s_{ul,l} y_l \tag{41}$$

FG learns the local similarity between the FVs based on easily accessible label information corresponding to the $s_{l,l}$ components before LP using (42)

$$s_{l,l} = \begin{cases} s_{ij}, & \text{if } y_i \text{ and } y_j \text{ are known and } i = j \\ 0, & \text{Otherwise} \end{cases} \tag{42}$$

The labeled–labeled similarity feature component $s_{l,l}$ stated in (42) helps in grouping S_{ij} into c block diagonal elements [33], thereby, eliminating the need for additional graph-cutting, hence boosting LP. Later, the label information P_{ul} for unlabeled FVs with u non-zero row elements is predicted using (43)

$$P_{ul} = \begin{cases} |p_i - p_j|, & s_{ul,l} \neq 0 \\ 0, & \text{Otherwise} \end{cases} \tag{43}$$

where $i \in [0, l]$ & $j \in [l+1, n]$ with u constraining the search range. This FGLP arrangement propagates label information y_l of $s_{l,l}$ directly to the first u element in each column of non-zero elements in $s_{ul,l}$. After completion of each class LP, the j value is incremented to $j + u$ to search for a new region, and the label information is propagated to the complete dataset using the optimized S_{ij}.

All the operations such as label learning, similarity, and projection matrix learning are done simultaneously. The mechanism discussed above is outlined in Algorithm 2 and also, mathematical explanation for an instance of the FV dataset is provided in Example 2 of the Appendix.

Finally, the rendered FDM is capable of handling multi-labels owing to the compaction of intraclass FVs with simultaneous decorrelation of interclass FVs an attribute owed to the standard deviation-based normalization of the FVs presented in (11). This discrimination characteristic is extended to the construction of optimal S_{ij} for handling the multiclass data. Also, when combined with LP, it warrants consistent scores for classifying multiclass or multi-label data by exploring the effect of label information, correlation, and local graph structure with the criteria that there should be at least one labeled FV for each class. To demonstrate the multiclass handling ability of FGLP, it is rigorously validated and relatively analyzed on the UCI and a few other datasets in Section III.

Algorithm 2 Optimization Algorithm for FGLP

Input:

1) Dataset $\mathcal{X}_{n\times m} = \{\mathcal{X}_l, \mathcal{X}_{ul}\} = \{x_i\} \mid l \ll ul$ for $i = 1, 2, ..., n$ where $\mathcal{X}_l$ corresponds to l labeled FVs, $\mathcal{X}_{ul}$ corresponds to ul unlabeled FVs, and $n = l + ul$.
2) Known Label Matrix y_l

Output:

1) Projection Matrix w_{ij} of dimensions $m \times \omega$
2) Predicted Labels $P = \begin{pmatrix} P_l \\ P_{ul} \end{pmatrix}$; where P_l corresponds to available labels, and P_{ul} corresponds to learned/predicted labels

Step 1: NHE-based fuzzy feature extraction with scaling
Step 2: Compute the fuzzy entropy (Importance Score v_j)
Step 3: Determine FDM d_{ij}^x using (11), for the dataset $\mathcal{X}$
Step 4: Determine the similarity Matrix $S_{ij} = 1 - d_{ij}^x$; $\mid S_{ii} = 0$ and set $w_{ij} = 0_{m\times\omega}$; where ω is the size of FV after FS
Step 5: Determine the following parameters by meeting KKT conditions
- η, η_1 from (30) and (38)
- γ, λ from (32) and (39)

Step 6: Initialize pseudo label matrix P with given labels P_l by treating P_{ul} as zeros
for $\forall$ classes in $\mathcal{X}$ **do**
  Step 7: Update S_{ij} using (33)
  Step 8: Update w_{ij} using (35)
  Step 9: Update P using (43)
end for

III. Results and Discussion

A. Experimental Settings

The performance of the proposed FGLP on different datasets on par with the traditional and recent contemporaries is investigated using relevant validation metrics and elaborated in the below sections.

1) Dataset Description: FGLPs effectiveness is examined by conducting numerous tests on heterogeneous real-world datasets in the context of LP for data categorization and FS. Especially the model is evaluated on 23 UCI datasets [65] with 14 being multiclass datasets and the other 9 corresponding to binary datasets. To analyze the suitability of the proposed model when extended to real-time applications, along with UCI datasets, a few other image, text, and digit heterogeneous-natured datasets are additionally engaged. The details of these datasets with the number of classes and size are briefly described in Table II.

The datasets engaged for investigation are composed of diverse entities with different categories defined with a varying number of FVs, lengths, and the number of classes as listed in Table II. Columns 4 and 5 listed under "Length of FV" in Table II demonstrate the potential of the introduced NHE-based feature extraction that has downed the FV length significantly, especially for the face, object, scene, and text datasets. Particularly, the data redundancy in FVs belonging to these datasets is well exploited by the introduced NHE-based feature extraction.

2) Implementation & Comparison Approaches: FGLPs efficacy is investigated by analyzing the test results obtained for the 32 diverse datasets listed in Table II with those obtained using various baseline and prevailing GSSL models like LGC [27], GFHF [28], MFA [29], SDA [34], FME [35], L1-Semi [36], NNSG [37], S^2LRR [1], AWSSL [11], ANSSL [11], FDEFS [38], DLA [39], SSWTDS [31], SPGO [30], LPSGL [33], GCSE [40], and DFEFP [41]. The traditional graph-based models LGC [27], and GFHF [28], attained the affinity matrix using the Gaussian weighting that requires the tuning of hyper-parameter σ for structuring. Also, the diagonal elements of the affinity matrix are set to zero to avoid self-looping in the graph. Similarly for SDA and FME, KNN graphs were constructed

TABLE II
DESCRIPTION OF DATASETS

S. No.	Dataset	# Samples (FVs)	Length of FV		# Classes
			Before Scaling	After Scaling	
1	Breast	699	9	9	2
2	German	1000	20	20	2
3	Ionosphere	351	34	34	2
4	Pima	768	8	8	2
5	Vote	435	16	16	2
6	Crx	690	15	15	2
7	Heart	270	13	13	2
8	Isolet	1560	617	256	2
9	Monkl	432	6	6	2
10	Balance	147	4	4	3
11	Dermatology	120	34	34	6
12	E-Coli	100	7	7	5
13	Glass	54	9	9	6
14	Iris	150	4	4	3
15	Seeds	210	7	7	3
16	Soybean	40	35	35	4
17	Wine	144	13	13	3
18	Zoo	28	16	16	7
19	Multi-Features	2000	216	216	10
20	Page Blocks	140	10	10	5
21	Optical Digits	1566	64	64	9
22	Segments	2310	19	19	7
23	Pen Digits	3024	16	16	9
24	ORL	400	644	256	40
25	MSRA25	1799	256	256	12
26	Yale	165	1024	256	15
27	Coil-20	1440	1024	256	20
28	USPS	1854	1024	256	10
29	8 Sports	1040	1024	256	8
30	Scene 15	1950	1024	256	15
31	Dig1-10	1797	64	64	10
32	CANE-9	1080	856	256	9

with k being set to 10, and edge weights are calculated utilizing the Gaussian kernel. The hyper-parameters for NNSG are taken from the set $\left\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1\right\}$ and for SPGO and LPSGL from $\left\{10^{-6}, 10^{-5}, ..., 10^{5}, 10^{6}\right\}$ similar to [33]. Also, the importance score and regularization parameters for LPSGL and DFEFP are fixed and taken from the set $\left\{1, 10, 10^{2}, 10^{3}, 10^{4}\right\}$ respectively. As the proposed framework is parameter-free, herein, the importance score is tuned by the fuzzy entropy, and the value of u is varied dataset-wise based on the number of unlabeled FVs in a class.

B. Relative Analysis

Apart from the parameter setting, the quantity of randomly selected labeled and unlabeled samples from the dataset remains the same for all methodologies used for relative analysis. Herein, the mean classification accuracy along with standard deviation over 10 random splits are reported for uniform relative analysis. Accordingly, if x_i is the unlabeled FV from the dataset $\mathcal{X}$, and g_i, y_i are the predicted and ground truth labels, then the task of labeling is set as in (44) and the model's accuracy is assessed using (45)

$$\delta\left(x_i, g_i\right) = \begin{cases} 1 & if\, g_i = y_i \\ 0 & elsewhere \end{cases} \tag{44}$$

$$A\left(y, g\right) = \frac{\sum_{i=l+1}^{n} \delta\left(x_i, g_i\right)}{n} \tag{45}$$

1) Label Propagation Models: To evaluate FGLPs proficiency, tests were widely conducted on 23 datasets representing binary and multiclass data from the UCI machine learning repository with the achievements relatively recorded with the trending and conventional predecessors. Initially, the FGLP model operates on binary datasets, and the achieved outcomes are rated with the graph-based models [11]. Subsequently, a maximum of 10% of FVs are selected as seed labels by fixing u as the number of unlabeled images per class. As the FVs count in every class is different, the labeled information is fixed for each class and is graphically illustrated in Fig. 3.

The accuracy curves presented in Fig. 3 outperform their counterparts which are owed to the fuzzy-based edge and node characterization ensuring interpretability, flexibility, and adaptability irrespective of the diverse-natured datasets. Moreover, the probability-based edge formation in FGs plays an important role in LP with the propagation of only the first u elements, leading to a reduction in the false classification. To demonstrate the inherent characteristics of FGLP aligning with the aforesaid qualities this analysis is performed as the data intricacies of multiclass UCI datasets are extremely high in comparison with the binary counterparts. Accordingly, the efficacy analysis is extended with the labeled information confined between 10% to 70% followed by the achieved accuracy tabulation in Table III along with the recent semi-supervised multi-category classification [31].

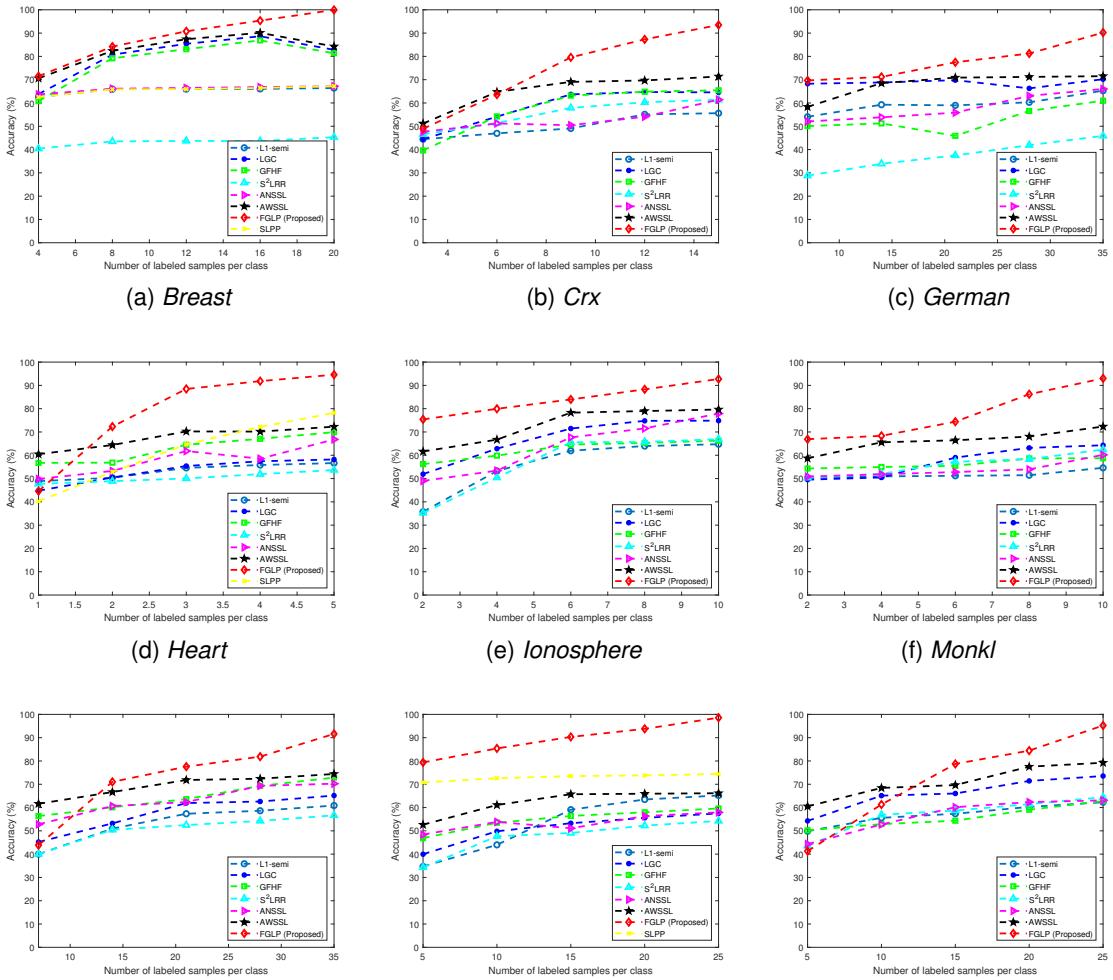
FGLPs accuracies recorded in Table III demonstrate its effectiveness in the context of available labels. If the label information is more than 30% of the dataset size FGLP dominates both linear and non-linear versions of SSL and even with 10% labels its performance matches with the non-linear SSWTDS. This dominance is owed to FDM which bounds the data distribution within the standard deviation thereby, reducing outliers. Also, the fuzzy-based characterization of edges and nodes in structuring FG confines their membership value in the range $[0\ \ 1]$ ensuring interpretability.

FGLPs' suitability and applicability in real-world applications are investigated by engaging heterogeneous datasets namely MSRA25, ORL, Coil-20, and USPS consisting of face and object images along with text and digit datasets like CNAE-9 and Digi-1-10. Specifically, for the text dataset CNAE-9 the FVs are extracted utilizing term frequency and inverse document frequency (TFIDF) followed by NHE-based processing and later subjected to accuracy assessment. The accomplishments are compared with the recent graph-based LP models [30], [33]–[36] and reported in Table IV.

Table IV numerically depicts the dominance of FGLP with the other 8 models, while, the case of single labeled sample in Dig 1-10 and Coil-20 declines due to the global scaling nature of NHE which neglects data intricacies if the labels are very less thereby, insisting the need for adaptive FV scaling constrained by local data characteristics. Rather, FGLPs accuracy surpasses its peers when the labeled samples were increased. However, a steady increase in the accuracy is witnessed in the other datasets and remains very high even with less labeled information which is owed to FG and the similarity matrix. Also, the tuned similarity matrix combining SoftMax with cross-entropy ensured swift data adaptability that increases propagation accuracy. As the overall model building and processing is done in the fuzzy domain, the data values are constrained in $[0\ \ 1]$ which makes the prediction and propagation model more flexible by prioritizing edge formation and determining the nearest neighbor at ease.

2) Embedding Models: Additionally, the analysis is extended to a few other image datasets like 8 sports, Scene 15, Coil-20, and Extended YALE from the scene, object, and face categories and compared with the recent embedding models [38]–[41] in Table V.

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2023.3323093



(a) *Breast* (b) *Crx* (c) *German*

(d) *Heart* (e) *Ionosphere* (f) *Monkl*

(g) *Isolet* (h) *Pima* (i) *Vote*

Fig. 3. Relative comparison of 9 different two-class UCI datasets

Accuracies registered by the proposed model have outperformed the others as illustrated in Table V. This quality is attributed to the standard deviation-based normalization and the max term in (11) that had promised better accuracies by reducing outliers and ensuring interclass separability in FVs. Also, the cross-entropy-based cost function with SoftMax is another major reason for increasing propagation accuracy.

3) Ranking by Statistical Testing: To justify these relative analyses additionally, a pairwise analysis using a non-parametric statistical hypothesis test namely the Wilcoxon signed ranked test was conducted on all the datasets. To test the hypothesis a significance level (α) is assumed 0.05 and later inferences were made based on the obtained probability values $(p-values)$ [66]. The choice $\alpha = 0.05$ denotes the labeled sample size which is fair for relative analysis and if decreased adds to the sample strength which makes LP complex. Also, if the attained $p-value$ for a pair of models in a dataset is less than α then it indicates that there is a difference in the models' performance outlining the superiority over the other. Table VI presents the Wilcoxon signed rank results $(p-values)$ corresponding to the datasets and models from Table III.

The superiority of FGLP is witnessed in Table VI over its peers with the attained $p-values$ (row-wise) lesser than α. Also, a graphical ranking based on Wilcoxon's notations is given as $FGLP > Non-Linear\ SSWTDS > Linear\ SSWTDS$. The ranking test is extended to the other LP [30], [33]–[35], [37] and embedding [38]–[41] methodologies with FGLP and tabulated in Table VII, and Table VIII respectively.

From Table VII, it is inferred that the FGLP is performing better when compared to all its predecessors accordingly the models are ranked as $FGLP > LPSGL > FME > SPGO\_LRR > SDA > SPGO\_ADP > SPGO\_SSC > SPGO\_KNN$.

The same superiority is observed in Table VIII with the models ranked as $FGLP > GCSE > DFEFP > DLA > FDEFS$. From these Wilcoxon tests, it is inferred that the performance of the proposed FGLP is consistent with different types of data, irrespective of the number of classes, and size of the dataset. This consistency is owed to the optimal similarity matrix, which is responsible for decorrelating the interclass FVs by maintaining intraclass firmness that makes LP effective.

C. Effect of Feature Selection

To investigate the strength of the introduced FS in Section II-D, it is specifically validated on the Coil20, ORL, and, YALE image datasets as their FVs are larger in dimensions compared with the FVs of the other datasets. Relatively, the experiments are conducted by randomly selecting 10% of the labeled FVs from the dataset followed by the selection of the top $\{10\%,\ 30\%,\ 50\%,\ 70\%,\ 90\%\}$ features to maintain uniformity with its peers [67]–[69], and presented in Table IX.

TABLE III
RELATIVE ANALYSIS OF MULTICLASS CLASSIFICATION RESULTS ATTAINED USING FGLP (BOLD CORRESPONDS TO BEST, AND ITALICS CORRESPOND TO SECOND BEST)

Dataset	Accuracy (Mean ± Standard Deviation) for 10 Folds with different labeled data samples (in %)											
	10%			30%			50%			70%		
	Linear SSWTDS	Non-Linear SSWTDS	FGLP	Linear SSWTDS	Non-Linear SSWTDS	FGLP	Linear SSWTDS	Non-Linear SSWTDS	FGLP	Linear SSWTDS	Non-Linear SSWTDS	FGLP
Balance	**88.16 ± 2.29**	*87.84 ± 3.95*	82.34 ± 2.19	88.32 ± 2.84	**90.66 ± 2.69**	*89.84 ± 2.73*	88.48 ± 2.81	*91.04 ± 2.52*	**92.16 ± 0.54**	88.48 ± 2.81	*92.16 ± 2.88*	**95.32 ± 1.25**
Dermatology	*87.44 ± 5.72*	**88.25 ± 5.08**	85.17 ± 3.84	*95.10 ± 4.75*	**96.21 ± 4.65**	92.67 ± 2.74	*97.36 ± 3.56*	**97.74 ± 3.18**	95.41 ± 0.38	**98.50 ± 2.60**	97.94 ± 2.16	*98.41 ± 0.38*
Ecoli	*81.66 ± 6.83*	78.59 ± 7.41	**81.80 ± 1.30**	82.25 ± 6.35	*85.54 ± 5.29*	**88.40 ± 1.95**	82.57 ± 6.87	*85.65 ± 6.97*	**91.40 ± 2.19**	83.32 ± 7.79	*86.31 ± 7.20*	**94.80 ± 2.39**
Glass	45.70 ± 6.57	*56.58 ± 8.11*	**81.11 ± 3.80**	59.82 ± 9.02	*62.85 ± 9.02*	**87.04 ± 1.31**	61.68 ± 9.45	*63.51 ± 7.01*	**92.59 ± 2.62**	63.21 ± 9.83	*66.24 ± 7.19*	**95.93 ± 1.22**
Iris	84.00 ± 6.44	**90.00 ± 12.67**	*84.40 ± 6.48*	*93.66 ± 4.49*	**96.00 ± 3.44**	88.27 ± 8.91	**96.00 ± 4.66**	*96.00 ± 3.44*	92.93 ± 1.92	97.00 ± 2.89	*97.33 ± 3.44*	**97.60 ± 2.43**
Seeds	*88.29 ± 4.07*	**88.57 ± 7.84**	86.37 ± 2.28	*93.27 ± 4.66*	**94.88 ± 4.59**	92.13 ± 4.16	93.81 ± 3.92	*95.24 ± 4.49*	**96.81 ± 2.25**	94.18 ± 3.26	*95.63 ± 3.71*	**98.31 ± 1.01**
Soybean	*94.00 ± 9.66*	**100**	88.00 ± 2.09	*98.00 ± 6.32*	**100**	93.00 ± 0.51	*98.00 ± 6.32*	**100**	97.18 ± 5.41	98.68 ± 1.32	**100**	*99.58 ± 3.41*
Wine	92.12 ± 8.48	**97.78 ± 3.88**	*92.85 ± 2.72*	*98.04 ± 3.65*	97.78 ± 3.88	**98.61 ± 1.63**	98.82 ± 3.72	*98.33 ± 3.75*	**99.05 ± 2.07**	99.26 ± 1.91	*99.26 ± 1.91*	**100**
Zoo	70.00 ± 10.54	**88.18 ± 9.03**	*87.50 ± 2.53*	85.00 ± 7.07	*96.00 ± 5.26*	**97.60 ± 2.43**	92.00 ± 10.33	*97.00 ± 4.83*	**98.00 ± 2.12**	93.00 ± 7.28	*97.00 ± 6.75*	**98.54 ± 1.72**
Multiple Features	57.65 ± 6.84	62.90 ± 5.26	**64.20 ± 8.58**	85.10 ± 2.88	*88.75 ± 2.02*	**91.05 ± 1.41**	97.05 ± 1.21	*97.20 ± 1.23*	**97.85 ± 1.63**	97.75 ± 1.32	*98.10 ± 0.68*	**98.40 ± 1.95**
Page Blocks	91.63 ± 1.34	**93.02 ± 1.41**	*92.56 ± 3.13*	93.09 ± 1.19	*93.97 ± 1.28*	**94.93 ± 1.22**	93.09 ± 1.19	*93.97 ± 1.28*	**96.16 ± 3.23**	93.09 ± 1.19	*94.90 ± 1.15*	**97.41 ± 0.58**
Optical Digits	*89.57 ± 2.57*	**92.33 ± 1.17**	87.56 ± 2.73	90.66 ± 2.15	**98.00 ± 0.76**	*94.38 ± 4.10*	91.58 ± 2.26	**98.19 ± 0.63**	*98.17 ± 1.06*	92.36 ± 2.65	*98.55 ± 0.71*	**99.71 ± 0.21**
Segment	58.57 ± 11.46	*76.52 ± 13.19*	**77.85 ± 0.73**	78.85 ± 9.92	*86.58 ± 5.19*	**95.34 ± 3.29**	86.67 ± 6.66	*88.88 ± 2.50*	**96.73 ± 1.5**	87.26 ± 4.89	*89.62 ± 2.81*	**98.57 ± 1.42**
Pen Digits	87.85 ± 0.74	**89.65 ± 1.12**	*89.48 ± 0.73*	*91.25 ± 2.65*	91.16 ± 0.87	**92.22 ± 2.26**	92.01 ± 2.24	*93.62 ± 2.21*	**95.2 ± 2.38**	93.05 ± 1.28	*94.49 ± 0.77*	**98.57 ± 1.42**

From Table IX, it is evident that the proposed label learning process attains higher accuracies even at reduced dimensions. Irrespective of the FS percentage FGLP dominates its predecessors in the ORL and YALE datasets and particularly in Coil20 for the 30% case it supersedes others. However, for the other cases in Coil20, FGLP closely tails top performers and falls marginally behind them. Upon observing the dominators in Coil20, it can be seen that they are majorly KNN-based FS schemes whose fixed choice of K has influenced the achievement, whereas they largely decline in the other FS cases to the FGLP-based FS in the other datasets which is an indication of the optimization introduced across every stage of FS. Upon analyzing experiments, it is further inferred that to have a better balance between the FV size, and accuracy, at least 10% of features should be selected whereas no such significance is noticed in its competitors. Overall, it is claimed that the introduced fuzzy entropy clearly distinguished the labeled FVs from unlabeled FVs. Also, the introduced lasso penalty tuned by the Lagrangian assists in efficient learning of the thereby effectively minimizing the FV size without compromising on the accuracy. Further to justify the model's simplicity and amicability for real-world scenarios computational complexities along the time and space are dealt with in Section IV.

IV. COMPLEXITY AND INTERPRETABILITY ANALYSIS

The computational complexity of the presented method is described with the $\mathcal{O}-$ for analyzing the incurred time and space. The presented method's complexity considers the mathematical steps involved in different stages dealing with feature extraction followed by learning of similarity, projection, and label matrices. In contrast, most SSL approaches involve individual data training and testing that highly escalates these complexities. The suggested methodology's complexity is detailed in Sections IV-A and IV-B.

A. Time Complexity

Formulating time complexity commences with the extraction of NHE-based feature extraction based on the data entity being handled. For instance, when dealing with images, the n images of the dataset (face dataset and handwritten numbers) having the dimensions $p \times q$ incur the time complexity prescribed in (46) to compute the NHE

$$\mathcal{O}\left(n\left(pq+b\right)\right) \tag{46}$$

$b-$ Number of bins.

Similarly, for n text documents firstly, the TFIDF features are extracted in $\mathcal{O}\left(ng\log\left(ng\right)\right)$, where g is the length of sequence or $N-$gram. Later, NHE scales these TFIDF features that incur $O\left(n\left(a+b\right)\right)$ operations, where a is the number of elements in the TFIDF vector. Thus, the total time required for feature scaling of text documents is stated in (47)

$$O\left(n\left(a+b\right)\right)+O\left(ng\log\left(ng\right)\right)\approx O\left(n\left(a+b\right)\right) \tag{47}$$

Therefore, the computational time for calculating the distance between every FV with all FVs in the dataset requires $\mathcal{O}\left(nm\right)$ operations. $n-$ Number of FVs in the data, $m-$ Number of features for each FV. The complexity rises to $\mathcal{O}\left(n^2m\right)$ when constructing the distance matrix from the FVs. As the proposed work considers only, L features for data distinction based on their intensity variations, hence, the computational time for determining d_{ij}^{x} is given in (48)

$$\mathcal{O}\left(n^2L\right) \tag{48}$$

For calculating the similarity matrix this computational time is again reduced to a very large extent based on u nearest neighbors constrained by $u \ll L$ and presented in (49)

$$\mathcal{O}\left(n^2u\right) \tag{49}$$

This article has been accepted for publication in IEEE Transactions on Fuzzy Systems. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/TFUZZ.2023.3323093

TABLE IV
PERFORMANCE COMPARISON ON FACE, OBJECT, TEXT, AND DIGIT DATASETS (BOLD CORRESPONDS TO BEST, AND ITALICS CORRESPOND TO SECOND BEST)

Type	Dataset	Labeled Samples	Accuracy (Mean ± Standard Deviation) for 10 Folds							
			FGLP	LPSGL (2021)	SPGO_KNN (2020)	SPGO_SSC (2020)	SPGO_LRR (2020)	SPGO_ADP (2020)	NNSG (2015)	FME (2010)	SDA (2007)
Face	MSRA25	1	*80.0 ± 2.3*	**81.4 ± 4.5**	69.0 ± 5.9	72.4 ± 4.7	72.2 ± 5.8	69.2 ± 7.0	59.8 ± 4.4	56.5 ± 4.8	56.5 ± 5.1
		2	**93.4 ± 0.4**	*91.8 ± 4.5*	82.1 ± 4.5	85.7 ± 5.3	85.5 ± 4.7	83.1 ± 5.5	80.9 ± 4.5	77.6 ± 7.0	78.0 ± 5.6
		3	**96.2 ± 2.6**	*95.9 ± 2.5*	90.0 ± 5.1	89.5 ± 3.8	91.6 ± 4.2	89.9 ± 4.1	89.6 ± 3.9	88.1 ± 6.2	88.7 ± 4.2
	ORL	1	**71.5 ± 3.1**	*69.8 ± 4.3*	52.4 ± 4.5	59.0 ± 4.1	59.2 ± 3.8	64.0 ± 2.4	58.7 ± 3.9	47.8 ± 4.1	43.0 ± 5.2
		2	**81.0 ± 0.3**	*80.5 ± 3.5*	63.7 ± 4.4	73.1 ± 3.4	73.4 ± 4.3	73.0 ± 4.1	76.1 ± 3.8	66.2 ± 3.5	66.5 ± 4.8
		3	**90.0 ± 3.2**	*86.9 ± 3.2*	73.9 ± 4.3	77.3 ± 4.1	79.9 ± 3.6	78.1 ± 3.8	82.4 ± 5.5	75.3 ± 5.3	79.6 ± 4.1
Object	Coil-20	1	78.0 ± 3.1	**81.4 ± 2.3**	77.0 ± 4.8	71.8 ± 4.1	65.9 ± 3.6	*79.7 ± 3.0*	76.2 ± 2.5	67.4 ± 3.0	65.2 ± 2.6
		2	**93.2 ± 2.3**	*84.4 ± 1.7*	79.1 ± 2.9	75.2 ± 2.0	70.9 ± 2.9	81.3 ± 2.6	80.0 ± 2.6	77.0 ± 2.5	74.4 ± 2.5
		3	**95.1 ± 2.6**	*87.3 ± 1.7*	84.0 ± 2.8	80.0 ± 3.5	79.3 ± 3.6	85.6 ± 2.2	83.6 ± 1.7	80.8 ± 2.2	79.2 ± 1.7
	USPS20	1	**74.2 ± 1.9**	*69.6 ± 6.0*	33.7 ± 4.2	39.0 ± 4.9	56.7 ± 5.6	37.3 ± 4.6	57.8 ± 6.8	63.3 ± 4.6	61.4 ± 4.1
		2	**81.3 ± 2.1**	*78.7 ± 3.3*	41.9 ± 5.1	40.0 ± 3.8	67.7 ± 5.1	42.5 ± 4.7	65.4 ± 3.9	71.4 ± 3.9	71.5 ± 3.9
		3	**89.8 ± 5.1**	*83.0 ± 2.1*	45.5 ± 3.7	44.7 ± 4.1	71.5 ± 3.7	45.9 ± 2.9	72.6 ± 3.0	76.1 ± 3.7	76.2 ± 3.1
Text	CNAE-9	1	**79.1 ± 3.1**	*70.0 ± 3.8*	38.0 ± 4.8	31.1 ± 5.9	45.0 ± 9.2	36.7 ± 4.7	30.7 ± 7.5	51.8 ± 5.9	50.5 ± 8.1
		2	**83.5 ± 2.8**	*77.0 ± 4.0*	47.0 ± 6.6	41.4 ± 6.6	54.0 ± 3.4	46.9 ± 5.2	46.3 ± 8.1	68.0 ± 5.5	59.2 ± 8.7
		3	**88.9 ± 3.6**	*80.0 ± 3.6*	51.6 ± 4.2	45.9 ± 5.3	56.8 ± 3.9	52.7 ± 4.6	56.9 ± 7.1	72.4 ± 3.9	67.1 ± 7.0
Hand Written Digit	Dig 1-10	1	*78.7 ± 0.4*	**79.9 ± 4.9**	44.1 ± 3.6	47.9 ± 5.0	67.1 ± 5.4	45.6 ± 4.3	64.5 ± 4.7	74.8 ± 4.9	68.4 ± 5.2
		2	**89.1 ± 5.1**	*86.5 ± 2.2*	51.8 ± 4.7	53.7 ± 4.6	75.2 ± 4.9	55.8 ± 3.7	74.0 ± 4.4	79.2 ± 4.2	78.8 ± 4.1
		3	**93.8 ± 3.4**	*88.8 ± 1.5*	58.3 ± 4.3	59.3 ± 4.9	79.6 ± 2.3	59.0 ± 4.1	79.4 ± 3.3	82.2 ± 3.0	82.3 ± 2.4

TABLE V
PERFORMANCE COMPARISON ON SCENE, OBJECT, AND FACE DATASETS (BOLD CORRESPONDS TO BEST, AND ITALICS CORRESPOND TO SECOND BEST)

Type	Dataset	Labeled Samples (in %)	Accuracy (Mean) for 10 Folds				
			FGLP	DFEFP (2022)	GCSE (2021)	DLA (2020)	FDEFS (2019)
Scene	8-Sports	50	66.35	**67.41**	*67.04*	65.05	65.3
		70	**73.08**	*70.6*	69.97	66.98	68.5
	Scene-15	50	**76.67**	67.07	*67.26*	64.15	66.18
		70	**88**	69.52	*70.36*	67.9	66.75
Object	Coil-20	10	**96.96**	95.1	*96.6*	94.55	94.03
		20	**99.93**	98.2	*99.5*	96.46	93.26
Face	Extended Yale	20	92.12	**96.05**	*94.74*	93.75	94.14
		40	95.52	*98.32*	**98.53**	97.81	97.2

TABLE VI
WILCOXON SIGNED RANK TESTS WITH $\alpha = 0.05$ FOR THE METHODS FROM TABLE III

	Linear SSWTDS	Non-Linear SSWTDS	FGLP
FGLP	0.0119	0.0332	1
Non-Linear SSWTDS	0.0005	1	0.6922
Linear SSWTDS	1	0.9996	0.9899

TABLE VII
WILCOXON SIGNED RANK TESTS WITH $\alpha = 0.05$ FOR THE LP METHODS (IN TABLE IV)

	SDA	FME	NNSG	SPGO ADP	SPGO LRR	SPGO SSC	SPGO KNN	LPSGL	FGLP
FGLP	0.018	0.018	0.018	0.018	0.018	0.018	0.018	0.018	1
LPSGL	0.018	0.018	0.018	0.018	0.018	0.018	0.018	1	0.990
SPGO KNN	0.953	0.953	0.953	0.982	0.953	0.663	1	0.990	0.990
SPGO SSC	0.953	0.953	0.990	0.982	0.970	1	0.417	0.990	0.990
SPGO LRR	0.050	0.896	0.050	0.015	1	0.015	0.017	0.990	0.990
SPGO ADP	0.896	0.929	0.896	1	0.896	0.030	0.030	0.990	0.990
NNSG	0.663	0.799	1	0.147	0.583	0.018	0.071	0.990	0.990
FME	0.018	1	0.265	0.104	0.147	0.071	0.071	0.990	0.990
SDA	1	0.990	0.042	0.015	0.206	0.047	0.047	0.990	0.990

TABLE VIII
WILCOXON SIGNED RANK TESTS WITH $\alpha = 0.05$ FOR THE EMBEDDING METHODS (IN TABLE V)

	FDEFS	DLA	GCSE	DFEFP	FGLP
FGLP	0.0106	0.0106	0.0291	0.0291	1
DFEFP	0.05	0.05	0.7081	1	0.8193
GCSE	0.05	0.05	1	0.4276	0.8193
DLA	0.5724	1	0.9777	0.9777	0.9498
FDEFS	1	0.5724	0.9777	0.9777	0.9498

Similarly w_{ij} learning with u nearest neighbors requires $\mathcal{O}(n\omega u)$ with $n \times \omega$ being its dimensions. Further, the incorporated optimization model downscales the computing time to finally achieve the total time (T_{total}) for LP using FG given in (50).

$$T_{total} = \mathcal{O}(n(pq+b)) + \mathcal{O}\left(n^2 L\right) + \mathcal{O}\left(n^2 u\right) + \mathcal{O}(n\omega u)$$

$$T_{total} \approx \mathcal{O}\left(n^2 L\right) \tag{50}$$

To signify the reduced realization time of FGLP additionally, LP experiments are performed on MATLAB R2021a running on Intel(R) Core i3-6100U CPU @ 2.30GHz and 8GB RAM with the outcomes relatively presented in Table X along with time complexities.

The complexities of the widely popular and recent methods are extremely higher in comparison with FGLP as witnessed in Table X. FGLPs' reduced complexity highly benefits implementation thereby, signifying the models' simplicity, which is an essential quality desired for real-time extensions.

TABLE IX
EFFECT OF FS ON PROPAGATION ACCURACY

Methodology/ Datasets	Accuracy with various FV lengths for 10 random splits with 10% of known labels														
	10%	30%	50%	70%	90%	10%	30%	50%	70%	90%	10%	30%	50%	70%	90%
	Coil20					ORL					YALE				
SemiFree-CART	85.876	90.46	87.42	88.2	86.57	55.56	55	56.22	58.62	58.57	39.79	*54.88*	50.03	44.94	43.34
SemiFree-KNN	**96.23**	95.71	*95.34*	94.78	94.66	*71.36*	*75.01*	*76.97*	*77.58*	77.12	41.86	*54.88*	*53.41*	50.22	51.65
SFS-CART	89.466	89.11	88.35	88.16	88.03	36.99	46.68	48.93	55.26	54.34	28.31	36.53	34.94	39.97	39.85
SFS-KNN	93.77	95.76	95.12	**95.96**	**95.15**	71.2	69.92	75.22	75.22	*79.02*	*46.71*	45.12	51.53	*53.37*	*54.96*
AGLRM-CART	82.456	86.43	86.45	87.57	87.14	42.6	50.05	57.55	57.6	55.61	31.63	41.38	38.5	39.91	38.07
AGLRM-KNN	90.525	*96.03*	**95.69**	*95.15*	*94.73*	68.32	73.16	76.81	77.02	76.81	41.74	51.65	50.06	51.74	51.53
FGLP (Proposed)	*96.04*	**96.74**	94.65	92.85	93.68	**94.25**	**93.00**	**95.50**	**92.75**	**91.50**	**74.36**	**76.36**	**79.39**	**76.97**	**78.18**

TABLE X
COMPARISON OF THE COMPUTATIONAL COMPLEXITY OF DIFFERENT METHODS

Methodology	Computational Complexity	CPU Runtime (milliseconds)
MFA [29]	$\mathcal{O}\left(n^3\right)$	725
SDA [34]	$\mathcal{O}\left(d^3\right)$	898
FME [35]	$\mathcal{O}\left(2cn^2\right)$	915
NNSG [37]	$\mathcal{O}\left(\begin{array}{c}\left(n^3+\max\left\{n^3,n^2c\right\}+in^3\right)t\\+(ndc)+\left(2d^2n+d^3+n^2d+n^3\right)\end{array}\right)$	4698
SPGO [30]	$\mathcal{O}\left(\max\left\{n^3,d^3,clu\right\}t\right)$	1088
LPSGL [33]	$\mathcal{O}\left(\max\left\{n^2,u^3\right\}t\right)$	962
GNMFLD [70]	$\mathcal{O}\left(t\left(mnk\right)+n^2m+lk\right)$	2518
FGLP (Proposed)	$\mathcal{O}\left(n^2L\right)$	657

B. Space Complexity

Similar to the time complexity, the space complexity of the proposal is also determined across all stages. In the first stage, the extracted NHE-based features (for n images or text documents) occupy $\mathcal{O}\left(nL\right)$ space. Later, FDM requires $\mathcal{O}\left(n^2\right)$ space for storing the similarity matrix. Thus, the overall space occupancy(S_{total}) is evaluated in (51)

$$S_{total} = \mathcal{O}\left(nL\right) + \mathcal{O}\left(n^2\right)$$
$$S_{total} \approx \mathcal{O}\left(n^2\right) \tag{51}$$

The evaluated time and space complexities in (50) and (51) of the proposed model are in quadratic order and are highly less in comparison with its peers having higher orders. Also, it is established that despite the computational reduction, the method does not compromise the classification accuracy, thereby, making it more favorable for real-time heterogeneous data classification.

C. Interpretability Analysis

The projection matrix w_{ij} learning minimizes the FV size, reducing computational complexity thereby increasing the FGLP's interpretability as both complement each other. Therefore, a mathematical metric defining interpretability index I of the fuzzy system is introduced in [43] is engaged in this work and presented in (52)

$$I = 1 - Q_F \tag{52}$$

Q_F representing the overall fuzzy complexity and is determined using (53)

$$Q_F = \frac{Q_{rules} + Q_{FS} + Q_{inputs}}{3} \tag{53}$$

Herein, Q_F completely rely on FG structuring encompassing the number of fuzzy relations for edge formations characterized by the number of fuzzy rules and sets. Based on these discussions Q_F is finally reduced to (54)

$$Q_F = \frac{Q_{rules} + Q_{FS}}{2} \tag{54}$$

Wherein Q_{rules} is the complexity related to the number of rules firing one at a time and Q_{FS} is the ratio of the number of fuzzy sets utilized at a time (fs_i) to the total number of fuzzy sets (T_{fs}), as presented in (55) and (56) respectively

$$Q_{rules} = \frac{1}{T_{rules}} \tag{55}$$

$$Q_{FS} = \frac{fs_i}{T_{fs}} \tag{56}$$

$T_{rules} = \frac{n(n-1)}{2}$ correspond to the total number of rules utilized in decisioning the edge formation and T_{fs} represents n fuzzy sets each of length L. By considering these factors, and utilizing 2 fuzzy sets at a time, Q_F is finally reorganized in (57)

$$Q_F = \frac{\frac{2}{n^2-n} + \frac{2L}{n}}{2}$$

$$Q_F = \frac{Ln^2 - Ln + n}{n^3 - n^2} \tag{57}$$

To better understand FGLPs reliability and expressiveness, the developed interpretability measure is examined on 8 datasets presented in Table XI.

TABLE XI
INTERPRETABILITY ANALYSIS

Dataset	# Samples (FVs)	Length of FV		I	
		Before Scaling	After Scaling	Before Scaling	After Scaling
Isolet	1560	617	256	0.6045	0.8359
USPS	1854	1024	256	0.4477	0.8619
Wine	144	13	13	0.9097	0.9097
Seeds	210	7	7	0.9666	0.9666
Monkl	432	6	6	0.9861	0.9861
8 Sports	1040	1024	256	0.6045	0.8359
Scene 15	1950	1024	256	0.0154	0.7538
CNAE-9	1080	856	256	0.4749	0.8687

From Table XI, it is evident that the interpretability involved in FG structuring is highly dependent on dataset dimensions, especially the reduction in FV size that led to the hike in interpretability thereby outlining FGLPs enhanced fairness in LP.

V. Conclusion

This research proposed a simple yet versatile FG-based paradigm for LP. The key contribution is the creation of the FDM that assists in grouping intraclass samples and deviating interclass data samples which is extremely crucial in the design of the FG for LP. Later to attain the optimal similarity matrix, the newly framed cost function engages the SoftMax function coupled with cross-entropy to offer wider data deviations in the interclass FVs. The coalescence of these techniques makes it more efficient than its predecessors, as evidenced in the accuracy analyses by a factor of 20% done on diverse heterogeneous datasets. Furthermore, the complexity analysis strongly demonstrates the appropriateness of this model for real-time applications without sacrificing performance. Although FGLP showcases higher accuracy values with and without FS even with lesser label information, the model's computational complexity increases with the dataset size. Also, it is observed that the model's accuracy decay if the labeled information is less in large datasets. Moreover, the interpretability, flexibility, and adaptability characteristics of FG require deep exploration for accelerating simultaneous learning to reduce the model's computational complexity.

References

[1] Li, C., Lin, Z., Zhang, H. & Guo, J. Learning Semi-Supervised Representation Towards a Unified Optimization Framework for Semi-Supervised Learning. *2015 IEEE International Conference On Computer Vision (ICCV)*. pp. 2767-2775, 2015.

[2] Ma, J., Xiao, B. & Deng, C. Graph based semi-supervised classification with probabilistic nearest neighbors. *Pattern Recognition Letters*. vol.133 pp. 94-101, 2020.

[3] Chen, L., Ravichandran, V. & Stolcke, A. Graph-based label propagation for semi-supervised speaker identification. *Interspeech 2021*. 2021, https://www.amazon.science/publications/graph-based-label-propagation-for-semi-supervised-speaker-identification

[4] Vanegas, J., Escalante, H. & González, F. Scalable multi-label annotation via semi-supervised kernel semantic embedding. *Pattern Recognition Letters*. vol.123 pp. 97-103, 2019.

[5] Seeger, M. Learning with labeled and unlabeled data. 2000.

[6] Dunlop, M., Slepčev, D., Stuart, A. & Thorpe, M. Large data and zero noise limits of graph-based semi-supervised learning algorithms. *Applied And Computational Harmonic Analysis*. vol.49, pp. 655-697, 2020.

[7] Chong, Y., Ding, Y., Yan, Q. & Pan, S. Graph-based semi-supervised learning: A review. *Neurocomputing*. vol.408 pp. 216-230 2020.

[8] He, F., Wang, R. & Jia, W. Fast semi-supervised learning with anchor graph for large hyperspectral images. *Pattern Recognition Letters*. vol.130 pp. 319-326 2020.

[9] He, F., Nie, F., Wang, R., Hu, H., Jia, W. & Li, X. Fast Semi-Supervised Learning With Optimal Bipartite Graph. *IEEE Transactions On Knowledge And Data Engineering*. vol.33, pp. 3245-3257, 2021.

[10] Liu, Q., Sun, Y., Wang, C., Liu, T. & Tao, D. Elastic Net Hypergraph Learning for Image Clustering and Semi-Supervised Classification. *IEEE Transactions On Image Processing*. vol.26, pp. 452-463, 2017.

[11] Nie, F., Shi, S. & Li, X. Semi-Supervised Learning with Auto-Weighting Feature and Adaptive Graph. *IEEE Transactions On Knowledge And Data Engineering*. vol.32, pp. 1167-1178 2020.

[12] Nie, F., Xiang, S., Jia, Y. & Zhang, C. Semi-supervised orthogonal discriminant analysis via label propagation. *Pattern Recognition*. vol.42, pp. 2615-2627, 2009.

[13] Elhamifar, E. & Vidal, R. Sparse Subspace Clustering: Algorithm, Theory, and Applications. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. vol.35, pp. 2765-2781, 2013.

[14] Ma, L., Crawford, M., Yang, X. & Guo, Y. Local-Manifold-Learning-Based Graph Construction for Semisupervised Hyperspectral Image Classification. *IEEE Transactions On Geoscience And Remote Sensing*. vol.53, pp. 2832-2844, 2015.

[15] Wan, J., Chen, H., Li, T., Sang, B. & Yuan, Z. Feature Grouping and Selection With Graph Theory in Robust Fuzzy Rough Approximation Space. *IEEE Transactions On Fuzzy Systems*. vol.31, pp. 213-225, 2023.

[16] Liu, C., Nie, F., Wang, R. & Li, X. Graph Based Soft-Balanced Fuzzy Clustering. *IEEE Transactions On Fuzzy Systems*. pp. 1-12, 2022.

[17] Samanta, S. & Pal, M. Fuzzy Planar Graphs. *IEEE Transactions On Fuzzy Systems*. vol.23, pp. 1936-1942, 2015.

[18] Zhang, Z., Liu, X. & Wang, L. Spectral clustering algorithm based on improved gaussian kernel function and beetle antennae search with damping factor. *Computational Intelligence And Neuroscience*. vol.2020 pp. 1-9, 2020.

[19] Tron, R. & Vidal, R. A Benchmark for the Comparison of 3-D Motion Segmentation Algorithms. *2007 IEEE Conference On Computer Vision And Pattern Recognition*. pp. 1-8, 2007.

[20] Goldberg, A., Zhu, X. & Wright, S. Dissimilarity in Graph-Based Semi-Supervised Classification. *Proceedings Of The Eleventh International Conference On Artificial Intelligence And Statistics*. vol.2 pp. 155-162 (2007-3-21), https://proceedings.mlr.press/v2/goldberg07a.html

[21] Yadav, R., Verma, S. & Venkatesan, S. Cross-covariance based affinity for graphs. *Applied Intelligence*. vol.51 pp. 3844-3864, 2021.

[22] Cao, Y., Zhou, Z., Hu, C., He, W. & Tang, S. On the Interpretability of Belief Rule-Based Expert Systems. *IEEE Transactions On Fuzzy Systems*. vol.29, pp. 3489-3503, 2021.

[23] Razak, T., Garibaldi, J., Wagner, C., Pourabdollah, A. & Soria, D. Toward a Framework for Capturing Interpretability of Hierarchical Fuzzy Systems—A Participatory Design Approach. *IEEE Transactions On Fuzzy Systems*. vol.29, pp. 1160-1172, 2021.

[24] Pourbahrami, S., Balafar, M., Khanli, L. & Kakarash, Z. A survey of neighborhood construction algorithms for clustering and classifying data points. *Computer Science Review*. vol.38 pp. 100315, 2020.

[25] Zhang, C., Lin, Y., Chen, C., Yao, H., Cai, H. & Fang, W. Fuzzy Representation Learning on Graph. *IEEE Transactions On Fuzzy Systems*. pp. 1-13, 2023.

[26] Xia, F., Sun, K., Yu, S., Aziz, A., Wan, L., Pan, S. & Liu, H. Graph learning: A survey. *IEEE Transactions On Artificial Intelligence*. vol.2, pp. 109-127, 2021.

[27] Zhou, D., Bousquet, O., Lal, T., Weston, J. & Schölkopf, B. Learning with Local and Global Consistency. *Advances In Neural Information Processing Systems*. vol.16, 2003.

[28] Zhu, X., Ghahramani, Z. & Lafferty, J. Semi-Supervised Learning Using Gaussian Fields and Harmonic Functions. *Proceedings Of The Twentieth International Conference On International Conference On Machine Learning*. pp. 912-919, 2003.

[29] Yan, S., Xu, D., Zhang, B., Zhang, H., Yang, Q. & Lin, S. Graph Embedding and Extensions: A General Framework for Dimensionality Reduction. *IEEE Transactions On Pattern Analysis And Machine Intelligence*. vol.29, pp. 40-51, 2007.

[30] Nie, F., Dong, X. & Li, X. Unsupervised and Semisupervised Projection With Graph Optimization. *IEEE Transactions On Neural Networks And Learning Systems*. vol.32, pp. 1547-1559, 2021.

[31] Saigal, P., Rastogi, R. & Chandra, S. Semi-supervised Weighted Ternary Decision Structure for Multi-category Classification. *Neural Processing Letters*. vol.52 pp. 1555-1582, 2020.

[32] Saigal, P., Khanna, V. & Rastogi, R. Divide and conquer approach for semi-supervised multi-category classification through localized kernel spectral clustering. *Neurocomputing*. vol.238 pp. 296-306, 2017.

[33] Wang, F., Zhu, L., Xie, L., Zhang, Z. & Zhong, M. Label propagation with structured graph learning for semi-supervised dimension reduction. *Knowledge-Based Systems*. vol.225 pp. 107130, 2021.

[34] Cai, D., He, X. & Han, J. Semi-supervised Discriminant Analysis. *2007 IEEE 11th International Conference On Computer Vision*. pp. 1-7, 2007.

[35] Nie, F., Xu, D., Tsang, I. & Zhang, C. Flexible Manifold Embedding: A Framework for Semi-Supervised and Unsupervised Dimension Reduction. *IEEE Transactions On Image Processing*. vol.19, pp. 1921-1932, 2010.

[36] Nie, F., Wang, H., Huang, H. & Ding, C. Unsupervised and semi-supervised learning via $l1-$norm graph. *2011 International Conference On Computer Vision*. pp. 2268-2273, 2011.

[37] Fang, X., Xu, Y., Li, X., Lai, Z. & Wong, W. Learning a Nonnegative Sparse Graph for Linear Regression. *IEEE Transactions On Image Processing*. vol.24, pp. 2760-2771, 2015.

[38] Zhu, R., Dornaika, F. & Ruichek, Y. Learning a discriminant graph-based embedding with feature selection for image categorization. *Neural Networks*. vol.111 pp. 35-46 2019.

[39] Zhu, R., Dornaika, F. & Ruichek, Y. Semi-supervised elastic manifold embedding with deep learning architecture. *Pattern Recognition*. vol.107 pp. 107425, 2020.

[40] Dornaika, F. Flexible data representation with feature convolution for semi-supervised learning. *Applied Intelligence*. vol.51, pp. 7690-7704, 2021.

[41] Dornaika, F. & Hoang, V. Deep data representation with feature propagation for semi-supervised learning. *International Journal Of Machine Learning And Cybernetics*. vol.14, pp. 1303-1316, 2023.

[42] Chen, Z., Cao, H. & Chang, K. GraphEBM: Energy-Based Graph Construction for Semi-Supervised Learning. *2020 IEEE International Conference On Data Mining (ICDM)*. pp. 62-71, 2020.

[43] Madhu, C. & M.S., S. Adaptive Bezier Curve-based Membership Function formulation scheme for Interpretable Edge Detection. *Applied Soft Computing*. vol.133 pp. 109968, 2023.

[44] Berton, L., De Paulo Faleiros, T., Valejo, A., Valverde-Rebaza, J. & De Andrade Lopes, A. RGCLI: Robust Graph that Considers Labeled Instances for Semi-Supervised Learning. *Neurocomputing*. vol.226 pp. 238-248, 2017.

[45] Alonso, J., Magdalena, L. & González-Rodríguez, G. Looking for a good fuzzy system interpretability index: An experimental approach. *International Journal Of Approximate Reasoning*. vol.51, pp. 115-134, 2009.

[46] Alcalá, R., Casillas, J., Cordón, O. & Herrera, F. Building Fuzzy Graphs: Features and Taxonomy of Learning for Non-Grid-Oriented Fuzzy Rule-Based Systems. *J. Intell. Fuzzy Syst.*. vol.11, pp. 99-119, 2002.

[47] Zadeh, L. Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets And Systems*. vol.90, pp. 111-127 (1997).

[48] Xu, W., Guo, D., Qian, Y. & Ding, W. Two-Way Concept-Cognitive Learning Method: A Fuzzy-Based Progressive Learning. *IEEE Transactions On Fuzzy Systems*. vol.31, pp. 1885-1899, 2023.

[49] Liu, Y., Miao, Y., Pantelous, A., Zhou, J. & Ji, P. On Fuzzy Simulations for Expected Values of Functions of Fuzzy Numbers and Intervals. *IEEE Transactions On Fuzzy Systems*. vol.29, pp. 1446-1459, 2021.

[50] Miao, Y., Wang, G., Zhou, J., Pantelous, A. & Wang, K. Comments on "Fuzzy Simulations for Expected Values of Functions of Fuzzy Numbers and Intervals". *IEEE Transactions On Fuzzy Systems*. vol.31, pp. 1756-1758, 2023.

[51] Liu, Z., Lai, Z., Ou, W., Zhang, K. & Zheng, R. Structured optimal graph based sparse feature extraction for semi-supervised learning. *Signal Processing*. vol.170 pp. 107456, 2020.

[52] Liu, Z., Wang, J., Liu, G. & Zhang, L. Discriminative low-rank preserving projection for dimensionality reduction. *Applied Soft Computing*. vol.85 pp. 105768 2019.

[53] Rosenfeld, A. FUZZY GRAPHS††The support of the Office of Computing Activities, National Science Foundation, under Grant GJ-32258X, is gratefully acknowledged, as is the help of Shelly Rowe in preparing this paper.. *Fuzzy Sets And Their Applications To Cognitive And Decision Processes*. pp. pp. 77-95, 1975.

[54] Jayalakshmi, T. & Santhakumaran, A. Statistical normalization and back propagation for classification. *International Journal Of Computer Theory And Engineering*. vol.3, pp. 1793-8201, 2011.

[55] Shin, J. & Jeong, J. Multiclass classification of hemodynamic responses for performance improvement of functional near-infrared spectroscopy-based brain–computer interface. *Journal Of Biomedical Optics*. vol.19, pp. 067009-067009, 2014.

[56] Chaira, T. & Ray, A. Construction of fuzzy edge image using Interval Type II fuzzy set. *International Journal Of Computational Intelligence Systems*. vol.7, pp. 686-695, 2014.

[57] Versaci, M. & Morabito, F. Image edge detection: A new approach based on fuzzy entropy and fuzzy divergence. *International Journal Of Fuzzy Systems*. vol.23, pp. 918-936, 2021.

[58] Flores-Vidal, P., Olaso, P., Gómez, D. & Guada, C. A new edge detection method based on global evaluation using fuzzy clustering. *Soft Computing*. vol.23 pp. 1809-1821 2019.

[59] Sitara, M., Akram, M. & Yousaf Bhatti, M. Fuzzy Graph Structures with Application. *Mathematics*. vol.7 2019.

[60] Ng, A., Jordan, M. & Weiss, Y. On spectral clustering: Analysis and an algorithm. *Advances In Neural Information Processing Systems*. vol.14, 2001.

[61] Luo, Y., Wong, Y., Kankanhalli, M. & Zhao, Q. $\mathcal{G}$ -Softmax: Improving Intraclass Compactness and Interclass Separability of Features. *IEEE Transactions On Neural Networks And Learning Systems*. vol.31, pp. 685-699, 2020.

[62] Qin, Z., Kim, D. & Gedeon, T. Rethinking Softmax with Cross-Entropy: Neural Network Classifier as Mutual Information Estimator. 2020.

[63] Zhao, S., Gong, M., Liu, T., Fu, H. & Tao, D. Domain Generalization via Entropy Regularization. *Advances In Neural Information Processing Systems*. vol.33 pp. 16096-16107, 2020.

[64] Hastie, T., Tibshirani, R. & Friedman, J. The elements of statistical learning: data mining, inference, and prediction. *Springer Series In Statistics*. vol.2 pp. 1-758, 2009.

[65] Blake, C. & Merz, C. UCI Machine Learning Repository. 1998, https://archive.ics.uci.edu/ml/index.php

[66] Di Leo, G. & Sardanelli, F. Statistical significance: p value, 0.05 threshold, and applications to radiomics—reasons for a conservative approach. *European Radiology Experimental*. vol.4, pp. 1-8, 2020.

[67] Liu, K., Li, T., Yang, X., Chen, H., Wang, J. & Deng, Z. SemiFREE: Semi-supervised Feature Selection with Fuzzy Relevance and Redundancy. *IEEE Transactions On Fuzzy Systems*. pp. 1-13, 2023.

[68] Li, X., Zhang, Y. & Zhang, R. Semisupervised Feature Selection via Generalized Uncorrelated Constraint and Manifold Embedding. *IEEE Transactions On Neural Networks And Learning Systems*. vol.33, pp. 5070-5079, 2022.

[69] Lai, J., Chen, H., Li, T. & Yang, X. Adaptive graph learning for semi-supervised feature selection with redundancy minimization. *Information Sciences*. vol.609 pp. 465-488, 2022.

[70] Xing, Z., Ma, Y., Yang, X. & Nie, F. Graph regularized nonnegative matrix factorization with label discrimination for data clustering. *Neurocomputing*. vol.440 pp. 297-309, 2021.

Cherukula Madhu is a Research Scholar in the School of Electronics Engineering, VIT, Vellore, India. His research interests include Data Analytics and Interpretable Machine Learning.

Sudhakar MS received the PhD degree in information and communication engineering from Madras Institute of Technology, Anna University, India, in 2014. He is currently serving the School of Electronics Engineering, Vellore Institute of Technology, India as an Associate professor with. His research interests include machine learning and its applications, such as computer vision, data mining, image processing, information retrieval, and pattern recognition. He has several years of industrial and R&D experience in tier-1 organizations and has published sufficient number of research articles in Science Citation Index Journals. He is currently serving as the PI for a research project with Indian Space Research Organization (ISRO).