Research Article

Purnawansyah, Aji Prasetya Wibawa*, Triyanna Widiyaningtyas, Haviluddin, Herdianti Darwis, and Huzain Azis

An in-depth exploration of supervised and semi-supervised learning on face recognition

https://doi.org/10.1515/comp-2025-0029
received March 20, 2024; accepted February 21, 2025

Abstract: This study aims to assess the effectiveness of various algorithms in the realms of supervised and semi-supervised learning applied to three multiclass facial image datasets: JAFFE, Georgia tech, and Yale. The datasets were partitioned into proportions of 80:20, 75:25, and 50:50 for supervised learning, while semi-supervised learning was conducted with labelled and unlabeled data ratios of 20:80, 25:75, and 50:50. The evaluated algorithms include convolutional neural networks (CNNs), decision tree, long short-term memory, K-nearest neighbors (K-NNs), multilayer perceptron, and support vector classifier (SVC), each with varying parameters. Experimental outcomes reveal that the performance of models depends on the dataset partitioning strategies and the type of algorithms used. Specifically, linear and polynomial SVC consistently yield favorable results in supervised learning, particularly demonstrating efficacy on the Georgia tech dataset. Conversely, on the JAFFE and Yale dataset, linear SVC and K-NN emerge as optimal choices. The inclusion of semi-supervised learning enhances insights, particularly evident in the Georgia tech dataset, where the combination of labeled and unlabeled data significantly improves accuracy, especially when leveraging linear SVC algorithm. Although there are some instances of sub-optimal performance in certain algorithms like CNN on specific datasets, this research provides comprehensive insights into the effectiveness of various models in contexts of limited-label learning. The implications of these findings are crucial in advancing the development of adaptive and robust facial recognition systems, especially in navigating datasets characterized by diverse variations and complexities.

Keywords: face expression recognition, face recognition, multiclass, supervised learning, semi-supervised learning

*** Corresponding author: Aji Prasetya Wibawa**, Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Jl. Semarang, Malang, 65145, Indonesia, e-mail: aji.prasetya.ft@um.ac.id
Purnawansyah: Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Jl. Semarang, Malang, 65145, Indonesia; Department of Computer Science, Universitas Muslim Indonesia, Jl. Urip Sumoharjo, Makassar, 90231, Indonesia, e-mail: purnawansyah.2205349@students.um.ac.id, purnawansyah@umi.ac.id
Triyanna Widiyaningtyas: Department of Electrical Engineering and Informatics, Universitas Negeri Malang, Jl. Semarang, Malang, 65145, Indonesia, e-mail: triyannaw.ft@um.ac.id
Haviluddin: Department of Computer Science, Universitas Mulawarman Jl. Kuaro Gunung Kelua, Samarinda, 75119, Indonesia, e-mail: haviluddin@unmul.ac.id
Herdianti Darwis: Department of Computer Science, Universitas Muslim Indonesia, Jl. Urip Sumoharjo, Makassar, 90231, Indonesia, e-mail: herdianti.darwis@umi.ac.id
Huzain Azis: Department of Computer Science, Universitas Muslim Indonesia, Jl. Urip Sumoharjo, Makassar, 90231, Indonesia, e-mail: huzain.azis@umi.ac.id

1 Introduction

In the rapidly evolving landscape of technological advancement, the realms of facial recognition and facial expression recognition (FER) assumed heightened significance within human–computer interaction systems and affective computing domains [1,2]. The imperative for robust classification models, particularly in the multifaceted arena of multiclass scenarios, becomes increasingly pronounced. Multiclass classification [3], pivotal in applications such as facial recognition and emotion analysis, presents intricate challenges necessitating specialized methodologies. Hence, this investigation endeavors to delve deeper into multiclass classification, elucidating the comparative efficacy of supervised and semi-supervised learning paradigms [4]. While extant literature has extensively scrutinized fundamental emotion recognition, we contend that multi-class FER warrants distinct attention. The discernment between nuanced and intricate expressions often poses a formidable hurdle in FER frameworks [5]. Therefore, this inquiry aspires to augment the comprehension of diverse supervised and semi-supervised learning classifications, poised to enhance accuracy and reliability in FER endeavors.

Within the framework of this research endeavor, an array of models, including decision tree [6,7], K-nearest

neighbor (K-NN) [8,9], support vector classifier (SVC) [10], multilayer perceptron (MLP) [11], long short-term memory (LSTM) [12,13], and CNN [12,14,15] are enlisted to meticulously assess and compare the efficacy of diverse supervised and semi-supervised learning paradigms in the domain of multi-class classification [16,17]. Specifically, our scrutiny extends to the performance of these models across varying ratios of labeled to unlabeled data [18], the exploration of the impact stemming from different classifier types and their parameter configurations, and an evaluation of their aptitude for generalization across heterogeneous datasets. This study confines its investigation to three prominent multiclass datasets: Yale, JAFFE, and Georgia tech, thereby providing a focused analysis within a well-defined experimental scope.

The forthcoming outcomes of this scholarly investigation endeavor to provide nuanced insights into the comparative efficacy of supervised and semi-supervised learning models within the intricate domain of multi-class classification [19,20]. By delineating optimal pre-processing configurations and methodologies, the research aspires to contribute to the refinement of classifiers characterized by heightened precision and adaptability, thus pushing the boundaries of machine learning prowess in multiclass classification contexts [21]. Employing a comprehensive suite of training validation using cross-validation (CV) method and several testing evaluative metrics, including accuracy, Kullback–Leibler (KL) divergence, Jaccard index, *F*1-score, cosine similarity, and receiver operating characteristic (ROC)–area under the curve (AUC) [16,18,22], this study encapsulates its rigorous analytical endeavors within the dynamic and evolving landscape of multiclass classification research.

While this study harbors optimistic prospects for unveiling novel insights into multiclass FER [1], it is imperative to acknowledge the potential constraints imposed by the utilization of solely three datasets. It is recognized that extrapolating the findings to analogous datasets necessitates additional inquiry. Therefore, within the confines of these acknowledged limitations, the scholarly contribution of this investigation is envisaged to manifest through the meticulous exposition and juxtaposition of diverse classification methodologies within the milieu of multiclass FER.

In the forthcoming elaboration of this article, a meticulous exposition of the employed experimental methodology, the delineated experimental configuration, the resultant empirical findings, and a nuanced analysis elucidating the efficacy of varied supervised learning classifiers in multiclass FER will be provided [23]. Through this endeavor, it is anticipated that this research endeavor will engender a significant stride toward the development of a more sophisticated and dependable FER system.

2 Method

This research follows a series of steps to evaluate the performance of various learning algorithms on the prepared dataset. The first step involves inputting the image dataset. The dataset then undergoes preprocessing [11], including resizing to equalize the size, normalization to standardize the scale, and splitting the data into two parts: labeled and unlabeled for semi-supervised learning, and training and testing data for supervised learning with composition [24]. The second step includes the selection of learning algorithms, encompassing supervised learning for labeled data and semi-supervised learning that makes use of unlabeled data [23]. In the third step [25], various classification algorithms are applied, such as decision tree, MLP, SVC (linear Kernel [26], polynomial, and radial basis function [RBF]), K-NN [27] (distance: Euclidean, Manhattan, Minkowski, Hamming, Chebyshev, Mahalanobis, Correlation, and cosine similarity [28,29]), LSTM [30], and CNN [13,14]. Before MLP is being applied, preprocessing using StandardScaler and MinMaxScaler is specifically used to normalise and transform data features. Step 4 involves model training and evaluation using metrics such as CV, accuracy, KL, intersection (Jaccard), *F*1-score, cosine, and ROC-AUC [18,31]. Other metrics that can be used are shown in Figure 1.

2.1 Dataset

There are three datasets used: JAFFE, Georgia tech, and Yale sample in Table 1. The JAFFE dataset consists of seven classes representing different emotional expressions on the human face, including happy, sad, angry, fearful, surprised, disgusted, and neutral [32–34]. The Georgia tech Dataset [35], with its 50 classes, covers a wider variety, encompassing variations in facial expressions, poses, and lighting conditions. Meanwhile, the Yale dataset has 15 classes [19,31], and like JAFFE, it focuses on variations in facial expressions and different lighting conditions.

2.2 Supervised learning and semi-supervised learning

Supervised learning [11,36], the primary paradigm in machine learning, focuses on training a model with a pre-labeled dataset, enabling the model to discern patterns and make accurate predictions on new data. In this

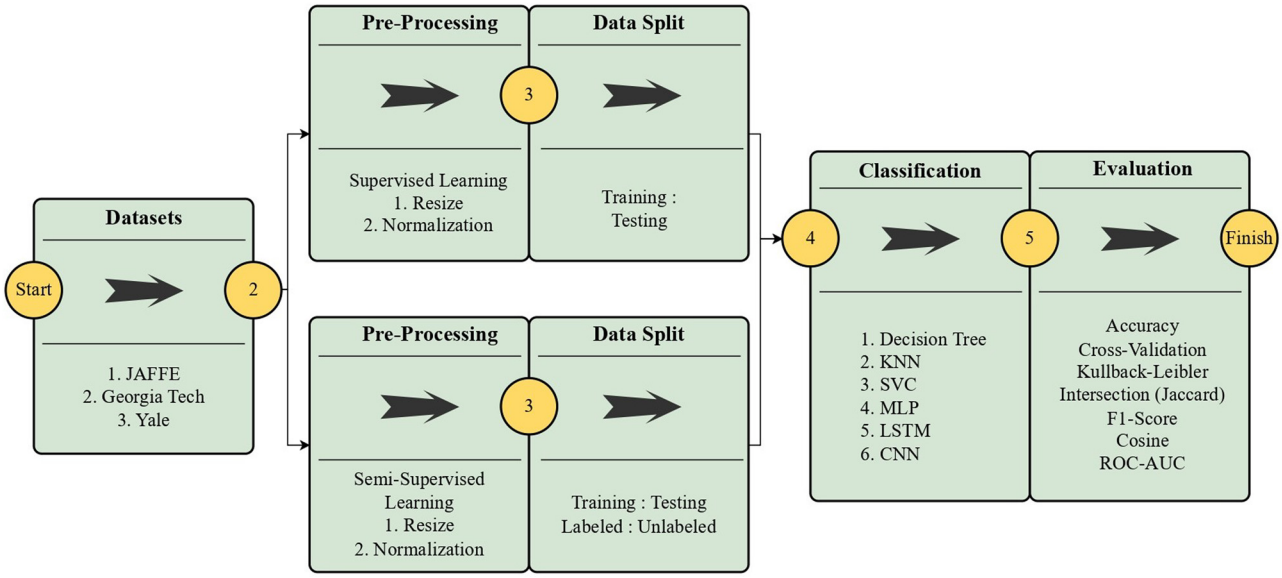
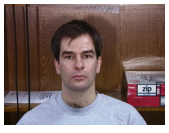


Figure 1: Research design.

research, supervised learning is applied to recognize emotions in human face images, utilizing a dataset labeled based on emotion categories. Various algorithms, including decision tree, K-NN, SVC, MLP, LSTM, and CNN, are tested on this datasets. Dataset partitioning with specific ratios provides a comprehensive understanding of model performance, depending on the allocation of training and test data. Experiments with dataset sharing ratios – such as 80:20, 75:25, and 50:50 – offer insights into the models' flexibility in handling data variations [4,10]. The results demonstrate that the model's performance in classifying emotions in images depends on the algorithm type and dataset sharing ratio. These findings contribute significantly to guiding the development of image-based emotion recognition systems by providing a deeper understanding of the model's adaptability to different datasets.

Table 1: Dataset information

Dataset	Number of classes	Images	Sample data
JAFFE	7	213	
Georgia tech	50	750	
Yale	15	165	

The research also delves into semi-supervised learning methods [31,37,38], incorporating a combination of pre-labeled and unlabeled data [39]. These methods aim to enhance model performance by leveraging additional information from unlabeled data. Initially, the dataset is divided into training and testing sets in an 80:20 ratio. Subsequently, a semi-supervised learning approach is employed, where the 80% training data are structured into various scenarios of labeled and unlabeled data compositions with specific dataset sharing ratios, namely, 20:80, 25:75, and 50:50. The combined data, consisting of labeled and pseudo-labeled data, is then utilized as the new training dataset to classify the testing data reserved for the performance evaluation of the algorithm as elucidated in Figure 2.

Furthermore, data training partitioning includes CV to validate training data in both supervised and semi-supervised learning analyses as depicted in Figure 2. This validation ensures the appropriateness of training data for subsequent testing data classification. Due to the limited data per class in the Yale dataset, CV is restricted to 4 and 5 folds for each dataset (JAFFE, Georgia tech, and Yale), as using more than 5 folds is impractical. Given the scenario of sparse labelled data, semi-supervised learning employs CV with 80% of training data derived from concatenated pre-labeled and pseudo-labeled sets, which is subsequently utilized in the classification process.

2.3 Classification

Six techniques are employed in classifying the FER datasets, specifically decision tree, K-NN with varied k and distance parameters, SVC (linear, polynomial, and RBF kernels), MLP, LSTM, and CNN.

The decision tree serves as a supervised learning model [6], delineating a set of decisions and their consequential outcomes. Through a sequential decision-making process rooted in data features, this model constructs decision rules that are straightforward to interpret, proving highly advantageous for comprehending the factors influencing classification. The construction of the decision tree involves utilizing a subset of data that satisfies the criterion where the value of attribute I in the data is 0. The mathematical representation can be derived by applying the formula (2).

$$D_v = \{(a, b) \in D : x_i = 0\} \tag{1}$$

K-NN represents a non-parametric approach that determines classification outcomes by considering the majority of labels assigned to the K-nearest neighbors within the feature space [40,41]. Numerous distance metrics, including Euclidean, Manhattan, Minkowski with parameters $p = 3$ and $p = 5$, Hamming, Chebyshev, Mahalanobis, correlation [42–44], and cosine similarity [45], are applied to gauge the proximity between data points formulated as provided in Table 2.

SVC is a supervised learning model designed to construct an optimal hyperplane for the purpose of delineating distinct classes within the dataset. Within the framework of SVC, diverse configurations are employed, encompassing linear and polynomial with degrees 2, 3, and 4, as well as RBF with variations in gamma (0.1, 0.01, 0.001, 1, and 10). These variations are applied to facilitate the fitting of the optimal separator, tailored to the characteristics inherent in the provided data

$$f(x) = \sum_{i=1}^{n} a_i y_i K(x_i, x) + b. \tag{2}$$

In a machine learning model, the Lagrange coefficient is denoted as a_i, the class label as y_i, the support vector as x_i, the kernel function as K, and the bias as b. These elements are used in the formulation of the model's decision function.

MLP is a supervised learning model consisting of multiple layers of neurons. By performing iteration-based learning to optimize weights and biases, MLP can handle complex classification tasks.

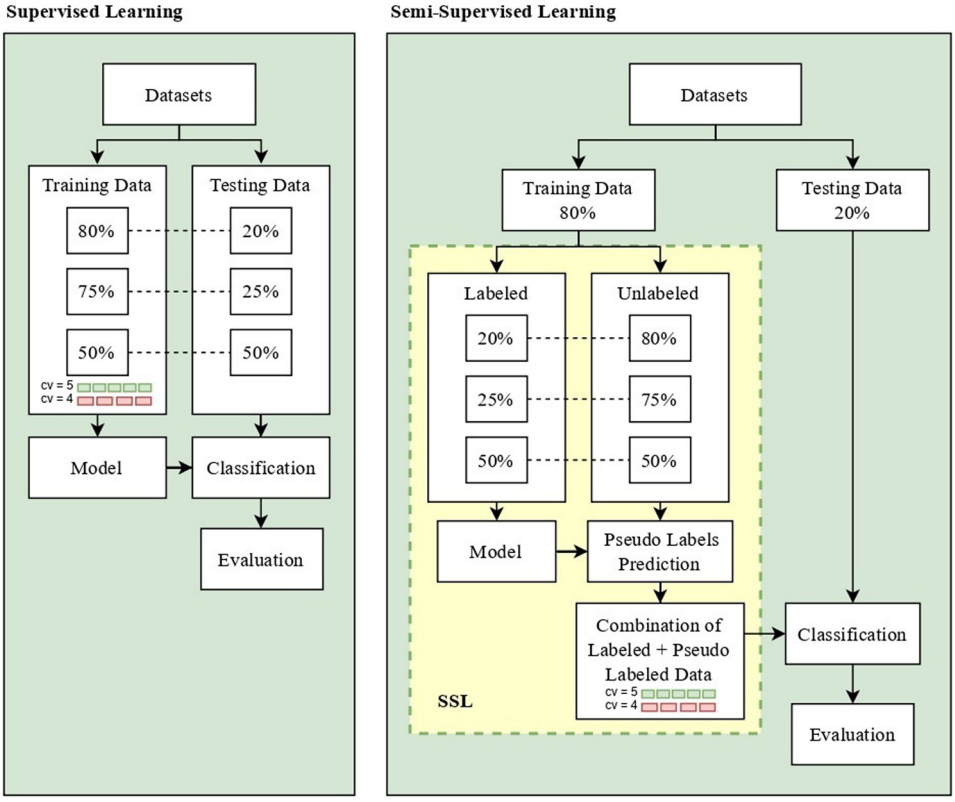


Figure 2: Data splitting scenarios.

Table 2: Formula distance of K-NN

Evaluation distance	Formula $d(x, y)$
Euclidean	$\sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$
Manhattan	$\sum_{i=1}^{n}\lvert x_i - y_i\rvert$
Chebyshev	$\max(\lvert x_i - y_i\rvert)$
Hamming	$\frac{1}{n}\sum_{i=1}^{n}(\delta_{x_i,y_i})$
Mahalanobis	$\sqrt{(x - y)^T S^{-1}(x - y)}$
Minkowski	$\left(\sum_{i=1}^{n}\lvert x_i - y_i\rvert^p\right)^{\frac{1}{p}}$
Correlation	$\frac{1}{2}\left(1 - \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}\right)$
Cosine similarity	$\frac{x \cdot y}{\lVert x\rVert \cdot \lVert y\rVert}$

$$y_{\text{pred}}(x) = \sigma(W^{(2)} \cdot \sigma(W^{(1)} \cdot x + b^{(1)}) + b^{(2)}). \quad (3)$$

Within this symbolic representation, x denotes the input vector, $W^{(1)}$ and $W^{(2)}$ represent the weight matrices, $b^{(1)}$ and $b^{(2)}$ signify the bias vectors, and σ denotes the activation function. This procedural approach enables the MLP to characterize intricate relationships within the data by employing hidden layers with activation functions, ultimately generating predictions denoted as $y_{\text{pred}}(x)$.

LSTM represents a category of recurrent neural networks (RNNs) proficient in addressing issues related to sequencing or temporal dynamics [46]. LSTMs possess the capability to retain and recall information from preceding inputs, rendering them well suited for classification assignments characterized by sequential data, such as textual analysis or temporally oriented data sets.

$$(h_t, c_t) = \text{LSTM}(x_t, h_{t-1}, c_{t-1}). \quad (4)$$

LSTM depicted in formula (3) encapsulates the mechanism of updating the hidden state h_t and cell state c_t at a given time step t. This update integrates the present input $x_{(t)}$, the preceding hidden state $h_{(t-1)}$, and the prior cell state $c_{(t-1)}$. Employing a gating mechanism, LSTM effectively regulates long- and short-term information, mitigating the vanishing gradient challenge inherent in RNNs. This procedural approach is embedded within the LSTM architecture to uphold and structure contextual information from past inputs in the time series [47,48].

CNN, as a deep learning model, proves to be exceptionally efficient in tasks related to image classification [49]. Leveraging convolution, pooling, and connected layers to their full potential, CNNs can autonomously derive hierarchical features from image data, rendering them apt for multiclass classification scenarios with a visual basis. The mathematical representation can be computed utilizing formula (1).

$$Y = \sigma(X \cdot F + b). \quad (5)$$

CNN integrates convolution, activation, and pooling procedures [11,15]. Let X denote the input, F the filter, b the bias, σ the activation function, and P the pooling operation [50].

Among the algorithms discussed, K-NN stands out for its simplicity and non-parametric nature, making it suitable for multi-class classification tasks but susceptible to computational inefficiency with large datasets. SVC exhibits versatility in high-dimensional spaces, albeit demanding careful kernel selection and computational resources. LSTM networks excel in capturing temporal dependencies, although they are complex and require substantial data for training. CNNs excel at image recognition by extracting hierarchical features but require large datasets and are sensitive to input variations. MLP offers flexibility across tasks but are prone to overfitting and require meticulous hyperparameter tuning. Decision trees provide interpretability but are susceptible to overfitting and instability with complex structures, necessitating careful pruning and parameter adjustment. Overall, understanding the unique characteristics and trade-offs of each algorithm is crucial for selecting the most appropriate model for a given problem domain. The parameter settings used for each classification method are detailed in Table 3.

2.4 Evaluation

The various evaluation metrics in this study include a number of criteria to measure model performance. CV helps measure the extent to which the model can be generalized to new data. In this research, CV is implemented to validate the training data before being applied to the classification process. It is conducted by splitting the training data into CV of 4 and 5, fitting a model, computing the accuracy for each iteration, and performing the average accuracy for each CV.

On the other hand, testing data evaluation is performed by calculating accuracy, KL, Intersection, *F*1-score, cosine similarity, and ROC-AUC. Accuracy gives an idea of how accurate the model is in classifying the data. KL evaluates the extent to which the model's predicted probability distribution is close to the true distribution [51]. Intersection (Jaccard) measures the similarity between the actual data set and the predicted results. *F*1-score provides a balance between precision and recall, very important in cases of class imbalance. Cosine similarity measures the extent to which

Table 3: Parameter settings

Algorithm	Distance/layer/function	Parameter	Value
Decision tree	—	Random_state	42
K-NN	Euclidean	k	1, 2, 3
	Manhattan		
	Hamming		
	Chebyshev		
	Mahalanobis		
	Correlation		
	Cosine Similarity		
	Minkowski	k	1, 2, 3
		Power	3,5
SVC	Linear	Probability	True
		C	1
	RBF	Probability	True
		Gamma	0.001; 0.01; 0.1; 1; 10
	Poly	Probability	True
		C	1
		Degree	2, 3, 4
MLP	—	Hidden_layer_sizes	(100, 50)
		Max_iter	100
		Random_state	42
LSTM	1st Layer	Units	128
		Input_shape	(150, 150*3)
		Return_sequences	True
	2nd Layer	Units	64
		Return_sequences	True
	3rd Layer	Units	32
	Dense (Hidden)	Units	64
		Activation	Relu
	Dense (Output)	Units	Number_of_classes
		Activation	Softmax
	Compile	Optimizer	Adam
		Loss	Sparse_categorical_crossentropy
		Metrics	Accuracy
	KerasClassifier	Build	Build_lstm_model
		Epochs	10
		Batch_size	32
		Verbose	0
		Random_state	42
CNN	Conv2D	Filters	32
		Kernel_size	(3,3)
		Activation	Relu
		Input_shape	(150, 150, 3)
	MaxPooling2D	Pool_size	(2,2)
	Flatten	—	—
	Dense (hidden)	Units	64
		Activation	Relu
	Dense (Output)	Units	number_of_classes
		activation	Softmax
	Compile	Optimizer	Adam
		loss	sparse_categorical_crossentropy
		metrics	Accuracy
	KerasClassifier	Build	Build_cnn_model
		Epochs	10
		Batch_size	32
		Verbose	0
		Random_state	42

the modelled data representation vector is similar to the actual data vector. Finally, ROC-AUC assesses the model's ability to distinguish between positive and negative classes based on the ROC curve and area under the curve (AUC) [52]. These metrics provide a holistic picture of the model's effectiveness in the context of various aspects of evaluation.

In Table 4, we provide the mathematical representation for each assessment, accompanied by its elucidation. The accuracy metric gauges the proportion of accurate predictions relative to the overall evaluated data, where true positive (TP), true negative (TN), false positive (FP), and false negative (FN) are defined [8,14]. Here, k represents the number of folds, Model_i signifies the model trained on the ith fold, and $1/k$ denotes the average of these models. In addition, P and Q denote two probability distributions, and i is the index of an element within the distribution. A and B represent two sets, $J \cdot K$ corresponds to the dot product between vectors J and K, and $\|J\| \cdot \|K\|$ denotes the Euclidean norm of vector A. The variable d typically signifies the decision threshold.

Various evaluation metrics are employed in this research to assess the performance of models. Accuracy, a fundamental metric, provides a simple measure of overall correctness but may overlook class imbalances and misclassification costs. CV offers robust performance estimates by averaging over multiple folds, yet its computational demands can be prohibitive. K-L Divergence measures dissimilarity between probability distributions, but it necessitates knowledge of the true underlying distribution. The Jaccard index is beneficial for imbalanced datasets but is sensitive to class imbalance and misclassification costs. $F1$-score balances precision and recall, yet its efficacy depends on the chosen threshold [53]. Cosine similarity captures direction similarity but disregards magnitude differences, while ROC-AUC provides a comprehensive measure of classifier performance but may be unsuitable for imbalanced datasets and fails to encapsulate overall model performance [54,55].

Table 4: Formula of each evaluation metric

Evaluation metric	Formula
Accuracy	$\frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$
Cross validation	$\frac{1}{k}\sum_{i=1}^{k}\text{Model}_i$
KL divergence	$D_{\text{KL}}(P\|\|Q) = \sum_i P(i)\log\left(\frac{P(i)}{Q(i)}\right)$
Intersection (Jaccard) similarity	$J(A,B) = \frac{\|A \cap B\|}{\|A \cup B\|}$
$F1$-score	$\frac{2 \cdot \frac{\text{TP}}{\text{TP}+\text{FP}} \cdot \frac{\text{TP}}{\text{TP}+\text{FN}}}{\left(\frac{\text{TP}}{\text{TP}+\text{FP}} + \frac{\text{TP}}{\text{TP}+\text{FN}}\right)}$
Cosine similarity	$\frac{\mathbf{J} \cdot \mathbf{K}}{\|\mathbf{J}\| \times \|\mathbf{K}\|}$
ROC-AUC	$\int_0^1 \left(\frac{\text{TP}}{\text{TP}+\text{FN}}\right) d\left(\frac{\text{FP}}{\text{FP}+\text{FN}}\right)$

3 Result and discussion

This study evaluates the performance of 6 classification algorithms across 18 different scenarios, encompassing types of learning (supervised and semi-supervised learning); datasets (JAFFE, Georgia tech, and Yale); and data splitting for each type of learning as detailed in Tables 5–22. Overall, the findings indicate that the training data across all datasets were effectively validated through the CV results across all 18 scenarios. Furthermore, a comparison between the accuracy results of training validation and testing evaluation in all scenarios revealed that the differences were not significantly large. Consequently, there is no indication of potential overfitting.

The performance of several supervised learning models employing an 80:20 split for training and testing are shown in Tables 5–7. The experimental results on the JAFFE dataset reveal that the linear SVC model achieves the highest accuracy in Table 5 at 0.84, followed by polynomial SVC models with degrees of 3 and 4. However, upon evaluation using the KL divergence metric, LSTM model exhibits a negative score (−0.25), indicating improved performance with decreasing scores. This suggests that a lower KL divergence value corresponds to better model performance [51]. On the basis of research [52], some argue that theoretically and empirically, ROC-AUC is superior to accuracy metrics for evaluating performance. These findings offer crucial insights into model selection for emotion recognition tasks on the JAFFE dataset and highlight the potential of LSTM models in addressing such issues.

The Georgia tech dataset revealed that SVC linear in Table 6, polynomial degrees of 3 and 4 have the highest accuracy of 1.00 (100%). However, evaluation using Kullback indicates that the model with RBF kernel and gamma 0.1 has a lower value, specifically 0.0001, showing a significant improvement in model performance. This suggests that the decrease in the KL value signifies an enhancement in the model's alignment with empirical data. In Table 7, the Yale dataset (80:20), both linear and polynomial SVC (degrees of 2, 3, 4), achieved an accuracy of 0.91. However, assessment using KL divergence revealed that RBF with gamma 10 had a value of −0.29, indicating better performance with a lower value. Consequently, despite the high accuracy of SVC, RBF is preferred.

Table 5: Training validation and testing results of supervised JAFFE (80:20)

					Training validation		Testing evaluation					
					CV		Accuracy	K–L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
JAFFE	Decision tree				0.48	0.50	0.42	2.64	0.30	0.43	0.76	0.66
	K-NN	k	1	Euclidean	0.71	0.72	0.67	2.64	0.52	0.67	0.96	0.81
				Manhattan	0.73	0.74	0.67	2.64	0.53	0.67	0.94	0.81
				Minkowski p = 3	0.69	0.70	0.72	2.65	0.58	0.72	0.97	0.84
				Minkowski p = 5	0.66	0.69	0.70	2.65	0.54	0.69	0.94	0.82
				Hamming	0.67	0.67	0.63	2.64	0.46	0.61	0.94	0.78
				Chebyshev	0.51	0.50	0.58	2.65	0.46	0.61	0.94	0.76
				Mahalanobis	0.75	0.75	0.72	2.65	0.57	0.71	0.96	0.84
				Correlation	0.72	0.72	0.70	2.65	0.56	0.69	0.97	0.82
				Cosine similarity	0.73	0.72	0.72	2.65	0.58	0.71	0.97	0.84
			3	Euclidean	0.44	0.51	0.49	2.07	0.32	0.47	0.87	0.81
				Manhattan	0.41	0.50	0.51	2.04	0.34	0.49	0.90	0.82
				Minkowski p = 3	0.47	0.52	0.42	2.04	0.28	0.41	0.88	0.78
				Minkowski p = 5	0.44	0.47	0.53	2.02	0.37	0.51	0.90	0.82
				Hamming	0.35	0.44	0.44	1.96	0.28	0.41	0.83	0.79
				Chebyshev	0.40	0.44	0.47	1.92	0.29	0.45	0.81	0.84
				Mahalanobis	0.49	0.50	0.47	2.07	0.30	0.45	0.86	0.83
				Correlation	0.41	0.51	0.47	2.12	0.31	0.45	0.81	0.82
				Cosine similarity	0.41	0.50	0.51	2.15	0.35	0.49	0.89	0.81
			5	Euclidean	0.28	0.34	0.21	1.64	0.12	0.20	0.84	0.21
				Manhattan	0.26	0.31	0.23	1.63	0.14	0.23	0.82	0.72
				Minkowski p = 3	0.29	0.40	0.28	1.57	0.17	0.28	0.86	0.76
				Minkowski p = 5	0.30	0.35	0.35	1.54	0.21	0.33	0.85	0.79
				Hamming	0.23	0.28	0.23	1.48	0.12	0.20	0.84	0.70
				Chebyshev	0.32	0.34	0.42	1.49	0.26	0.40	0.77	0.81
				Mahalanobis	0.41	0.44	0.40	1.61	0.25	0.37	0.76	0.83
				Correlation	0.24	0.33	0.30	1.62	0.19	0.31	0.82	0.76
				Cosine similarity	0.26	0.35	0.26	1.64	0.16	0.26	0.84	0.76
	SVC	Linear			**0.80**	**0.79**	**0.84**	**0.10**	**0.74**	**0.83**	**0.99**	**0.98**
		Poly	Degree	2	0.68	0.62	0.74	−0.15	0.62	0.74	0.97	0.94
				3	**0.79**	**0.77**	**0.81**	**−0.12**	**0.70**	**0.81**	**0.99**	**0.97**
				4	**0.76**	**0.77**	**0.74**	**−0.14**	**0.60**	**0.74**	**0.95**	**0.97**
		RBF	Gamma	0.1	0.16	0.15	0.14	**−0.24**	0.02	0.03	0.83	0.23
				0.01	0.51	0.47	0.44	**−0.24**	0.30	0.44	0.89	0.86
				0.001	0.55	0.54	0.47	−0.20	0.32	0.45	0.94	0.87
				1	0.20	0.32	0.21	**−0.24**	0.06	0.10	0.81	0.50
				10	0.14	0.14	0.14	**−0.24**	0.01	0.03	0.83	0.50
	MLP	MLP			0.19	0.19	0.14	0.42	0.01	0.03	0.83	0.53
		StandardScaler			0.65	0.70	0.65	2.37	0.53	0.66	0.93	0.89
		MinMaxScaler			0.18	0.19	0.16	1.00	0.04	0.06	0.85	0.51
	LSTM				0.14	0.16	0.14	**−0.25**	0.04	0.08	0.77	0.52
	CNN				0.30	0.40	0.63	−0.07	0.46	0.60	0.94	0.89

CV: cross validation, K–L: Kullback–Leibler, I–J: intersection (Jaccard).

The results from supervised learning experiments employing a 75:25 dataset split, as depicted in Tables 8–10, exhibit varied performance. The JAFFE dataset in Table 8 indicates that linear SVC achieves the highest performance with an accuracy of 0.85, followed by polynomial SVC with degrees 3 and 4 at 0.78 and 0.76, respectively. However, when evaluated with the KL metric, RBF SVC with various gamma values (0.1, 0.01, 1, and 10) demonstrates the best performance. On the Georgia tech dataset in Table 9, the linear SVC method, polynomial SVC with degrees 2, 3, 4, and K-NN

Table 6: Training validation and testing results of supervised Georgia tech (80:20)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.81	0.81	0.83	0.41	0.73	0.81	0.94	0.91
	K-NN	*k*	1	Euclidean	0.99	0.99	0.99	0.002	0.99	0.99	0.99	**1.00**
				Manhattan	**0.99**	**0.99**	**1.00**	**3.82**	**1.00**	**1.00**	**1.00**	**1.00**
				Minkowski p = 3	0.99	0.99	0.99	0.002	0.99	0.99	0.99	**1.00**
				Minkowski p = 5	0.97	0.97	0.95	0.01	0.92	0.95	0.99	0.98
				Hamming	0.99	0.99	0.99	0.002	0.99	0.99	**1.00**	**1.00**
				Chebyshev	0.90	0.91	0.91	0.40	0.85	0.90	0.98	0.95
				Mahalanobis	0.97	0.97	0.97	0.01	0.94	0.97	0.99	0.98
				Correlation	0.99	0.99	0.99	0.002	0.99	**1.00**	0.99	**1.00**
				Cosine similarity	0.99	0.99	0.99	0.002	0.99	0.99	0.99	**1.00**
			3	Euclidean	0.96	0.96	0.97	0.009	0.95	0.97	0.99	**1.00**
				Manhattan	0.97	0.97	0.98	0.006	0.97	0.98	**1.00**	**1.00**
				Minkowski p = 3	0.96	0.96	0.97	0.02	0.91	0.94	0.99	0.99
				Minkowski p = 5	0.96	0.96	0.95	0.02	0.91	0.94	0.99	0.99
				Hamming	0.96	0.96	0.97	0.007	0.96	0.97	**1.00**	**1.00**
				Chebyshev	0.82	0.82	0.85	0.07	0.76	0.84	0.97	0.97
				Mahalanobis	0.95	0.95	0.95	0.02	0.91	0.94	**1.00**	0.99
				Correlation	0.96	0.96	0.98	0.01	0.97	0.98	**1.00**	**1.00**
				Cosine similarity	0.97	0.97	0.97	0.009	0.95	0.97	**1.00**	**1.00**
			5	Euclidean	0.96	0.96	0.97	0.009	0.95	0.97	**1.00**	**1.00**
				Manhattan	0.97	0.77	0.98	0.006	0.97	0.98	**1.00**	**1.00**
				Minkowski	0.96	0.96	0.97	0.009	0.95	0.97	0.99	**1.00**
				Minkowski p = 3	0.91	0.92	0.94	0.02	0.90	0.93	0.98	**1.00**
				Minkowski p = 5	0.90	0.91	0.93	0.03	0.88	0.98	0.99	0.99
				Chebyshev	0.82	0.82	0.85	0.07	0.76	0.84	**1.00**	**1.00**
				Mahalanobis	0.95	0.95	0.95	0.02	0.91	0.94	**1.00**	0.99
				Correlation	0.96	0.96	0.98	0.01	0.97	0.98	0.99	**1.00**
				Cosine similarity	0.97	0.97	0.97	0.009	0.95	0.97	**1.00**	**1.00**
	SVC	Linear			**0.99**	**0.99**	**1.00**	**0.03**	**1.00**	**1.00**	**1.00**	**0.99**
		Poly	Degree	2	0.99	0.99	0.99	0.03	0.98	0.99	0.99	0.99
				3	**0.99**	**0.99**	**1.00**	**0.04**	**1.00**	**1.00**	**1.00**	**0.99**
				4	**0.99**	**0.99**	**1.00**	**0.05**	**1.00**	**1.00**	**1.00**	**0.99**
		RBF	Gamma	0.1	0.27	0.15	0.26	**0.0001**	0.24	0.32	0.89	0.10
				0.01	0.68	0.61	0.63	0.08	0.61	0.69	0.93	0.93
				0.001	0.97	0.98	0.98	0.07	0.96	0.98	0.99	0.99
				1	0.72	0.36	0.67	0.001	0.64	0.68	0.91	0.48
				10	0.07	0.05	0.07	0.0003	0.05	0.07	0.86	0.50
	MLP	MLP			0.96	0.66	0.89	3.01	0.82	0.88	0.97	0.99
		StandardScaler			0.96	0.96	0.95	3.01	0.82	0.88	0.97	**1.00**
		MinMaxScaler			0.98	0.97	0.97	3.47	0.95	0.97	0.99	**1.00**
	LSTM				0.41	0.54	0.59	0.02	0.39	0.46	0.85	0.97
	CNN				0.97	0.97	0.99	0.001	0.98	0.99	**1.00**	**1.00**

with k = 1 using Manhattan and Hamming distances achieve high accuracies of 0.98. Meanwhile, on the Yale dataset in Table 10, linear SVC dominates with an accuracy of 0.81, followed by polynomial SVC with degree 2, and K-NN with k = 3 and 5 using Manhattan distance, each with an accuracy of 0.76.

In supervised learning experiments with a 50:50 dataset split as shown in Tables 11–13, the results indicate significant performance variations among algorithms and methods on each dataset. The JAFFE dataset, as shown in Table 11, the leading performance is held by Linear SVC, but in terms of the Kullback metric, the best performance is with RBF SVC

Table 7: Training validation and testing results of supervised Yale (80:20)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.56	0.54	0.73	5.97	0.63	0.72	0.89	0.85
	K-NN	*k*	1	Euclidean	0.70	0.71	0.78	5.97	0.64	0.77	0.92	0.88
				Manhattan	0.71	0.70	0.78	5.97	0.64	0.77	0.93	0.88
				Minkowski $p = 3$	0.69	0.69	0.78	5.97	0.64	0.76	0.92	0.88
				Minkowski $p = 5$	0.66	0.65	0.71	5.97	0.57	0.70	0.87	0.84
				Hamming	0.69	0.69	0.80	5.97	0.67	0.80	0.94	0.89
				Chebyshev	0.46	0.43	0.53	5.97	0.68	0.80	0.94	0.89
				Mahalanobis	0.69	0.69	0.80	5.97	0.67	0.80	0.94	0.89
				Correlation	0.74	0.72	0.87	5.97	0.77	0.86	0.95	0.92
				Cosine similarity	0.71	0.70	0.80	5.97	0.66	0.79	0.92	0.89
			3	Euclidean	0.68	0.68	0.76	5.72	0.63	0.75	0.92	0.93
				Manhattan	0.72	0.72	0.78	5.74	0.66	0.78	0.93	0.93
				Minkowski $p = 3$	0.63	0.61	0.78	5.68	0.64	0.77	0.89	0.94
				Minkowski $p = 5$	0.68	0.68	0.79	5.72	0.63	0.75	0.84	0.92
				Hamming	0.73	0.71	0.82	5.74	0.71	0.81	0.96	0.93
				Chebyshev	0.42	0.39	0.40	5.39	0.25	0.34	0.78	0.81
				Mahalanobis	0.73	0.71	0.82	5.74	0.71	0.81	0.96	0.89
				Corelation	0.67	0.68	0.77	5.71	0.65	0.34	0.78	0.94
				Cosine similarity	0.69	0.68	0.80	5.68	0.68	0.79	0.93	0.93
			5	Euclidean	0.66	0.66	0.76	5.45	0.61	0.74	0.91	0.95
				Manhattan	0.68	0.67	0.80	5.56	0.68	0.80	0.94	0.95
				Minkowski $p = 3$	0.61	0.60	0.73	5.43	0.58	0.77	0.89	0.94
				Minkowski $p = 5$	0.57	0.55	0.71	5.26	0.55	0.68	0.84	0.95
				Hamming	0.67	0.67	0.78	0.53	0.65	0.77	0.96	0.95
				Chebyshev	0.37	0.37	0.47	4.93	0.33	0.41	0.75	0.84
				Mahalanobis	0.67	0.67	0.78	5.52	0.65	0.77	0.96	0.95
				Correlation	0.65	0.67	0.80	5.38	0.68	0.41	0.74	0.94
				Cosine similarity	0.65	0.67	0.77	5.36	0.63	0.73	0.92	0.97
	SVC	Linear			**0.81**	**0.81**	**0.91**	**−0.10**	**0.84**	**0.91**	**0.98**	**0.99**
		Poly	Degree	2	**0.77**	**0.77**	**0.91**	**−0.15**	**0.85**	**0.91**	**0.97**	**0.99**
				3	**0.72**	**0.73**	**0.91**	**−0.14**	**0.84**	**0.91**	**0.78**	**0.98**
				4	**0.68**	**0.66**	**0.91**	**−0.14**	**0.85**	**0.91**	**0.99**	**0.97**
		RBF	Gamma	0.1	0.37	0.16	0.29	−0.28	0.23	0.27	0.52	0.46
				0.01	0.40	0.16	0.33	−0.28	0.27	0.32	0.52	0.22
				0.001	0.71	0.70	0.80	−0.23	0.68	0.80	0.93	0.97
				1	0.28	0.15	0.24	−0.28	0.18	0.22	0.84	0.45
				10	0.14	0.12	0.11	**−0.29**	0.05	0.07	0.85	0.43
	MLP	MLP			0.13	0.12	0.77	5.54	0.67	0.77	0.95	0.82
		StandardScaler			0.64	0.65	0.77	5.78	0.67	0.76	0.95	0.78
		MinMaxScaler			0.20	0.13	0.22	1.53	0.09	0.13	0.87	0.78
	LSTM				0.13	0.12	0.56	5.53	0.67	0.76	0.95	0.82
	CNN				0.64	0.65	0.55	5.78	0.67	0.76	0.95	0.78

with gamma and LSTM, which is −0.16. In the Georgia tech dataset in Table 12, overall good performance and good results in terms of the Kullback metric are obtained from SVC RBF with gamma 1, which is 0.004. Meanwhile, in the Yale dataset in Table 13, good performance is generally observed in linear SVC, with the use of ROC-AUC in SVC poly degree 4, and LSTM for the Kullback metric, where all uses of SVC RBF with gamma yield −0.25.

The outcomes of the semi-supervised learning experiments with a 20:80 ratio in Tables 14–16 reveal noteworthy variations in the performance of the models. In the JAFFE dataset, as shown in Table 14, the leading performance is

Table 8: Training validation and testing results of supervised JAFFE (75:25)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
JAFFE	Decision tree				0.48	0.47	0.48	2.07	0.32	0.48	0.85	0.70
	K-NN	*k*	1	Euclidean	0.74	0.72	0.67	2.07	0.51	0.67	0.95	0.81
				Manhattan	0.73	0.71	0.65	2.07	0.50	0.65	0.94	0.79
				Minkowski $p = 3$	0.70	0.68	0.70	2.07	0.56	0.70	0.96	0.83
				Minkowski $p = 5$	0.72	0.70	0.69	2.07	0.54	0.69	0.95	0.82
				Hamming	0.67	0.64	0.61	2.07	0.45	0.60	0.94	0.78
				Chebyshev	0.51	0.49	0.52	2.07	0.48	0.60	0.94	0.72
				Mahalanobis	0.75	0.72	0.72	2.07	0.58	0.72	0.94	0.84
				Correlation	0.73	0.71	0.67	2.07	0.52	0.67	0.95	0.81
				Cosine similarity	0.74	0.72	0.68	2.07	0.54	0.69	0.95	0.82
			3	Euclidean	0.46	0.43	0.41	1.61	0.25	0.39	0.83	0.79
				Manhattan	0.43	0.41	0.44	1.59	0.28	0.43	0.86	0.79
				Minkowski $p = 3$	0.45	0.43	0.37	1.60	0.23	0.36	0.84	0.77
				Minkowski $p = 5$	0.40	0.37	0.43	1.54	0.27	0.40	0.86	0.80
				Hamming	0.44	0.42	0.43	1.49	0.26	0.38	0.82	0.78
				Chebyshev	0.35	0.37	0.46	1.50	0.29	0.45	0.80	0.80
				Mahalanobis	0.53	0.52	0.43	1.60	0.26	0.40	0.79	0.82
				Correlation	0.47	0.46	0.41	1.63	0.26	0.45	0.80	0.80
				Cosine similarity	0.47	0.43	0.44	1.65	0.30	0.44	0.85	0.79
			5	Euclidean	0.29	0.29	0.24	1.29	0.14	0.23	0.83	0.73
				Manhattan	0.29	0.28	0.24	1.27	0.14	0.25	0.80	0.70
				Minkowski $p = 3$	0.32	0.34	0.28	1.20	0.16	0.27	0.84	0.74
				Minkowski $p = 5$	0.29	0.30	0.35	1.14	0.21	0.35	0.82	0.76
				Hamming	0.24	0.20	0.24	1.13	0.14	0.23	0.84	0.69
				Chebyshev	0.36	0.34	0.43	1.21	0.26	0.41	0.79	0.81
				Mahalanobis	0.37	0.40	0.43	1.21	0.28	0.43	0.79	0.83
				Correlation	0.29	0.30	0.30	1.25	0.18	0.30	0.81	0.74
				Cosine similarity	0.30	0.30	0.24	1.25	0.14	0.24	0.81	0.73
	SVC	Linear			**0.79**	**0.75**	**0.85**	**−0.13**	**0.75**	**0.85**	**0.97**	**0.98**
		Poly	Degree	2	0.66	0.65	0.63	−0.17	0.47	0.63	0.90	0.94
				3	**0.78**	**0.75**	**0.78**	**0.15**	**0.77**	**0.77**	**0.95**	**0.95**
				4	**0.78**	**0.73**	**0.76**	**−0.15**	**0.60**	**0.75**	**0.93**	**0.95**
		RBF	Gamma	0.1	0.15	0.14	0.15	**−0.23**	0.02	0.04	0.84	0.47
				0.01	0.50	0.48	0.71	**−0.23**	0.36	0.48	0.49	0.71
				0.001	0.55	0.52	0.43	−0.19	0.28	0.43	0.89	0.85
				1	0.18	0.17	0.15	**−0.23**	0.02	0.04	0.84	0.50
				10	0.14	0.14	0.15	**−0.23**	0.02	0.04	0.84	0.50
	MLP	MLP			0.19	0.16	0.15	0.24	0.02	0.04	0.84	0.56
		StandardScaler			0.66	0.65	0.57	1.71	0.44	0.57	0.91	0.85
		MinMaxScaler			0.18	0.18	0.17	0.44	0.04	0.07	0.84	0.53
	LSTM				0.14	0.14	0.20	−0.23	0.08	0.14	0.70	0.54
	CNN				0.41	0.35	0.74	−0.04	0.60	0.74	0.95	0.93

held by linear SVC, polynomial degree 3, and K-NN with $k = 1$ using Minkowski distance with $p = 5$, which is 0.38. However, when viewed in terms of the Kullback metric, the best performance is seen in RBF SVC with gamma 0.1, which is −0.16. In the Georgia tech dataset, as shown in Table 15, overall good performance and favorable results are obtained from both datasets used, namely, SVC polynomial degree and Manhattan. Meanwhile, in the Yale dataset, as presented in Table 16, good performance is generally observed in K-NN with $k = 1$ and 3 using Minkowski distance with $p = 5$ and correlation. However, in the usage of CV and cosine leading polynomial degree 2 and 3,

Table 9: Training validation and testing results Georgia tech (75:25)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.80	0.79	0.78	0.45	0.65	0.75	0.96	0.88
	K-NN	k	1	Euclidean	0.98	0.98	0.97	0.01	0.95	0.97	0.99	0.99
				Manhattan	**0.99**	**0.99**	**0.98**	**0.01**	**0.96**	**0.97**	**0.99**	**0.99**
				Minkowski p = 3	0.98	0.98	0.97	0.02	0.94	0.96	0.99	0.98
				Minkowski p = 5	0.97	0.97	0.94	0.03	0.89	0.92	0.99	0.97
				Hamming	**0.98**	**0.98**	**0.98**	**0.01**	**0.96**	**0.97**	**0.99**	**0.99**
				Chebyshev	0.89	0.89	0.89	0.41	0.81	0.87	0.98	0.94
				Mahalanobis	0.96	0.97	0.94	0.03	0.90	0.93	0.99	0.97
				Correlation	0.98	0.98	0.97	0.01	0.95	0.97	0.99	0.99
				Cosine similarity	0.99	0.99	0.97	0.01	0.95	0.97	0.99	0.99
			3	Euclidean	0.96	0.97	0.96	0.02	0.93	0.96	0.99	0.99
				Manhattan	0.96	0.97	0.97	0.02	0.95	0.97	0.99	0.99
				Minkowski p = 3	0.95	0.96	0.94	0.04	0.89	0.93	0.98	0.98
				Minkowski p = 5	0.95	0.96	0.97	0.02	0.95	0.97	0.99	0.99
				Hamming	0.95	0.96	0.97	0.02	0.95	0.97	0.99	0.99
				Chebyshev	0.84	0.83	0.82	0.09	0.72	0.80	0.96	0.96
				Mahalanobis	0.94	0.96	0.93	0.04	0.88	0.92	0.98	0.97
				Correlation	0.96	0.96	0.97	0.03	0.95	0.97	0.97	0.99
				Cosine similarity	0.96	0.97	0.96	0.02	0.96	0.96	0.99	0.99
			5	Euclidean	0.90	0.91	0.93	0.04	0.88	0.92	0.98	0.99
				Manhattan	0.90	0.92	0.94	0.04	0.90	0.93	0.98	0.99
				Minkowski p = 3	0.90	0.91	0.93	0.05	0.88	0.92	0.97	0.99
				Minkowski p = 5	0.89	0.90	0.90	0.06	0.83	0.89	0.96	0.98
				Hamming	0.77	0.78	0.78	0.10	0.67	0.77	0.95	0.96
				Chebyshev	0.91	0.92	0.93	0.03	0.89	0.92	0.98	0.99
				Mahalanobis	0.91	0.90	0.90	0.06	0.83	0.89	0.97	0.97
				Correlation	0.90	0.91	0.92	0.04	0.87	0.91	0.98	0.99
				Cosine similarity	0.90	0.92	0.95	0.04	0.92	0.94	0.98	0.99
	SVC	Linear			**0.98**	**0.99**	**0.98**	**0.05**	**0.96**	**0.98**	**1.00**	**0.99**
		Poly	Degree	2	**0.98**	**0.99**	**0.98**	**0.04**	**0.96**	**0.98**	**1.00**	**0.99**
				3	**0.98**	**0.99**	**0.98**	**0.05**	**0.97**	**0.98**	**1.00**	**0.99**
				4	**0.98**	**0.99**	**0.98**	**0.05**	**0.96**	**0.98**	**1.00**	**0.99**
		RBF	Gamma	0.1	0.06	0.21	0.32	8.04	0.29	0.28	0.89	0.07
				0.01	0.52	0.64	0.62	0.06	0.60	0.68	0.93	0.55
				0.001	0.97	0.97	0.95	0.07	0.93	0.96	0.99	**1.00**
				1	0.12	0.53	0.68	**0.001**	0.66	0.69	0.91	0.47
				10	0.02	0.06	0.07	9.87	0.05	0.07	0.86	0.49
	MLP	MLP			0.92	0.96	0.94	2.03	0.89	0.92	0.99	0.99
		StandardScaler			0.94	0.95	0.95	4.26	0.91	0.94	0.98	0.99
		MinMaxScaler			0.95	0.96	0.94	2.07	0.91	0.94	0.99	0.99
	LSTM				0.51	0.54	0.45	0.01	0.32	0.37	0.85	0.96
	CNN				0.96	0.97	0.93	0.04	0.87	0.92	0.97	0.99

Kullback from gamma 0.1, and in ROC-AUC, linear SVC with 0.93 stands out.

The experimental results are presented in Tables 17–19 semi-supervised 25:75. In the JAFFE dataset, as shown in Table 17, overall good performance is observed with linear SVC and K-NN k = 1 using Chebyshev distance and cosine similarity. In the Georgia tech dataset, as shown in Table 18, overall good performance is observed with linear SVC and polynomial degree, as well as K-NN k = 1 using Manhattan distance, particularly in the utilization of ROC and AUC metrics where the CNN algorithm excels. Meanwhile, in the Yale dataset as depicted in Table 19, good

Table 10: Training validation and testing results of supervised Yale (75:25)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.60	0.64	0.61	6.40	0.47	0.61	0.89	0.82
	K-NN	*k*	1	Euclidean	0.77	0.78	0.69	0.64	0.56	0.69	0.93	0.85
				Manhattan	0.77	0.78	0.71	6.40	0.58	0.71	0.93	0.85
				Minkowski $p = 3$	0.76	0.77	0.74	6.40	0.62	0.73	0.95	0.87
				Minkowski $p = 5$	0.74	0.74	0.71	6.40	0.59	0.69	0.92	0.86
				Hamming	0.73	0.74	0.69	6.40	0.57	0.70	0.93	0.84
				Chebyshev	0.40	0.41	0.55	6.40	0.57	0.70	0.93	0.84
				Mahalanobis	0.73	0.74	0.69	6.40	0.57	0.70	0.93	0.84
				Correlation	0.80	0.78	0.71	6.40	0.60	0.71	0.95	0.86
				Cosine similarity	0.77	0.78	0.69	6.40	0.56	0.69	0.91	0.85
			3	Euclidean	0.74	0.75	0.64	6.01	0.52	0.65	0.88	0.89
				Manhattan	**0.77**	**0.78**	**0.76**	**6.10**	**0.65**	**0.76**	**0.93**	**0.90**
				Minkowski $p = 3$	0.71	0.71	0.64	5.95	0.54	0.66	0.89	0.88
				Minkowski $p = 5$	0.64	0.64	0.60	5.94	0.46	0.60	0.88	0.88
				Hamming	0.74	0.75	0.74	6.05	0.61	0.73	0.92	0.90
				Chebyshev	0.38	0.38	0.33	5.86	0.22	0.33	0.79	0.83
				Mahalanobis	0.74	0.75	0.74	6.05	0.61	0.73	0.92	0.90
				Correlation	0.78	0.77	0.69	6.02	0.57	0.33	0.79	0.90
				Cosine similarity	0.76	0.76	0.71	6.01	0.61	0.73	0.93	0.89
			5	Euclidean	0.70	0.70	0.71	5.71	0.62	0.73	0.93	0.91
				Manhattan	**0.75**	**0.75**	**0.76**	**5.79**	**0.70**	**0.79**	**0.94**	**0.92**
				Minkowski $p = 3$	0.71	0.71	0.64	5.95	0.54	0.66	0.89	0.88
				Minkowski $p = 5$	0.67	0.65	0.71	5.55	0.43	0.56	0.92	0.89
				Hamming	0.70	0.69	0.71	5.83	0.58	0.68	0.94	0.90
				Chebyshev	0.40	0.41	0.40	5.47	0.30	0.41	0.84	0.83
				Mahalanobis	0.70	0.69	0.71	5.83	0.58	0.68	0.94	0.90
				Correlation	0.76	0.75	0.69	5.73	0.57	0.41	0.84	0.89
				Cosine similarity	0.74	0.72	0.71	5.61	0.61	0.73	0.93	0.90
	SVC	Linear			**0.83**	**0.85**	**0.81**	**−0.12**	**0.71**	**0.80**	**0.97**	**0.98**
		Poly	Degree	2	**0.80**	**0.81**	**0.76**	**−0.14**	**0.65**	**0.75**	**0.97**	**0.95**
				3	0.78	0.79	0.69	−0.15	0.54	0.67	0.93	0.92
				4	0.77	0.76	0.69	−0.15	0.54	0.67	0.92	0.90
		RBF	Gamma	0.1	0.11	0.15	0.14	**−0.29**	0.11	0.17	0.74	0.38
				0.01	0.11	0.14	0.17	**−0.29**	0.13	0.19	0.73	0.22
				0.001	0.73	0.75	0.71	−0.19	0.60	0.72	0.94	0.92
				1	0.10	0.10	0.14	**−0.29**	0.11	0.17	0.74	0.39
				10	0.10	0.10	0.14	**−0.29**	0.11	0.17	0.74	0.39
	MLP	MLP			0.13	0.08	0.17	1.34	0.02	0.04	0.80	0.78
		StandardScaler			0.67	0.69	0.64	6.17	0.49	0.63	0.91	0.71
		MinMaxScaler			0.16	0.16	0.05	2.34	0.007	0.01	0.81	0.63
	LSTM				0.13	0.08	0.57	2.14	0.007	0.014	0.81	0.63
	CNN				**0.70**	**0.71**	**0.77**	**0.45**	**0.65**	**0.70**	**0.89**	**0.88**

performance is generally observed with K-NN $k = 1$ using correlation, followed by CNN and SVC polynomial degree 3, but in terms of ROC-AUC and cosine, linear SVC and K-NN $k = 1$ using Hamming distance show superiority.

In this experiment, a study was conducted on the semi-supervised learning method with a dataset split of 50:50, as outlined in Tables 20–22. On the JAFFE dataset, as shown in Table 20, the overall good performance is achieved by linear SVC, polynomial of degree 3, and K-NN with $k = 1$ using Minkowski distance with $p = 3$; the leading metric in ROC-AUC is CNN. In the Georgia tech dataset in Table 21, overall good performance and favorable results from both

Table 11: Training validation and testing results of supervised JAFFE (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
					4	5						
JAFFE	Decision tree				0.25	0.32	0.55	1.00	0.38	0.54	0.87	0.74
	K-NN	*k*	1	Euclidean	0.55	0.60	0.67	1.00	0.52	0.68	0.94	0.81
				Manhattan	0.55	0.59	0.65	1.00	0.49	0.66	0.94	0.80
				Minkowski p = 3	0.55	0.59	0.67	1.00	0.52	0.68	0.94	0.81
				Minkowski p = 5	0.57	0.61	0.64	1.00	0.47	0.63	0.94	0.79
				Hamming	0.49	0.55	0.59	1.00	0.42	0.59	0.93	0.76
				Chebyshev	0.43	0.44	0.45	1.00	0.42	0.59	0.93	0.68
				Mahalanobis	0.58	0.63	0.63	1.00	0.47	0.63	0.94	0.78
				Correlation	0.55	0.60	0.66	1.00	0.51	0.67	0.95	0.80
				Cosine similarity	0.55	0.60	0.67	1.00	0.52	0.68	0.95	0.81
			3	Euclidean	0.28	0.30	0.40	0.71	0.25	0.40	0.84	0.76
				Manhattan	0.26	0.29	0.35	0.72	0.21	0.34	0.84	0.73
				Minkowski p = 3	0.30	0.32	0.45	0.72	0.29	0.45	0.85	0.78
				Minkowski p = 5	0.24	0.28	0.47	0.71	0.30	0.46	0.87	0.79
				Hamming	0.24	0.25	0.30	0.72	0.17	0.29	0.85	0.69
				Chebyshev	0.35	0.38	0.42	0.71	0.26	0.40	0.78	0.73
				Mahalanobis	0.37	0.38	0.51	0.74	0.35	0.50	0.88	0.79
				Correlation	0.27	0.30	0.41	0.70	0.25	0.40	0.78	0.76
				Cosine similarity	0.25	0.28	0.40	0.70	0.25	0.39	0.83	0.75
			5	Euclidean	0.16	0.18	0.22	0.47	0.12	0.21	0.80	0.69
				Manhattan	0.15	0.15	0.19	0.48	0.10	0.17	0.78	0.67
				Minkowski p = 3	0.16	0.18	0.24	0.46	0.14	0.23	0.84	0.70
				Minkowski p = 5	0.16	0.18	0.34	0.46	0.21	0.34	0.84	0.74
				Hamming	0.17	0.18	0.24	0.46	0.14	0.24	0.84	0.63
				Chebyshev	0.26	0.23	0.36	0.50	0.22	0.36	0.80	0.73
				Mahalanobis	0.34	0.36	0.37	0.54	0.23	0.36	0.83	0.78
				Correlation	0.18	0.17	0.23	0.43	0.13	0.22	0.81	0.68
				Cosine similarity	0.18	0.16	0.23	0.46	0.13	0.23	0.82	0.67
	SVC	Linear			**0.63**	**0.66**	**0.79**	**−0.14**	**0.67**	**0.80**	**0.96**	**0.95**
		Poly	Degree	2	0.43	0.48	0.68	−0.15	0.52	0.68	0.93	0.92
				3	**0.62**	**0.65**	**0.77**	**−0.14**	**0.63**	**0.77**	**0.95**	**0.94**
				4	**0.60**	**0.66**	**0.76**	**−0.14**	**0.62**	**0.76**	**0.94**	**0.93**
		RBF	Gamma	0.1	0.13	0.18	0.15	**−0.16**	0.02	0.04	0.84	0.46
				0.01	0.24	0.41	0.43	**−0.16**	0.31	0.46	0.88	0.62
				0.001	0.30	0.38	0.50	**−0.16**	0.34	0.51	0.91	0.79
				1	0.17	0.24	0.15	**−0.16**	0.02	0.04	0.84	0.50
				10	0.11	0.14	0.15	**−0.16**	0.02	0.04	0.84	0.50
	MLP	MLP			0.15	0.13	0.24	0.03	0.11	0.15	0.89	0.55
		StandardScaler			0.45	0.54	0.67	0.92	0.52	0.68	0.92	0.88
		MinMaxScaler			0.13	0.19	0.19	0.09	0.04	0.08	0.81	0.56
	LSTM				0.17	0.22	0.15	**−0.16**	0.03	0.06	0.29	0.52
	CNN				0.28	0.26	0.40	−0.14	0.26	0.39	0.88	0.75

datasets are attained by linear SVC, polynomial of degree, and K-NN with k = 1 using Manhattan distance, with MLP algorithm excelling in ROC and AUC metrics. Meanwhile, in the Yale dataset in Table 22, good performance is generally observed in K-NN with k = 1 using Minkowski distance with p = 5 and linear SVC, but in the use of Kullback, *F*1-score, cosine, and ROC-AUC metrics, the leading algorithm is SVC with polynomial of degree 4.

Examining the accuracy performance depicted in Figure 3 on the JAFFE dataset yields noteworthy insights, particularly concerning the application of SVC with a linear kernel. The experimental outcomes reveal that the optimal

Table 12: Training validation and testing results of supervised Georgia tech (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.74	0.70	0.76	0.06	0.63	0.75	0.94	0.88
	K-NN	*k*	1	Euclidean	**0.96**	**0.94**	**0.97**	**0.01**	**0.95**	**0.97**	**0.99**	**0.99**
				Manhattan	**0.96**	**0.96**	**0.98**	**0.01**	**0.96**	**0.98**	**1.00**	**0.99**
				Minkowski $p = 3$	**0.95**	**0.92**	**0.97**	**0.01**	**0.94**	**0.96**	**0.99**	**0.98**
				Minkowski $p = 5$	0.94	0.93	0.95	0.02	0.91	0.95	0.99	0.97
				Hamming	0.95	0.95	0.96	0.01	0.94	0.96	0.99	0.98
				Chebyshev	0.84	0.82	0.88	0.04	0.80	0.88	0.98	0.94
				Mahalanobis	0.94	0.93	0.96	0.01	0.93	0.96	0.99	0.98
				Correlation	**0.96**	**0.94**	**0.97**	**0.01**	**0.95**	**0.97**	**0.99**	**0.99**
				Cosine similarity	**0.96**	**0.95**	**0.97**	**0.01**	**0.96**	**0.97**	**0.99**	**0.99**
			3	Euclidean	0.87	0.88	0.93	0.03	0.87	0.92	0.98	0.99
				Manhattan	0.89	0.89	0.92	0.03	0.86	0.91	0.98	0.99
				Minkowski $p = 3$	0.85	0.87	0.93	0.03	0.88	0.93	0.97	0.99
				Minkowski $p = 5$	0.85	0.86	0.91	0.03	0.85	0.91	0.97	0.98
				Hamming	0.88	0.89	0.93	0.02	0.88	0.92	0.98	0.98
				Chebyshev	0.73	0.75	0.80	0.07	0.70	0.79	0.95	0.95
				Mahalanobis	0.86	0.87	0.91	0.04	0.85	0.91	0.98	0.98
				Correlation	0.87	0.88	0.93	0.03	0.88	0.93	0.98	0.99
				Cosine similarity	0.87	0.88	0.93	0.03	0.88	0.93	0.97	0.99
			5	Euclidean	0.77	0.79	0.85	0.05	0.76	0.84	0.96	0.99
				Manhattan	0.78	0.80	0.86	0.05	0.77	0.84	0.96	0.99
				Minkowski $p = 3$	0.76	0.78	0.85	0.06	0.76	0.84	0.96	0.99
				Minkowski $p = 5$	0.74	0.76	0.83	0.06	0.73	0.82	0.96	0.98
				Hamming	0.77	0.80	0.85	0.04	0.77	0.85	0.96	0.99
				Chebyshev	0.64	0.66	0.73	0.11	0.60	0.72	0.95	0.96
				Mahalanobis	0.81	0.80	0.88	0.10	0.81	0.88	0.95	0.98
				Correlation	0.77	0.79	0.84	0.04	0.75	0.83	0.95	0.99
				Cosine similarity	0.77	0.79	0.85	0.05	0.76	0.84	0.96	0.99
	SVC	Linear			**0.96**	**0.95**	**0.97**	**0.05**	**0.96**	**0.97**	**1.00**	**0.99**
		Poly	Degree	2	**0.96**	**0.95**	**0.97**	**0.05**	**0.96**	**0.97**	**1.00**	**0.99**
				3	**0.96**	**0.95**	**0.98**	**0.05**	**0.96**	**0.97**	**1.00**	**0.99**
				4	**0.96**	**0.95**	**0.97**	**0.06**	**0.95**	**0.97**	**1.00**	**1.00**
		RBF	Gamma	0.1	0.07	0.16	0.33	8.17	0.31	0.43	0.69	0.08
				0.01	0.37	0.46	0.66	0.07	0.64	0.75	0.86	0.88
				0.001	**0.94**	**0.93**	**0.97**	**0.08**	**0.95**	**0.97**	**1.00**	**1.00**
				1	0.11	0.33	0.72	**0.004**	0.71	0.76	0.92	0.50
				10	0.02	0.03	0.06	5.86	0.04	0.05	0.86	0.50
	MLP	MLP			0.93	0.91	0.94	0.79	0.90	0.94	0.99	0.99
		StandardScaler			0.90	0.90	0.94	2.05	0.89	0.94	0.98	0.99
		MinMaxScaler			0.93	0.91	0.95	0.68	0.92	0.95	0.99	0.99
	LSTM				0.29	0.36	0.26	0.06	0.14	0.18	0.83	0.91
	CNN				**0.91**	**0.93**	**0.96**	**0.02**	**0.93**	**0.96**	**0.99**	**1.00**

accuracy is attained when employing linear SVC. Intriguingly, a discernible pattern emerges wherein the reduction in the parameter k leads to an increase in model accuracy, underscoring the efficacy of linear SVC in handling the JAFFE dataset. Furthermore, investigations into SVC with various kernels yield substantive results. Notably, a lower RBF parameter gamma corresponds to enhanced accuracy results, suggesting a direct relationship between decreased gamma values and improved model performance. These observations offer valuable perspectives on optimizing SVC parameters to enhance classification accuracy for JAFFE datasets (Figure 4).

Table 13: Training validation and testing results of supervised Yale (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.62	0.57	0.47	3.12	0.33	0.43	0.83	0.75
	K-NN	*k*	1	Euclidean	0.75	0.75	0.73	3.12	0.60	0.74	0.93	0.88
				Manhattan	0.76	0.76	0.73	3.12	0.61	0.75	0.92	0.87
				Minkowski p = 3	0.76	0.75	0.71	3.13	0.57	0.71	0.93	0.87
				Minkowski p = 5	0.74	0.70	0.71	3.12	0.57	0.71	0.92	0.87
				Hamming	**0.74**	**0.74**	**0.76**	**3.12**	**0.63**	**0.76**	**0.93**	**0.87**
				Chebyshev	0.45	0.46	0.40	3.12	0.63	0.75	0.93	0.87
				Mahalanobis	**0.74**	**0.74**	**0.76**	**3.12**	**0.63**	**0.76**	**0.93**	**0.87**
				Correlation	**0.74**	**0.75**	**0.77**	**3.12**	**0.65**	**0.78**	**0.94**	**0.89**
				Cosine similarity	0.75	0.76	0.73	3.12	0.63	0.78	0.94	0.89
			3	Euclidean	0.68	0.66	0.70	2.90	0.55	0.70	0.88	0.89
				Manhattan	0.70	0.68	0.72	2.90	0.59	0.73	0.90	0.91
				Minkowski p = 3	0.63	0.65	0.61	2.80	0.45	0.59	0.87	0.90
				Minkowski p = 5	0.54	0.58	0.58	2.80	0.41	0.56	0.86	0.89
				Hamming	0.73	0.74	0.67	2.90	0.54	0.68	0.92	0.87
				Chebyshev	0.41	0.40	0.33	2.80	0.21	0.30	0.78	0.76
				Mahalanobis	0.73	0.74	0.67	2.90	0.54	0.68	0.92	0.87
				Correlation	**0.69**	**0.68**	**0.76**	**2.90**	**0.64**	**0.30**	**0.78**	**0.89**
				Cosine similarity	0.65	0.69	0.66	2.80	0.52	0.67	0.89	0.90
			5	Euclidean	0.58	0.59	0.57	2.60	0.43	0.56	0.86	0.88
				Manhattan	0.68	0.64	0.70	2.60	0.57	0.70	0.91	0.90
				Minkowski p = 3	0.63	0.65	0.61	2.80	0.45	0.59	0.87	0.90
				Minkowski p = 5	0.54	0.58	0.71	2.60	0.57	0.71	0.92	0.90
				Hamming	0.70	0.70	0.69	2.56	0.55	0.69	0.93	0.90
				Chebyshev	0.34	0.40	0.37	2.57	0.26	0.36	0.84	0.81
				Mahalanobis	0.70	0.70	0.69	2.56	0.55	0.69	0.93	0.90
				Correlation	**0.58**	**0.62**	**0.76**	**2.63**	**0.64**	**0.36**	**0.84**	**0.91**
				Cosine similarity	0.59	0.64	0.66	2.63	0.52	0.67	0.89	0.90
	SVC	Linear			**0.79**	**0.80**	**0.81**	**−0.20**	**0.70**	**0.81**	**0.96**	**0.95**
		Poly	Degree	2	0.77	0.79	0.75	−0.21	0.61	0.74	0.94	0.93
				3	0.75	0.75	0.71	−0.21	0.56	0.70	0.91	0.90
				4	0.73	0.75	0.69	−0.21	0.53	0.67	0.91	**0.98**
		RBF	Gamma	0.1	0.13	0.12	0.08	**−0.25**	0.06	0.10	0.86	0.41
				0.01	0.14	0.12	0.08	**−0.25**	0.06	0.10	0.86	0.08
				0.001	0.64	0.64	0.57	**−0.25**	0.46	0.59	0.87	0.86
				1	0.13	0.13	0.08	**−0.25**	0.06	0.10	0.86	0.42
				10	0.13	0.13	0.08	**−0.25**	0.06	0.10	0.86	0.42
	MLP	MLP			0.14	0.16	0.11	1.05	0.03	0.04	0.66	0.66
		StandardScaler			0.68	0.68	0.59	3.00	0.42	0.57	0.87	0.71
		MinMaxScaler			0.13	0.13	0.16	1.42	0.05	0.07	0.86	0.74
	LSTM				0.62	0.63	0.65	6.10	0.51	0.31	**0.98**	0.88
	CNN				0.74	0.72	0.70	5.61	0.61	0.73	0.93	0.90

In general, the preeminent dataset employed in this study is the Georgia tech dataset as in Figure 5, surpassing the other two datasets in terms of relevance and significance. Nevertheless, a noteworthy observation emerges concerning the application of SVM with a RBF kernel and a gamma parameter set to 10, as the model's accuracy exhibits a markedly low level in this specific configuration.

In the context of the experimental analysis depicted, the application of linear SVC to the Yale dataset manifests the highest level of accuracy. It is discerned that with each augmentation in the degree of polynomial usage, there

Table 14: Training validation and testing results of semi-supervised JAFFE (20:80)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	Algorithm				4	5						
JAFFE	Decision tree				0.89	0.90	0.27	2.52	0.14	0.23	0.75	0.57
	K-NN	*k*	1	Euclidean	0.87	0.86	0.36	2.52	0.23	0.37	0.74	0.62
				Manhattan	0.86	0.86	0.36	2.52	0.24	0.37	0.70	0.62
				Minkowski $p = 3$	0.80	0.81	0.36	2.52	0.22	0.36	0.77	0.62
				Minkowski $p = 5$	**0.77**	**0.78**	**0.38**	**2.52**	**0.25**	**0.39**	**0.79**	**0.64**
				Hamming	0.84	0.84	0.29	2.52	0.16	0.27	0.72	0.59
				Chebyshev	0.58	0.62	0.31	2.52	0.18	0.30	0.82	0.60
				Mahalanobis	0.88	0.86	0.36	2.52	0.24	0.37	0.71	0.62
				Correlation	0.83	0.80	0.33	2.52	0.21	0.34	0.74	0.61
				Cosine similarity	0.83	0.80	0.35	2.52	0.22	0.36	0.76	0.62
			3	Euclidean	0.73	0.71	0.16	2.07	0.05	0.10	0.64	0.59
				Manhattan	0.75	0.75	0.20	2.13	0.09	0.15	0.59	0.59
				Minkowski $p = 3$	0.67	0.65	0.18	2.06	0.07	0.11	0.75	0.59
				Minkowski $p = 5$	0.62	0.62	0.22	2.13	0.09	0.17	0.79	0.60
				Hamming	0.69	0.69	0.18	2.04	0.09	0.15	0.46	0.59
				Chebyshev	0.43	0.43	0.40	1.93	0.09	0.15	0.46	0.63
				Mahalanobis	0.77	0.77	0.20	2.28	0.10	0.18	0.73	0.58
				Correlation	0.72	0.70	0.16	2.10	0.06	0.10	0.74	0.57
				Cosine similarity	0.70	0.72	0.18	2.06	0.07	0.12	0.66	0.57
			5	Euclidean	0.63	0.65	0.16	1.84	0.06	0.10	0.66	0.59
				Manhattan	0.74	0.74	0.18	1.88	0.07	0.13	0.62	0.58
				Minkowski $p = 3$	0.56	0.56	0.18	1.81	0.07	0.13	0.66	0.58
				Minkowski $p = 5$	0.52	0.53	0.22	1.65	0.09	0.16	0.75	0.58
				Hamming	0.68	0.65	0.13	1.67	0.07	0.12	0.55	0.56
				Chebyshev	0.39	0.43	0.18	1.43	0.21	0.15	0.72	0.57
				Mahalanobis	0.58	0.58	0.20	1.83	0.09	0.16	0.77	0.57
				Correlation	0.64	0.64	0.13	1.78	0.04	0.08	0.67	0.54
				Cosine similarity	0.73	0.75	0.13	1.83	0.04	0.08	0.63	0.54
	SVC	Linear			**0.85**	**0.86**	**0.38**	**−0.02**	**0.24**	**0.38**	**0.74**	**0.61**
		Poly	Degree	2	0.73	0.77	0.31	−0.04	0.18	0.29	0.79	0.59
				3	**0.84**	**0.87**	**0.38**	**−0.03**	**0.24**	**0.38**	**0.74**	**0.58**
				4	0.82	0.85	0.36	−0.01	0.22	0.36	0.74	0.59
		RBF	Gamma	0.1	0.38	0.38	0.13	**−0.16**	0.02	0.03	**0.84**	0.48
				0.01	0.41	0.40	0.20	−0.14	0.07	0.12	0.78	0.54
				0.001	0.83	0.79	0.31	−0.09	0.17	0.27	0.72	0.54
				1	0.56	0.56	0.13	−0.15	0.02	0.03	**0.84**	0.49
				10	0.83	0.83	0.13	−0.04	0.01	0.03	**0.84**	0.50
	MLP	MLP			0.62	0.83	0.13	1.41	0.02	0.03	**0.84**	0.52
		StandardScaler			0.71	0.73	0.29	2.39	0.16	0.25	0.72	0.62
		MinMaxScaler			0.71	0.68	0.13	2.10	0.03	0.05	0.83	0.51
	LSTM				0.83	0.83	0.13	−0.08	0.01	0.03	**0.84**	0.49
	CNN				0.76	0.77	0.36	0.35	0.21	0.34	0.74	**0.65**

is a corresponding decrease in accuracy. Conversely, a lower value of RBF gamma yields superior accuracy. These empirical observations underscore the nuanced influence of polynomial degree and RBF gamma values on the accuracy of the linear SVC algorithm when applied to the specified Yale dataset, as evidenced by the graphical representation in Figure 5.

Within the scope of this study, the primary datasets under scrutiny comprise JAFFE, Georgia tech, and Yale. Among these datasets, notable variations in performance are evident, with the Georgia tech dataset notably exhibiting the highest average performance. A thorough examination of data originating from these three sources facilitates a more comprehensive comprehension of each

Table 15: Training validation and testing results of semi-supervised Georgia tech (20:80)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.63	0.61	0.51	2.37	0.39	0.47	0.90	0.75
	K-NN	*k*	1	Euclidean	0.98	0.98	0.91	0.35	0.85	0.90	0.96	0.95
				Manhattan	**0.99**	**0.98**	**0.92**	**0.35**	**0.87**	**0.91**	**0.96**	**0.96**
				Minkowski $p = 3$	0.97	0.97	0.91	0.35	0.86	0.90	0.96	0.95
				Minkowski $p = 5$	0.89	0.97	0.89	0.36	0.83	0.88	0.96	0.94
				Hamming	0.98	0.97	0.91	0.06	0.87	0.91	0.97	0.95
				Chebyshev	0.90	0.89	0.74	1.03	0.62	0.71	0.90	0.87
				Mahalanobis	0.98	0.98	0.89	0.39	0.84	0.88	0.94	0.94
				Correlation	0.98	0.99	0.91	0.36	0.86	0.90	0.96	0.95
				Cosine similarity	0.99	0.99	0.91	0.36	0.86	0.90	0.96	0.95
			3	Euclidean	0.93	0.94	0.67	0.56	0.54	0.62	0.90	0.90
				Manhattan	0.92	0.92	0.68	0.54	0.57	0.64	0.89	0.90
				Minkowski $p = 3$	0.92	0.92	0.67	0.55	0.54	0.62	0.89	0.90
				Minkowski $p = 5$	0.92	0.92	0.67	1.12	0.52	0.62	0.91	0.89
				Hamming	0.90	0.91	0.69	0.50	0.57	0.64	0.91	0.90
				Chebyshev	0.74	0.74	0.53	1.53	0.39	0.48	0.88	0.84
				Mahalanobis	0.91	0.91	0.63	1.22	0.54	0.61	0.87	0.87
				Correlation	0.93	0.93	0.67	0.56	0.55	0.63	0.90	0.90
				Cosine similarity	0.92	0.91	0.67	0.84	0.54	0.62	0.90	0.90
			5	Euclidean	0.85	0.84	0.43	0.47	0.26	0.33	0.84	0.92
				Manhattan	0.83	0.83	0.42	0.75	0.25	0.32	0.86	0.92
				Minkowski $p = 3$	0.84	0.84	0.39	0.78	0.23	0.29	0.83	0.91
				Minkowski $p = 5$	0.81	0.82	0.37	0.82	0.20	0.26	0.84	0.91
				Hamming	0.81	0.81	0.41	0.71	0.25	0.32	0.87	0.89
				Chebyshev	0.65	0.63	0.42	1.55	0.25	0.33	0.86	0.85
				Mahalanobis	0.74	0.74	0.41	1.23	0.30	0.36	0.84	0.88
				Correlation	0.84	0.84	0.45	0.49	0.28	0.34	0.85	0.92
				Cosine similarity	0.85	0.83	0.40	0.50	0.24	0.30	0.83	0.91
	SVC	Linear			**0.98**	**0.99**	**0.92**	**0.05**	**0.87**	**0.91**	**0.97**	**0.95**
		Poly	Degree	2	0.99	0.99	0.90	0.05	0.85	0.88	0.95	0.96
				3	**0.99**	**0.99**	**0.92**	**0.06**	**0.87**	**0.91**	**0.96**	**0.96**
				4	**0.98**	**0.98**	**0.92**	**0.07**	**0.87**	**0.91**	**0.96**	**0.96**
		RBF	Gamma	0.1	0.30	0.30	0.02	0.99	0.0004	0.001	0.84	0.44
				0.01	0.62	0.63	0.13	0.64	0.91	0.11	0.14	0.87
				0.001	0.95	0.95	0.87	0.12	0.84	0.88	0.96	0.92
				1	0.57	0.57	0.02	0.96	0.0003	0.001	0.86	0.50
				10	0.80	0.80	0.02	1.52	0.0003	0.001	0.86	0.50
	MLP	MLP			0.69	0.72	0.16	1.22	0.09	0.12	0.79	0.94
		StandardScaler			0.88	0.90	0.83	0.24	0.73	0.81	0.95	0.96
		MinMaxScaler			0.69	0.69	0.17	1.38	0.10	0.12	0.82	0.95
	LSTM				0.46	0.53	0.06	0.02	0.09	0.12	0.85	0.89
	CNN				0.86	0.89	0.65	2.69	0.52	0.58	0.92	0.95

dataset's characteristics, thereby elucidating the specific advantages inherent to the Georgia tech dataset within the context of this research. These findings constitute a significant contribution to the judicious selection of datasets for the analytical and model development aspects delineated in this article.

Overall, this proves that supervised learning is a prevalent machine learning technique effective for tasks requiring labeled data. However, its reliance on abundant labeled data poses challenges due to the associated costs and difficulty in acquisition. In contrast, semi-supervised learning integrates both labeled and unlabeled data,

Table 16: Training validation and testing results of semi-supervised Yale (20:80)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.53	0.53	0.16	5.97	0.11	0.15	0.84	0.55
	K-NN	*k*	1	Euclidean	0.79	0.80	0.62	5.97	0.45	0.58	0.87	0.80
				Manhattan	0.82	0.81	0.62	5.97	0.45	0.58	0.87	0.80
				Minkowski $p = 3$	0.78	0.78	0.58	5.97	0.39	0.51	0.88	0.77
				Minkowski $p = 5$	**0.78**	**0.78**	**0.76**	**5.97**	**0.33**	**0.40**	**0.89**	**0.80**
				Hamming	0.82	0.82	0.62	5.97	0.48	0.59	0.89	0.80
				Chebyshev	0.65	0.64	0.27	5.97	0.17	0.26	0.77	0.61
				Mahalanobis	0.82	0.82	0.62	5.97	0.48	0.59	0.89	0.80
				Correlation	**0.81**	**0.81**	**0.64**	**5.97**	**0.49**	**0.58**	**0.88**	**0.81**
				Cosine similarity	0.79	0.79	0.62	5.97	0.44	0.55	0.89	0.80
			3	Euclidean	0.67	0.67	0.31	5.52	0.17	0.24	0.76	0.77
				Manhattan	0.70	0.69	0.38	5.56	0.27	0.33	0.75	0.76
				Minkowski $p = 3$	0.63	0.64	0.31	5.55	0.20	0.27	0.74	0.77
				Minkowski $p = 5$	**0.67**	**0.69**	**0.65**	**5.57**	**0.20**	**0.27**	**0.74**	**0.73**
				Hamming	0.56	0.57	0.36	5.48	0.25	0.31	0.79	0.73
				Chebyshev	0.51	0.50	0.20	5.40	0.10	0.15	0.68	0.72
				Mahalanobis	0.56	0.57	0.36	5.48	0.25	0.31	0.79	0.73
				Correlation	0.68	0.67	0.40	5.59	0.27	0.37	0.76	0.77
				Cosine similarity	0.71	0.72	0.47	5.57	0.31	0.40	0.78	0.81
			5	Euclidean	0.67	0.67	0.31	5.52	0.17	0.24	0.76	0.77
				Manhattan	0.62	0.60	0.29	5.38	0.19	0.26	0.66	0.82
				Minkowski $p = 3$	0.63	0.61	0.24	5.36	0.11	0.15	0.76	0.76
				Minkowski $p = 5$	0.63	0.62	0.27	5.27	0.12	0.18	0.75	0.74
				Hamming	0.67	0.69	0.33	5.35	0.22	0.26	0.77	0.83
				Chebyshev	0.45	0.46	0.13	5.04	0.03	0.06	0.39	0.73
				Mahalanobis	0.67	0.69	0.33	5.35	0.22	0.26	0.77	0.83
				Correlation	0.73	0.76	0.29	5.39	0.17	0.22	0.66	0.80
				Cosine similarity	0.72	0.73	0.29	5.35	0.13	0.18	0.68	0.82
	SVC	Linear			0.77	0.79	0.62	−0.08	0.45	0.58	0.89	**0.93**
		Poly	Degree	2	0.81	**0.83**	0.60	−0.10	0.44	0.56	0.89	0.90
				3	0.74	**0.83**	0.62	−0.09	0.47	0.60	**0.93**	0.89
				4	0.77	0.80	0.53	−0.09	0.39	0.53	0.91	0.86
		RBF	Gamma	0.1	0.23	0.25	0.09	**−0.12**	0.03	0.04	0.85	0.52
				0.01	0.30	0.30	0.09	−0.09	0.03	0.04	0.85	0.49
				0.001	0.76	0.76	0.40	−0.01	0.23	0.30	0.90	0.74
				1	0.79	0.79	0.09	0.11	0.03	0.04	0.85	0.48
				10	0.82	0.82	0.07	0.12	0.004	0.008	0.85	0.50
	MLP	MLP			0.20	0.17	0.13	1.79	0.06	0.06	0.76	0.66
		StandardScaler			0.66	0.67	0.60	1.89	0.01	0.32	0.66	0.77
		MinMaxScaler			0.21	0.23	0.20	2.30	0.21	0.32	0.77	0.87
	LSTM				0.33	0.31	0.32	5.97	0.56	0.11	0.88	0.32
	CNN				0.56	0.57	0.55	5.97	0.10	0.16	0.77	0.50

aiming to enhance model performance. While this approach offers cost-effectiveness and flexibility, handling noise in unlabeled data presents ongoing challenges ending up in less performance. Nonetheless, semi-supervised learning presents a promising alternative for various machine learning applications, particularly for certain selected and deep-based algorithms [56,57].

3.1 Statistical test results

Table 23 evaluates various metrics used to assess model performance in supervised learning and semi-supervised learning tasks, aiming to understand the effectiveness of each metric in distinguishing model performance [58]. The null hypothesis established for the Friedman test asserts

Table 17: Training validation and testing results of semi-supervised JAFFE (25:75)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
JAFFE	Decision tree				0.77	**0.82**	0.20	2.52	0.11	0.20	0.77	0.53
	K-NN	*k*	1	Euclidean	0.80	0.81	0.38	2.52	0.26	0.40	0.76	0.64
				Manhattan	0.77	0.78	0.36	2.52	0.24	0.37	0.72	0.62
				Minkowski p = 3	0.80	0.81	0.36	2.52	0.22	0.36	0.77	0.62
				Minkowski p = 5	0.77	0.78	0.38	2.52	0.25	0.39	0.79	0.64
				Hamming	0.77	0.80	0.29	2.52	0.17	0.29	0.73	0.59
				Chebyshev	**0.65**	**0.68**	**0.40**	**2.52**	**0.26**	**0.41**	**0.89**	**0.65**
				Mahalanobis	0.77	0.79	0.33	2.52	0.27	0.36	0.78	0.61
				Correlation	**0.81**	0.80	0.38	2.52	0.25	0.39	0.78	0.64
				Cosine similarity	**0.79**	**0.82**	**0.40**	**2.52**	**0.28**	**0.42**	**0.80**	**0.65**
			3	Euclidean	0.64	0.67	0.16	2.04	0.06	0.12	0.70	0.60
				Manhattan	0.71	0.73	0.18	2.06	0.08	0.13	0.59	0.61
				Minkowski p = 3	0.73	0.71	0.16	2.07	0.06	0.10	0.64	0.59
				Minkowski p = 5	0.75	0.75	0.20	2.13	0.09	0.15	0.60	0.59
				Hamming	0.65	0.65	0.18	2.17	0.07	0.13	0.58	0.61
				Chebyshev	0.43	0.38	0.33	1.00	0.19	0.31	0.74	0.67
				Mahalanobis	0.62	0.61	0.27	2.13	0.14	0.24	0.78	0.60
				Correlation	0.67	0.68	0.18	2.08	0.08	0.14	0.76	0.59
				Cosine similarity	0.66	0.64	0.27	2.04	0.14	0.23	0.73	0.62
			5	Euclidean	0.64	0.67	0.16	2.04	0.12	0.69	0.69	0.60
				Manhattan	0.62	0.62	0.18	1.81	0.07	0.13	0.72	0.57
				Minkowski p = 3	0.67	0.65	0.18	2.06	0.07	0.11	0.75	0.59
				Minkowski p = 5	0.62	0.62	0.22	2.13	0.10	0.17	0.79	0.60
				Hamming	0.58	0.63	0.18	1.71	0.08	0.15	0.64	0.60
				Chebyshev	0.39	0.38	0.24	1.43	0.13	0.22	0.76	0.65
				Mahalanobis	0.58	0.62	0.18	1.86	0.06	0.10	0.76	0.58
				Correlation	0.64	0.65	0.13	1.82	0.05	0.09	0.73	0.54
				Cosine similarity	0.65	0.67	0.16	1.80	0.05	0.08	0.71	0.58
	SVC	Linear			**0.78**	**0.80**	**0.40**	**−0.01**	**0.28**	**0.42**	**0.77**	**0.68**
		Poly	Degree	2	0.65	0.70	0.31	−0.08	0.18	0.30	0.81	0.64
				3	0.75	0.77	0.38	−0.04	0.26	0.40	0.76	0.65
				4	0.77	0.80	0.38	−0.03	0.25	0.40	0.74	0.65
		RBF	Gamma	0.1	0.69	0.69	0.16	−0.13	0.02	0.04	0.84	0.50
				0.01	0.69	0.69	0.16	−0.13	0.02	0.04	0.84	0.50
				0.001	0.69	0.68	0.38	−0.11	0.24	0.38	0.78	0.58
				1	0.45	0.45	0.13	**−0.21**	0.02	0.03	0.84	0.51
				10	0.69	0.69	0.16	−0.13	0.02	0.04	0.84	0.51
	MLP	MLP			0.56	0.58	0.18	1.68	0.04	0.08	0.86	0.52
		StandardScaler			0.74	0.75	0.38	2.41	0.26	0.40	0.80	0.65
		MinMaxScaler			0.47	0.64	0.18	1.36	0.06	0.10	0.80	0.51
	LSTM				0.69	0.67	0.20	−0.15	0.05	0.09	0.83	0.54
	CNN				0.69	0.74	0.33	0.16	0.21	0.33	0.78	**0.68**

that all classification models exhibit equivalent performance. Rejecting this null hypothesis indicates that there is a significant performance difference in one or more of the models with a significance level at α = 0.05 and a total of 42 models compared. In each test, the p-value indicates the probability that observed differences arose by chance where the p-values in all dataset scenarios are less than 0.05 in both supervised and semi-supervised learning, and this suggests that the difference is significant, justifying the rejection of the null hypothesis.

Following the Friedman test's identification of statistically significant performance differences across multiple

Table 18: Training validation and testing results of semi-supervised Georgia tech (25:75)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.62	0.65	0.53	1.58	0.39	0.49	0.87	0.76
	K-NN	*k*	1	Euclidean	0.99	0.99	0.91	0.76	0.86	0.90	0.96	0.95
				Manhattan	**1.00**	**1.00**	**0.92**	**0.76**	**0.87**	**0.90**	**0.97**	**0.96**
				Minkowski $p = 3$	0.98	0.98	0.91	0.76	0.85	0.89	0.96	0.95
				Minkowski $p = 5$	0.98	0.98	0.88	0.78	0.81	0.86	0.96	0.94
				Hamming	0.97	0.97	0.91	0.41	0.87	0.90	0.98	0.95
				Chebyshev	0.90	0.90	0.77	1.56	0.67	0.75	0.94	0.88
				Mahalanobis	0.98	0.98	0.89	1.13	0.83	0.87	0.95	0.94
				Correlation	0.99	0.99	0.91	0.76	0.86	0.90	0.96	0.95
				Cosine similarity	0.99	0.99	0.91	0.76	0.86	0.90	0.96	0.95
			3	Euclidean	0.94	0.94	0.73	0.52	0.64	0.71	0.92	0.93
				Manhattan	0.93	0.94	0.77	0.83	0.67	0.75	0.93	0.94
				Minkowski $p = 3$	0.93	0.92	0.72	0.85	0.62	0.71	0.92	0.92
				Minkowski $p = 5$	0.94	0.93	0.72	0.88	0.62	0.71	0.93	0.92
				Hamming	0.92	0.93	0.75	0.46	0.66	0.73	0.93	0.93
				Chebyshev	0.79	0.78	0.63	0.95	0.50	0.58	0.90	0.89
				Mahalanobis	0.90	0.90	0.72	0.92	0.62	0.70	0.91	0.89
				Correlation	0.92	0.93	0.75	0.45	0.66	0.74	0.90	0.94
				Cosine similarity	0.93	0.93	0.73	0.50	0.64	0.72	0.92	0.93
			5	Euclidean	0.90	0.90	0.61	0.63	0.48	0.55	0.90	0.95
				Manhattan	0.88	0.89	0.65	0.57	0.54	0.61	0.90	0.96
				Minkowski $p = 3$	0.88	0.89	0.61	0.64	0.47	0.55	0.90	0.95
				Minkowski $p = 5$	0.87	0.88	0.60	1.00	0.46	0.54	0.90	0.94
				Hamming	0.88	0.86	0.61	0.57	0.48	0.56	0.90	0.94
				Chebyshev	0.74	0.75	0.53	0.99	0.40	0.48	0.88	0.89
				Mahalanobis	0.83	0.82	0.66	0.97	0.57	0.64	0.89	0.93
				Correlation	0.99	0.99	0.91	0.76	0.86	0.90	0.96	0.95
				Cosine similarity	0.91	0.91	0.64	0.59	0.52	0.60	0.90	0.95
	SVC	Linear			**0.99**	**0.99**	**0.92**	**0.06**	**0.87**	**0.90**	**0.96**	**0.96**
		Poly	Degree	2	0.99	0.99	0.91	0.07	0.85	0.88	0.96	0.96
				3	**1.00**	**1.00**	**0.92**	**0.72**	**0.87**	**0.90**	**0.97**	**0.96**
				4	0.99	0.99	0.91	0.08	0.87	0.90	0.97	0.96
		RBF	Gamma	0.1	0.11	0.11	0.03	0.27	0.01	0.01	0.86	0.28
				0.01	0.33	0.34	0.19	0.31	0.17	0.22	0.87	0.74
				0.001	0.97	0.96	0.86	0.15	0.82	0.86	0.94	0.93
				1	0.19	0.19	0.03	0.18	0.01	0.02	0.86	0.50
				10	0.74	0.74	0.02	1.25	0.0003	0.0007	0.86	0.50
	MLP	MLP			0.64	0.64	0.08	1.82	0.57	0.08	0.87	0.95
		StandardScaler			0.92	0.93	0.80	0.55	0.70	0.77	0.92	0.96
		MinMaxScaler			0.63	0.65	0.21	1.02	0.14	0.18	0.76	0.95
	LSTM				0.53	0.56	0.02	0.03	0.13	0.17	0.86	0.89
	CNN				0.95	0.95	0.85	3.45	0.76	0.82	0.95	**0.97**

datasets, the Nemenyi post hoc test is employed to detect specific differences between each pair of models. With 42 models under evaluation, this leads to $42 \times 42 = 1{,}764$ pairwise comparison models. The Nemenyi post hoc test assesses each model pair to ascertain whether their performance ranks differ significantly, thereby offering a more detailed insight into which models perform better or worse relative to others across the datasets. This thorough comparison enables precise conclusions regarding the comparative effectiveness of individual models.

Table 19: Training validation and testing results of semi-supervised Yale (25:75)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.56	0.60	0.40	5.97	0.33	0.41	0.85	0.68
	K-NN	*k*	1	Euclidean	0.87	0.92	0.71	5.71	0.58	0.72	0.87	0.85
				Manhattan	0.91	0.92	0.73	5.97	0.60	0.73	0.87	0.86
				Minkowski p = 3	0.86	0.86	0.69	5.97	0.55	0.69	0.89	0.83
				Minkowski p = 5	0.86	0.86	0.60	5.97	0.45	0.58	0.89	0.78
				Hamming	0.83	0.86	0.71	5.97	0.62	0.72	**0.90**	0.85
				Chebyshev	0.57	0.62	0.40	5.97	0.28	0.39	0.82	0.68
				Mahalanobis	0.83	0.86	0.71	5.97	0.62	0.72	**0.90**	0.85
				Correlation	**0.87**	**0.85**	**0.78**	**5.97**	**0.67**	**0.78**	**0.87**	**0.88**
				Cosine similarity	0.88	0.87	0.71	5.97	0.58	0.71	0.86	0.85
			3	Euclidean	0.75	0.76	0.62	5.64	0.46	0.58	0.83	0.87
				Manhattan	0.82	0.82	0.67	5.66	0.53	0.64	0.85	0.89
				Minkowski p = 3	0.70	0.67	0.60	5.65	0.44	0.55	0.80	0.87
				Minkowski p = 5	0.62	0.64	0.47	5.52	0.31	0.42	0.73	0.82
				Hamming	0.73	0.72	0.49	5.65	0.38	0.45	0.78	0.81
				Chebyshev	0.47	0.49	0.24	5.35	0.15	0.22	0.66	0.73
				Mahalanobis	0.73	0.72	0.49	5.65	0.38	0.45	0.78	0.81
				Correlation	0.70	0.72	0.64	5.61	0.50	0.62	0.81	0.90
				Cosine similarity	0.69	0.70	0.67	5.60	0.52	0.65	0.86	0.90
			5	Euclidean	0.75	0.76	0.62	0.64	0.46	0.58	0.83	0.87
				Manhattan	0.59	0.61	0.60	5.30	0.50	0.59	0.83	0.92
				Minkowski p = 3	0.64	0.66	0.60	5.19	0.41	0.53	0.75	0.90
				Minkowski p = 5	0.52	0.55	0.53	5.13	0.37	0.46	0.76	0.88
				Hamming	0.55	0.57	0.31	5.33	0.21	0.26	0.80	0.85
				Chebyshev	0.41	0.46	0.67	4.88	0.30	0.76	0.78	0.78
				Mahalanobis	0.55	0.57	0.31	5.33	0.21	0.26	0.80	0.85
				Correlation	0.62	0.59	0.56	5.28	0.38	0.48	0.78	0.90
				Cosine similarity	0.66	0.63	0.58	5.41	0.43	0.51	0.80	0.94
	SVC	Linear			**0.94**	**0.93**	0.73	−0.11	0.62	0.72	0.83	**0.98**
		Poly	Degree	2	0.91	0.90	0.71	−0.11	0.61	0.70	0.83	0.95
				3	**0.88**	**0.89**	**0.76**	**−0.10**	**0.64**	**0.74**	**0.87**	**0.94**
				4	0.82	0.81	0.71	−0.11	0.61	0.71	0.85	0.92
		RBF	Gamma	0.1	0.21	0.19	0.13	**−0.24**	0.07	0.11	0.84	0.46
				0.01	0.21	0.19	0.13	**−0.24**	0.07	0.11	0.84	0.46
				0.001	0.80	0.82	0.71	−0.12	0.57	0.68	0.83	0.92
				1	0.67	0.67	0.11	0.001	0.05	0.07	0.84	0.47
				10	0.78	0.78	0.07	0.05	0.004	0.08	0.85	0.51
	MLP	MLP			0.35	0.34	0.07	3.66	0.005	0.009	0.86	0.71
		StandardScaler			0.57	0.45	0.42	1.20	0.30	0.57	0.72	0.71
		MinMaxScaler			0.38	0.30	0.42	0.89	0.23	0.34	0.86	0.61
	LSTM				0.37	0.37	0.36	5.71	0.22	0.45	0.75	0.66
	CNN				**0.78**	**0.76**	**0.77**	**5.70**	**0.22**	**0.40**	**0.77**	**0.66**

In the context of supervised learning, metrics like CV 5 and accuracy exhibit high test statistics and a large number of significant combinations (e.g., 274 and 192 model pairs, out of 1,764, respectively) obtained from Nemenyi post hoc test, demonstrating their strength in identifying performance differences among models.

In semi-supervised learning analysis, there is a slight variation in metrics' ability to differentiate performance. CV 4, accuracy, and *F*1-score continue to demonstrate strong statistical power, with CV 4 achieving 236 significant combinations and a high test statistic, while accuracy and *F*1-score show similarly robust results. These metrics, with

Table 20: Training validation and testing results of semi-supervised JAFFE (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
JAFFE	Decision tree				0.59	0.63	0.51	2.52	0.35	0.49	0.88	0.71
	K-NN	*k*	1	Euclidean	0.78	0.74	0.64	2.52	0.47	0.63	0.91	0.79
				Manhattan	0.78	0.80	0.62	2.52	0.46	0.62	0.89	0.78
				Minkowski $p = 3$	**0.74**	**0.76**	**0.69**	**2.52**	**0.53**	**0.68**	**0.91**	**0.82**
				Minkowski $p = 5$	0.75	0.74	0.64	2.52	0.47	0.63	0.93	0.79
				Hamming	0.77	0.79	0.51	2.52	0.36	0.51	0.92	0.71
				Chebyshev	0.68	0.69	0.53	2.52	0.37	0.53	0.90	0.73
				Mahalanobis	0.74	0.74	0.62	2.52	0.45	0.60	0.89	0.78
				Correlation	0.78	0.75	0.67	2.52	0.50	0.66	0.93	0.81
				Cosine similarity	0.80	0.77	0.64	2.52	0.47	0.63	0.94	0.79
			3	Euclidean	0.65	0.68	0.31	1.95	0.16	0.25	0.79	0.77
				Manhattan	0.65	0.64	0.29	1.99	0.15	0.25	0.77	0.75
				Minkowski $p = 3$	0.61	0.65	0.33	1.92	0.20	0.31	0.82	0.74
				Minkowski $p = 5$	0.54	0.55	0.36	1.96	0.21	0.33	0.79	0.80
				Hamming	0.62	0.63	0.27	2.13	0.13	0.22	0.79	0.69
				Chebyshev	0.53	0.55	0.40	1.95	0.24	0.38	0.72	0.72
				Mahalanobis	0.58	0.61	0.33	1.98	0.18	0.29	0.73	0.75
				Correlation	0.62	0.60	0.29	1.90	0.14	0.23	0.79	0.75
				Cosine similarity	0.64	0.68	0.33	1.90	0.18	0.29	0.81	0.77
			5	Euclidean	0.65	0.68	0.31	1.95	0.16	0.25	0.79	0.77
				Manhattan	0.53	0.55	0.24	1.60	0.12	0.19	0.80	0.69
				Minkowski $p = 3$	0.56	0.54	0.22	1.45	0.10	0.17	0.80	0.71
				Minkowski $p = 5$	0.53	0.54	0.31	1.52	0.19	0.30	0.77	0.75
				Hamming	0.59	0.61	0.24	1.66	0.12	0.21	0.84	0.66
				Chebyshev	0.43	0.46	0.33	1.46	0.21	0.34	0.77	0.75
				Mahalanobis	0.52	0.54	0.29	1.57	0.15	0.24	0.79	0.71
				Correlation	0.52	0.53	0.20	1.55	0.08	0.14	0.69	0.69
				Cosine similarity	0.48	0.51	0.27	1.50	0.14	0.23	0.79	0.69
	SVC	Linear			**0.81**	**0.84**	**0.73**	**−0.09**	**0.59**	**0.74**	**0.95**	**0.85**
		Poly	Degree	2	0.72	0.74	0.64	−0.13	0.52	0.67	0.92	0.85
				3	**0.81**	**0.80**	**0.69**	**−0.09**	**0.54**	**0.69**	**0.92**	**0.83**
				4	**0.83**	0.82	0.67	−0.10	0.52	0.68	0.92	0.84
		RBF	Gamma	0.1	0.50	0.51	0.13	−0.19	0.01	0.03	0.84	0.48
				0.01	0.42	0.42	0.13	**−0.20**	0.02	0.03	0.84	0.60
				0.001	0.67	0.69	0.44	−0.17	0.33	0.48	0.89	0.78
				1	0.29	0.29	0.13	−0.23	0.01	0.03	0.84	0.50
				10	0.57	0.57	0.13	−0.18	0.02	0.03	0.84	0.49
	MLP	MLP			0.45	0.27	0.13	1.45	0.01	0.03	0.84	0.51
		StandardScaler			0.62	0.61	0.53	2.32	0.38	0.54	0.87	0.77
		MinMaxScaler			0.34	0.32	0.20	0.06	0.08	0.13	0.80	0.54
	LSTM				0.42	0.37	0.18	−0.18	0.05	0.08	0.51	0.55
	CNN				0.70	0.69	0.58	−0.003	0.44	0.58	0.91	**0.88**

high counts of significant combinations and large test statistics, are particularly suitable for discerning model performance differences in semi-supervised contexts. In contrast, metrics like K-L divergence and cosine similarity consistently yield fewer significant combinations in both learning types, suggesting they may be less sensitive for comparing models within this framework. Overall, the table indicates that CV, accuracy, and *F*1-score are effective metrics for distinguishing models regardless of task type, with a clear conclusion that *p*-values below 0.05 signal a significant difference among the models.

Table 21: Training validation and testing results of semi-supervised Georgia tech (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Georgia tech	Decision tree				0.75	0.74	0.71	0.46	0.59	0.70	0.93	0.85
	K-NN	*k*	1	Euclidean	0.99	0.99	0.97	0.02	0.95	0.97	0.99	0.99
				Manhattan	**1.00**	**0.99**	**0.98**	**0.01**	**0.97**	**0.97**	**1.00**	**0.99**
				Minkowski $p = 3$	0.99	0.99	0.97	0.02	0.94	0.96	0.99	0.98
				Minkowski $p = 5$	0.98	0.98	0.92	0.40	0.87	0.91	0.98	0.96
				Hamming	**1.00**	0.38	0.95	0.38	0.93	0.94	0.98	0.98
				Chebyshev	0.94	0.94	0.88	0.05	0.82	0.88	0.86	0.94
				Mahalanobis	0.98	0.98	0.95	0.03	0.92	0.95	0.99	0.98
				Correlation	**1.00**	0.99	0.97	0.02	0.95	0.97	0.99	0.99
				Cosine similarity	**1.00**	0.99	0.97	0.01	0.95	0.97	0.99	0.99
			3	Euclidean	0.96	0.96	0.89	0.08	0.82	0.97	0.94	0.97
				Manhattan	0.96	0.98	0.89	0.41	0.83	0.87	0.95	0.97
				Minkowski $p = 3$	0.95	0.96	0.88	0.07	0.82	0.86	0.93	0.97
				Minkowski $p = 5$	0.96	0.97	0.88	0.81	0.87	0.95	0.96	0.96
				Hamming	0.97	0.97	0.89	0.06	0.83	0.88	0.96	0.98
				Chebyshev	0.86	0.84	0.76	0.14	0.65	0.74	0.92	0.95
				Mahalanobis	0.94	0.93	0.87	0.43	0.81	0.86	0.94	0.97
				Correlation	0.97	0.96	0.90	0.06	0.85	0.89	0.94	0.97
				Cosine similarity	0.97	0.97	0.88	0.08	0.82	0.86	0.93	0.97
			5	Euclidean	0.91	0.92	0.80	0.10	0.70	0.76	0.93	0.99
				Manhattan	0.92	0.93	0.82	0.10	0.73	0.78	0.94	0.99
				Minkowski $p = 3$	0.90	0.92	0.77	0.11	0.67	0.47	0.92	0.99
				Minkowski $p = 5$	0.88	0.88	0.77	0.09	0.67	0.75	0.94	0.98
				Hamming	0.92	0.93	0.81	0.07	0.73	0.78	0.94	0.99
				Chebyshev	0.80	0.15	0.70	0.14	0.59	0.67	0.91	0.95
				Mahalanobis	0.91	0.90	0.85	0.10	0.78	0.83	0.93	0.98
				Correlation	0.91	0.91	0.79	0.08	0.70	0.76	0.92	0.98
				Cosine similarity	0.90	0.91	0.79	0.10	0.69	0.75	0.92	0.99
	SVC	Linear			**0.99**	**0.99**	**0.98**	**0.04**	**0.97**	**0.98**	**0.98**	**0.99**
		Poly	Degree	2	**1.00**	**0.99**	**0.98**	**0.04**	**0.97**	**0.98**	**0.99**	**0.99**
				3	**1.00**	**1.00**	**0.98**	**0.04**	**0.97**	**0.98**	**1.00**	**0.99**
				4	**1.00**	**1.00**	**0.98**	**0.05**	**0.97**	**0.98**	**1.00**	**0.99**
		RBF	Gamma	0.1	0.32	0.32	0.03	0.35	0.01	0.01	0.86	0.42
				0.01	0.30	0.30	0.19	0.24	0.17	0.22	0.88	0.76
				0.001	0.96	0.97	0.92	0.11	0.88	0.91	0.96	0.99
				1	0.10	0.10	0.03	0.07	0.01	0.86	0.48	0.48
				10	0.49	0.49	0.02	0.75	0.0003	0.001	0.60	0.50
	MLP	MLP			0.93	0.93	0.87	0.07	0.80	0.85	0.97	0.99
		StandardScaler			0.95	0.95	0.91	0.10	0.85	0.89	0.95	**1.00**
		MinMaxScaler			0.97	0.93	0.91	0.07	0.86	0.89	0.95	**1.00**
	LSTM				0.42	0.60	0.65	0.03	0.20	0.24	0.87	0.91
	CNN				0.99	0.95	0.95	3.65	0.91	0.94	0.99	0.99

In addition to Nemenyi post hoc test, each model is ranked based on its performance in each dataset, and then these ranks are averaged across all datasets to obtain the model mean rank. This gives an overall ranking that reflects the model's performance on all datasets. The critical difference (CD) is determined using formula (6) to determine whether the performance differences between pairs of models are statistically significant. CD takes into the number of models ($k = 42$), the total number of dataset scenarios ($n = 9$), and the critical value ($q_\alpha = 1.64899$) involved using the studentized range distribution, which is obtained by using formula degree of freedom

Table 22: Training validation and testing results of semi-supervised Yale (50:50)

					Training validation		Testing evaluation					
					CV		Accuracy	K-L	I–J	*F*1-score	Cosine	ROC-AUC
Dataset	**Algorithm**				**4**	**5**						
Yale	Decision tree				0.54	0.50	0.44	5.97	0.31	0.42	0.76	0.70
	K-NN	*k*	1	Euclidean	0.90	0.88	0.73	5.97	0.59	0.73	0.90	0.86
				Manhattan	0.87	0.89	0.73	5.97	0.59	0.73	0.88	0.86
				Minkowski $p = 3$	0.86	0.84	0.71	5.97	0.56	0.71	0.91	0.85
				Minkowski $p = 5$	**0.84**	**0.84**	**0.86**	**5.97**	**0.56**	**0.71**	**0.91**	**0.85**
				Hamming	0.82	0.82	0.78	5.97	0.69	0.79	0.96	0.88
				Chebyshev	0.67	0.67	0.53	5.97	0.38	0.52	0.87	0.75
				Mahalanobis	0.82	0.82	0.78	5.97	0.69	0.79	0.96	0.88
				Correlation	0.86	0.87	0.80	5.97	0.69	0.79	0.91	0.89
				Cosine similarity	0.83	0.83	0.76	5.97	0.62	0.75	0.90	0.87
			3	Euclidean	0.83	0.83	0.73	5.73	0.60	0.71	0.91	0.90
				Manhattan	0.79	0.82	0.76	5.74	0.64	0.76	0.92	0.93
				Minkowski $p = 3$	0.78	0.74	0.73	5.76	0.59	0.71	0.94	0.89
				Minkowski $p = 5$	0.67	0.67	0.66	5.57	0.59	0.71	0.94	0.86
				Hamming	0.77	0.76	0.78	5.66	0.66	0.78	0.91	0.93
				Chebyshev	0.67	0.67	0.22	5.35	0.14	0.19	0.61	0.76
				Mahalanobis	0.77	0.76	0.78	5.66	0.66	0.78	0.91	0.93
				Correlation	0.82	0.80	0.73	5.69	0.59	0.71	0.90	0.91
				Cosine similarity	0.83	0.83	0.76	5.97	0.62	0.75	0.90	0.87
			5	Euclidean	0.83	0.83	0.73	5.73	0.60	0.71	0.91	0.90
				Manhattan	0.79	0.82	0.76	5.74	0.64	0.76	0.92	0.93
				Minkowski $p = 3$	0.78	0.74	0.73	5.76	0.59	0.71	0.94	0.89
				Minkowski $p = 5$	0.67	0.67	0.66	5.57	0.59	0.71	0.94	0.86
				Hamming	0.77	0.76	0.78	5.66	0.66	0.78	0.91	0.93
				Chebyshev	0.66	0.67	0.22	5.35	0.14	0.19	0.61	0.76
				Mahalanobis	0.77	0.76	0.78	5.66	0.66	0.78	0.91	0.93
				Correlation	0.82	0.80	0.73	5.69	0.59	0.71	0.90	0.92
				Cosine similarity	0.75	0.77	0.76	5.66	0.63	0.74	0.90	0.91
	SVC	Linear			**0.92**	**0.89**	**0.84**	**−0.12**	**0.75**	**0.84**	**0.95**	**0.99**
		Poly	Degree	2	0.88	0.88	0.78	−0.12	0.66	0.78	0.94	**0.99**
				3	0.87	0.87	0.73	−0.12	0.60	0.72	0.88	0.97
				4	0.87	0.87	0.67	**−0.80**	0.78	**0.88**	**0.99**	**0.99**
		RBF	Gamma	0.1	0.24	0.24	0.20	−0.23	0.05	0.07	0.85	0.42
				0.01	0.38	0.38	0.13	−0.19	0.07	0.11	0.84	0.37
				0.001	0.74	0.76	0.69	−0.15	0.63	0.70	0.90	0.96
				1	0.51	0.51	0.11	−0.16	0.05	0.07	0.84	0.47
				10	0.57	0.57	0.09	−0.13	0.03	0.04	0.85	0.48
	MLP	MLP			0.37	0.42	0.07	4.17	0.004	0.008	0.22	0.68
		StandardScaler			0.38	0.30	0.42	0.89	0.23	0.34	0.86	0.61
		MinMaxScaler			0.38	0.30	0.42	0.89	0.23	0.34	0.86	0.61
	LSTM				0.65	0.63	0.64	5.97	0.11	0.55	0.60	0.55
	CNN				**0.78**	**0.77**	**0.81**	**5.97**	**0.42**	**0.55**	**0.88**	**0.77**

df = $n \times (k - 1) = 369$. If the difference in mean ranks between two models is greater than the CD, those models are considered to perform significantly differently. These calculations help identify which models show significant performance differences, providing insights into which models are more effective.

$$\mathrm{CD} = q_\alpha \times \sqrt{\frac{k(k+1)}{6n}}. \tag{6}$$

Figures 6 and 7 elucidate the pairwise comparison results of 42 models, showing significant differences in supervised and semi-supervised learning, respectively.

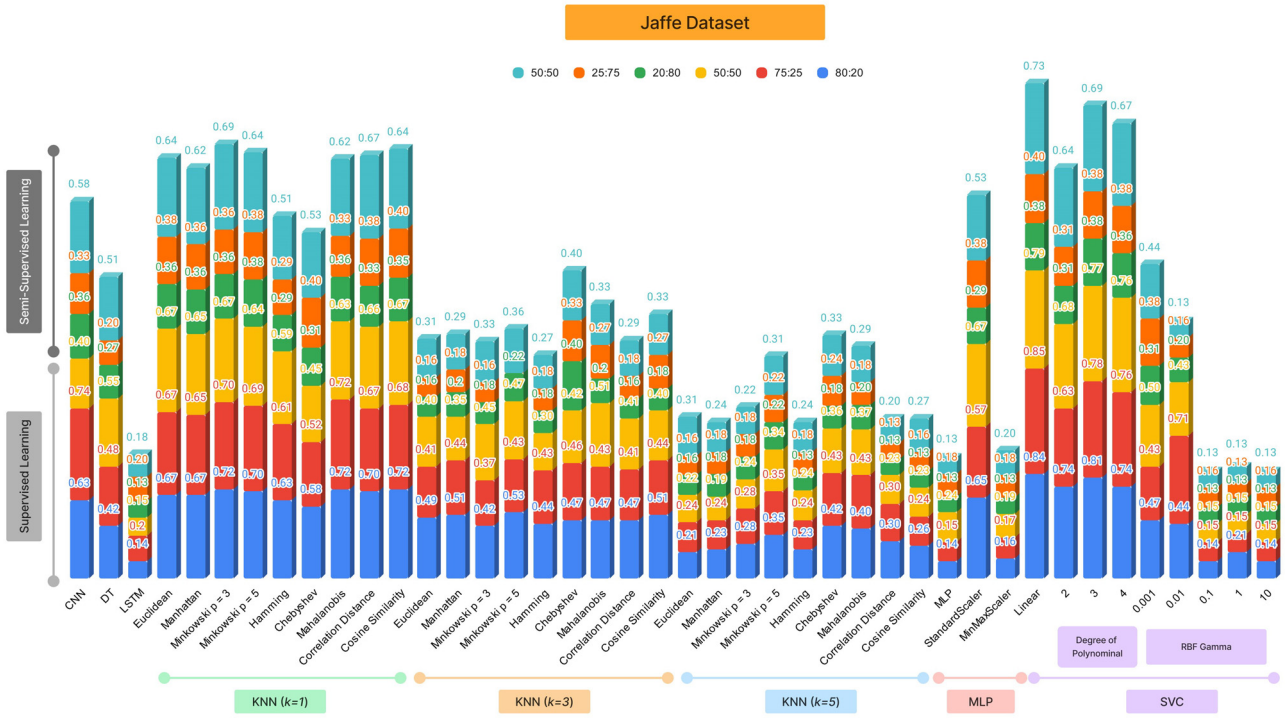


Figure 3: Visualization of accuracy performance on JAFFE dataset.

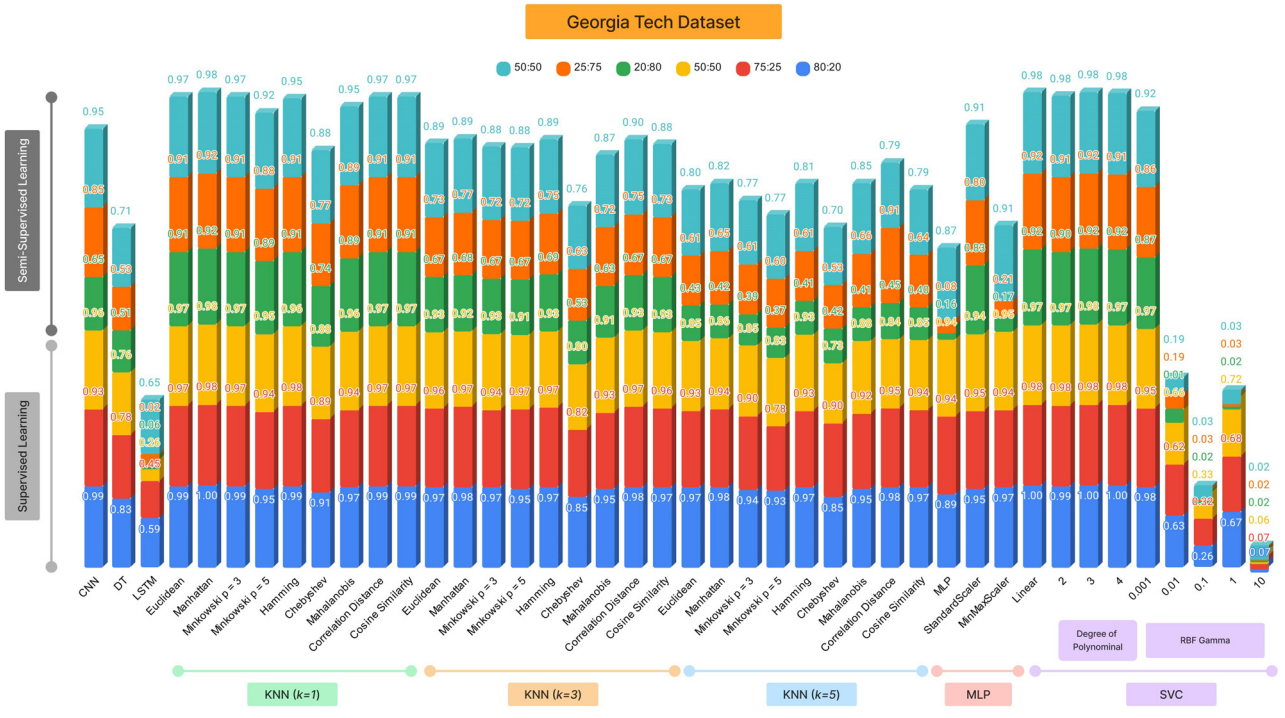


Figure 4: Visualization of accuracy performance on Georgia tech dataset.

Overall, it is evident that all 42 models differ significantly, as indicated by lines connecting the mean ranks of each model. Pairs with a mean rank difference smaller than the CD are significantly different from one another. The SVC method with linear and polynomial kernels is identified as the most significantly different methods. In addition, K-NN and CNN

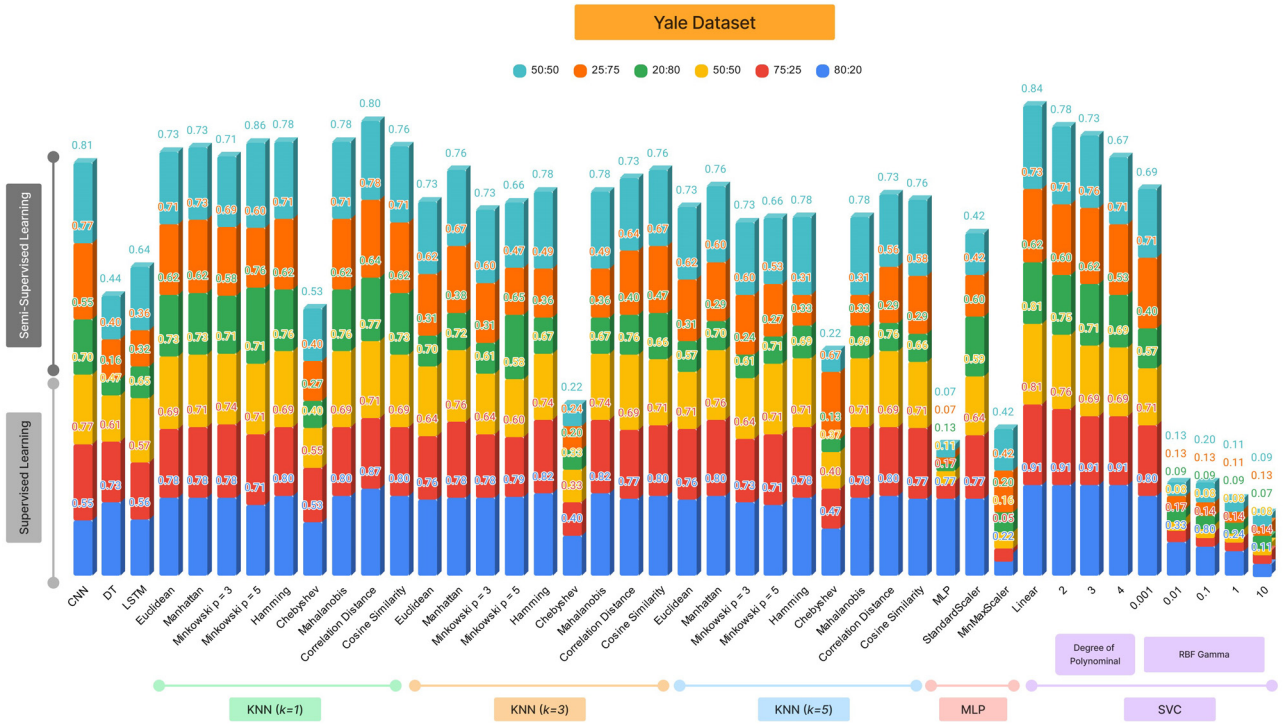


Figure 5: Visualization of accuracy performance on Yale dataset.

emerge as two algorithms with considerable significance in both supervised and semi-supervised learning analyses.

3.2 Limitations

This research investigates FER using multiple datasets, specifically the JAFFE, Georgia tech, and Yale datasets, to evaluate a comprehensive assessment of algorithmic performance across diverse scenarios with distinct data splits for both supervised and semi-supervised learning. The datasets are partitioned to create different training and testing ratios for each learning type. For supervised learning, the training-to-testing splits are set at 80:20, 75:25, and 50:50, while semi-supervised learning scenarios apply 20:80, 25:75, and 50:50 ratios. These splits reflect

Table 23: Friedman and Nemenyi post hoc test result

Scenario	Metric		Statistic	*p*-value	Significant Pairwise comparison models
Supervised learning	CV	4	292.090217	1.287475×10^{-39}	272
		5	293.421785	7.224596×10^{-40}	274
	Accuracy		253.620640	1.894047×10^{-32}	192
	K-L		117.661067	2.479582×10^{-09}	40
	I–J		242.619677	1.968756×10^{-30}	182
	*F*1-score		256.289117	6.106232×10^{-33}	196
	Cosine		208.886027	2.316541×10^{-24}	136
	ROC-AUC		201.071482	5.528787×10^{-23}	130
Semi-supervised learning	CV	4	277.750176	6.320082×10^{-37}	236
		5	265.451821	1.233029×10^{-34}	228
	Accuracy		275.992608	1.346039×10^{-36}	212
	K-L		170.975094	8.375519×10^{-18}	78
	I–J		276.032267	1.323284×10^{-36}	208
	*F*1-score		275.444938	1.703337×10^{-36}	202
	Cosine		145.266994	1.405621×10^{-13}	48
	ROC-AUC		210.804752	1.058638×10^{-24}	140

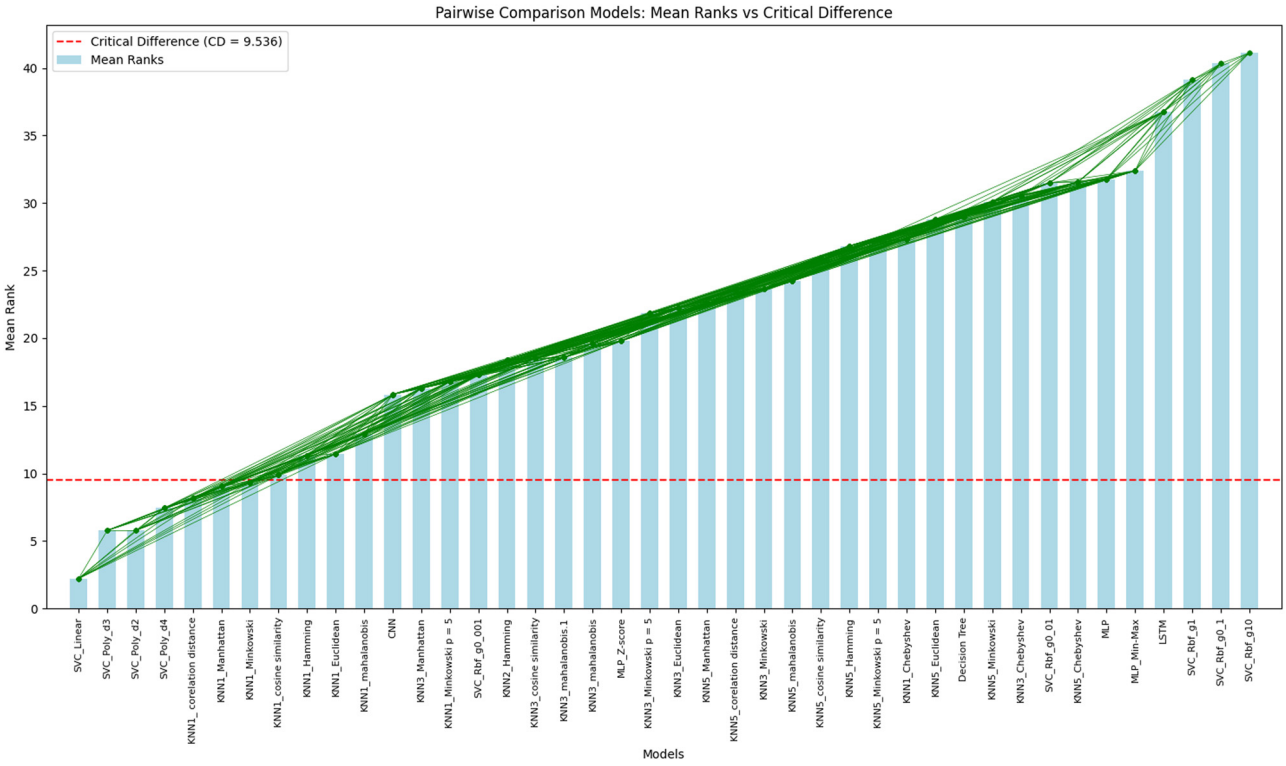


Figure 6: Pairwise comparison models of supervised learning: mean ranks vs critical difference.

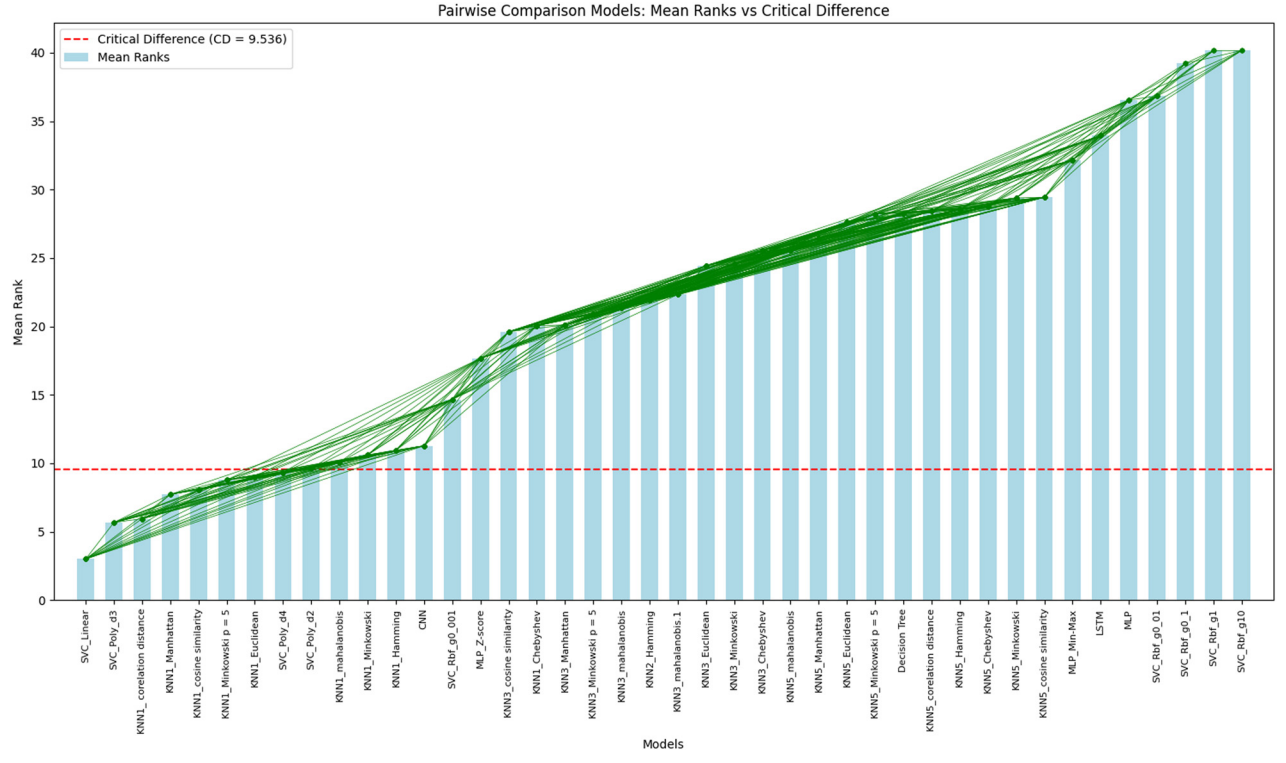


Figure 7: Pairwise comparison models of semi-supervised learning: mean ranks vs critical difference.

typical preferences in supervised learning, where the training dataset is typically equal to or larger than the testing dataset, as opposed to semi-supervised learning, where the training set is often smaller. This setup results in nine distinct scenarios for each learning approach, establishing a robust framework for model evaluation. Six classification algorithms with various parameter settings are applied across these scenarios, yielding a total of 42 models for each learning type. To rigorously assess performance, diverse evaluation metrics are employed: accuracy with 4 and 5-fold CV for training validation, and multiple metrics for testing evaluation, including accuracy, K-L divergence, I–J, *F*1-score, cosine similarity, and ROC-AUC. Statistical tests validate these findings, underscoring the superior performance of SVC with linear and polynomial kernels, which emerges as the most effective algorithm across both supervised and semi-supervised learning tasks within this facial recognition context.

While this study offers a comprehensive analysis, several limitations are acknowledged, highlighting avenues for future research. No specific feature extraction techniques were employed, which may limit the study's ability to capture critical features unique to facial expressions. Future studies could incorporate specialized feature extraction methods to enhance model sensitivity and performance. In addition, ensemble methods were not implemented, which could otherwise improve classification accuracy through model combination. Future research aims to address this gap by investigating the impact of ensemble methods on performance outcomes. The deep learning architectures utilized in this study – CNN, LSTM, and MLP – were relatively simple, which may account for the modest performance variation observed across deep learning models. Incorporating more complex architectures in subsequent research could provide greater insight into the effectiveness of advanced deep learning approaches in FER. In addition, with a substantial number of pairwise combinations – totaling 1,764 pairwise models – this study does not delve into the detailed significance of each model pair. Instead, the findings are limited to highlighting the most significant models in both supervised and semi-supervised analyses, specifically the linear SVC and polynomial SVC, followed by the remaining 40 models.

4 Conclusion

This study aims to evaluate various algorithms in supervised and semi-supervised learning applied to three multiclass facial image datasets: JAFFE, Georgia tech, and Yale. The datasets were divided into proportions of 80:20, 75:25, and 50:50 for supervised learning, and ratios of 20:80, 25:75, and 50:50 for semi-supervised learning. Evaluated algorithms include decision tree, K-NN, SVC, MLP, LSTM, and CNN, each with varying parameters. The research findings indicate that the model's performance depends on the dataset partitioning strategy and the choice of algorithms. SVC, especially linear SVC, and K-NN consistently yield favorable results in both supervised and semi-supervised learning, predominantly on the Georgia tech dataset. The integration of semi-supervised learning enhances accuracy, particularly noticeable on the Georgia tech dataset, where combining labeled and unlabeled data significantly improves accuracy, especially when using K-NN and linear SVC. Despite some algorithms, such as decision trees and CNN, showing suboptimal performance on certain datasets, this study provides comprehensive insights into the effectiveness of various models in limited-label learning scenarios. These findings are crucial for advancing the development of adaptive and robust facial recognition systems, especially when dealing with datasets characterized by diverse variations and complexities.

Funding information: This research was fully funded by Universitas Muslim Indonesia.

Author contributions: The authors have accepted responsibility for the entire content of this manuscript and approved its submission.

Conflict of interest: The authors declare that there is no conflict of interest regarding the publication of this article.

Data availability statement: The data used are included within the article.

References

[1] M. Sajjad, F. U. M. Ullah, M. Ullah, G. Christodoulou, F. A. Cheikh, M. Hijji, et al., "A comprehensive survey on deep facial expression recognition: challenges, applications, and future guidelines," *Alexandr. Eng. J.* vol. 68, pp. 817–840, 2023.

[2] S. Li and W. Deng, "Deep facial expression recognition: A survey," *IEEE Trans. Affective Comput.*, vol. 13, no. 3, pp. 1195–1215, 2020.

[3] S. K. Khangura and S. Akın, "Online available bandwidth estimation using multiclass supervised learning techniques," *Comput. Commun.*, vol. 170, no. November 2020, pp. 177–189, 2021.

[4] K. Wenger, K. Tirdad, A. D. Cruz, A. Mari, M. Basheer, C. Kuk, et al., "A semi-supervised learning approach for bladder cancer grading," *Machine Learn. Appl.* vol. 9, p. 100347, 2022.

[5] M. Murugappan, A. Mutawa, S. Sruthi, A. Hassouneh, A. Abdulsalam, S. Jerritta, et al., "Facial expression classification using KNN and decision tree classifiers," in *2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP)*, IEEE, 2020, pp. 1–6.
[6] J. Tanha, M. Van Someren, and H. Afsarmanesh, "Semi-supervised self-training for decision tree classifiers," *Int. J. Machine Learn. Cybernet.*, vol. 8, pp. 355–370, 2017.
[7] Q. Zhou, Y. Qi, H. Tang, and P. Wu, "Machine learning-based processing of unbalanced data sets for computer algorithms," *Open Comput. Sci.*, vol. 13, no. 1, p. 20220273, 2023.
[8] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, et al., "Effects of distance measure choice on k-nearest neighbor classifier performance: a review," *Big Data*, vol. 7, no. 4, pp. 221–248, 2019.
[9] A. N. Handayani, D. F. Laistulloh, I. A. E. Zaeni, M. Efendi, R. A. Asmara, and O. Fukuda, "Motion detection for children with cerebral palsy using k-nearest neighbor," in: *2022 International Conference on Electrical and Information Technology (IEIT)*, IEEE, 2022, pp. 65–68.
[10] J. S. Lee, H. M. Kim, S. I. Kim, and H. M. Lee, "Evaluation of structural integrity of railway bridge using acceleration data and semi-supervised learning approach," *Eng. Struct.*, vol. 239, p. 112330, 2021.
[11] W. Qian, F. Lauri, and F. Gechter, "Supervised and semi-supervised deep probabilistic models for indoor positioning problems," *Neurocomputing*, vol. 435, pp. 228–238, 2021.
[12] M. Abidin, M. Abidin, I. Nurtanio, and A. Achmad, "Deepfake detection in videos using long short-term memory and cnn resnext," *ILKOM Jurnal Ilmiah*, vol. 14, no. 3, pp. 178–185, 2022.
[13] T.-H. S. Li, P.-H. Kuo, T.-N. Tsai, and P.-C. Luan, "CNN and LSTM based facial expression analysis model for a humanoid robot," *IEEE Access*, vol. 7, pp. 93998–94011, 2019.
[14] H. Darwis, Z. Ali, Y. Salim, and P. L. L. Belluano, "Max feature map cnn with support vector guided softmax for face recognition," *JOIV: Int. J. Inform. Visual.*, vol. 7, no. 3, pp. 959–966, 2023.
[15] P. Purnawansyah, A. P. Wibawa, T. Widyaningtyas, H. Haviluddin, C. E. Hasihi, M. F. Teng, et al., "Comparative study of herbal leaves classification using hybrid of GLCM-SVM and GLCM-CNN," *ILKOM Jurnal Ilmiah*, vol. 15, no. 2, pp. 382–389, 2023.
[16] Y. Dong, K. Chen, Y. Peng, and Z. Ma, "Comparative study on supervised versus semi-supervised machine learning for anomaly detection of in-vehicle can network," in: *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, 2022, pp. 2914–2919.
[17] M. Bansal, A. Goyal, and A. Choudhary, "A comparative analysis of k-nearest neighbor, genetic, support vector machine, decision tree, and long short term memory algorithms in machine learning," *Decision Analyt. J.*, vol. 3, p. 100071, 2022.
[18] R. Guo, Y. Peng, W. Kong, and F. Li, "A semi-supervised label distribution learning model with label correlations and data manifold exploration," *J. King Saud Univ.-Comput. Inform. Sci.*, vol. 34, no. 10, pp. 10094–10108, 2022.
[19] G. Sun, Y. Wen, and Y. Li, "Instance segmentation using semi-supervised learning for fire recognition," *Heliyon*, vol. 8, no. 12, 2022.
[20] X. Yang, Z. Song, I. King, and Z. Xu, "A survey on deep semi-supervised learning," *IEEE Trans. Knowl. Data Eng.*, vol. 35, no. 9, pp. 8934–8954, 2022.
[21] I. Spiero, E. Schuit, O. Wijers, F. Hoebers, J. Langendijk, and A. Leeuwenberg, "Comparing supervised and semi-supervised machine learning approaches in ntcp modeling to predict complications in head and neck cancer patients," *Clin. Transl. Radiat. Oncol.* vol. 43, p. 100677, 2023.
[22] P. Y. Ristanti, A. P. Wibawa, and U. Pujianto, "Cosine similarity for title and abstract of economic journal classification," in: *2019 5th International Conference on Science in Information Technology (ICSITech)*, IEEE, 2019, pp. 123–127.
[23] G. K. Menezes, G. Astolfi, J. A. C. Martins, E. C. Tetila, A. da Silva Oliveira Junior, D. N. Gonçalves, et al., "Pseudo-label semi-supervised learning for soybean monitoring," *Smart Agricult. Tech.*, vol. 4, p. 100216, 2023.
[24] L.-Z. Guo, Z.-Y. Zhang, Y. Jiang, Y.-F. Li, and Z.-H. Zhou, "Safe deep semi-supervised learning for unseen-class unlabeled data," in: *International Conference on Machine Learning*, PMLR, 2020, pp. 3897–3906.
[25] S. Dabiri, C.-T. Lu, K. Heaslip, and C. K. Reddy, "Semi-supervised deep learning approach for transportation mode identification using gps trajectory data," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 1010–1023, 2019.
[26] J. Anil and L. Padma Suresh, "A novel fast hybrid face recognition approach using convolutional kernel extreme learning machine with hog feature extractor," *Measurement: Sensors*, vol. 30, p. 100907, 2023.
[27] A. Sultan, W. Makram, M. Kayed, and A. A. Ali, "Sign language identification and recognition: A comparative study," *Open Comput. Sci.*, vol. 12, no. 1, pp. 191–210, 2022.
[28] V. B. S. Prasath, H. A. A. Alfeilat, A. B. A. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, et al., *Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier - A Review*, pp. 1–39, 2017. arXiv preprint arXiv:1708.04321.
[29] A. M. Qamar, *Generalized cosine and similarity metrics: a supervised learning approach based on nearest neighbors*, PhD thesis, Université de Grenoble, 2010.
[30] R. Febrian, B. M. Halim, M. Christina, D. Ramdhan, and A. Chowanda, "Facial expression recognition using bidirectional LSTM-CNN," *Proc. Comput. Sci.*, vol. 216, pp. 39–47, 2023.
[31] M. Devgan, G. Malik, and D. K. Sharma, *Semi-supervised learning*, Springer, 2020.
[32] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in: *Proceedings Third EEE International Conference on Automatic Face and Gesture Recognition*, IEEE, 1998, pp. 200–205.
[33] M. J. Lyons, *Excavating ai re-excavated: Debunking a fallacious account of the jaffe dataset*, 2021, arXiv: http://arXiv.org/abs/arXiv:2107.13998.
[34] P. Purnawansyah, A. Adnan, H. Darwis, A. Wibawa, T. Widyaningtyas, and H. Haviluddin, "Ensemble semi-supervised learning in facial expression recognition," *Int. J. Adv. Intel. Inform.*, vol. 11, no. 1, pp. 1–24, 2025.
[35] S. I. Lestariningati, A. B. Suksmono, I. J. M. Edward, and K. Usman, "Group class residual 1-minimization on random projection sparse representation classifier for face recognition," *Electronics (Switzerland)*, vol. 11, no. 17, pp. 1–12, 2022.
[36] H. Hassan, Z. Ren, C. Zhou, M. A. Khan, Y. Pan, J. Zhao, and B. Huang, "Supervised and weakly supervised deep learning models for COVID-19 CT diagnosis: A systematic review," *Comput. Meth. Programs Biomed.*, vol. 218, p. 106731, 2022.
[37] S. S. Sawant and M. Prabukumar, "A review on graph-based semi-supervised learning methods for hyperspectral image classification," *Egypt. J. Remote Sens. Space Sci.*, vol. 23, no. 2, pp. 243–248, 2020.

[38] Y. Ouali, C. Hudelot, and M. Tami, *An overview of deep semi-supervised learning*, 2020, arXiv: http://arXiv.org/abs/arXiv:2006.05278.
[39] Z. Huang, L. Shen, J. Yu, B. Han, and T. Liu, "Flatmatch: Bridging labeled data and unlabeled data with cross-sharpness for semi-supervised learning," *Adv. Neural Inform. Proces. Syst.*, vol. 36, 2024, pp. 18474–18494.
[40] X. Liu, D. Zachariah, J. Wågberg, and T. B. Schön, *Reliable semi-supervised learning when labels are missing at random*, 2018, arXiv: http://arXiv.org/abs/arXiv:1811.10947.
[41] V. L. S. Lee, K. H. Gan, T. P. Tan, and R. Abdullah, "Semi-supervised learning for sentiment classification using small number of labeled data," *Proc. Comput. Sci.*, vol. 161, pp. 577–584, 2019.
[42] W. Sinhashthita and K. Jearanaitanakij, "Improving knn algorithm based on weighted attributes by pearson correlation coefficient and pso fine tuning," in: *2020-5th International Conference on Information Technology (InCIT)*, IEEE, 2020, pp. 27–32.
[43] X. Li and C. Xiang, "Correlation-based k-nearest neighbor algorithm," in: *2012 IEEE International Conference on Computer Science and Automation Engineering*, IEEE, 2012, pp. 185–187.
[44] I. Gunawan, T. Widyaningtyas, A. P. Wibawa, D. Darusalam, A. Pranolo, et al., "The performance of correlation-based support vector machine in illiteracy dataset," in: *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, IEEE, 2018, pp. 96–99.
[45] M. Muniswamaiah, T. Agerwala, and C. C. Tappert, *Applications of binary similarity and distance measures*, 2023. arXiv: http://arXiv.org/abs/arXiv:2307.00411.
[46] L. Zheng, W. Lu, and Q. Zhou, "Weather image-based short-term dense wind speed forecast with a convlstm-lstm deep learning model," *Build. Environ.*, vol. 239, p. 110446, 2023.
[47] A. Pranolo, Y. Mao, A. P. Wibawa, A. B. P. Utama, and F. A. Dwiyanto, "Robust lstm with tuned-pso and bifold-attention mechanism for analyzing multivariate time-series," *IEEE Access*, vol. 10, pp. 78423–78434, 2022.
[48] A. W. Saputra, A. P. Wibawa, U. Pujianto, A. B. P. Utama, and A. Nafalski, "LSTM-based multivariate time-series analysis: A case of journal visitors forecasting," *ILKOM Jurnal Ilmiah*, vol. 14, no. 1, pp. 57–62, 2022.
[49] H. Darwis, R. Puspitasari, Purnawansyah, W. Astuti, D. Atmajaya, and M. Hasnawi, "A deep learning approach for improving waste classification accuracy with resnet50 feature extraction," in: *2025 19th International Conference on Ubiquitous Information Management and Communication (IMCOM)*, 2025, pp. 1–8.
[50] S. Minaee, M. Minaei, and A. Abdolrashidi, "Deep-emotion: Facial expression recognition using attentional convolutional network," *Sensors*, vol. 21, no. 9, p. 3046, 2021.
[51] S. Ji, Z. Zhang, S. Ying, L. Wang, X. Zhao, and Y. Gao, "Kullback-Leibler divergence metric learning," *IEEE Trans. Cybernet.*, vol. 52, no. 4, pp. 2047–2058, 2022.
[52] M. Hossin, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Mining Knowl. Manag. Process*, vol. 5, pp. 01–11, 2015.
[53] V. Labatut and H. Cherifi, *Accuracy measures for the comparison of classifiers*, 2012, arXiv: http://arXiv.org/abs/arXiv:1207.3790.
[54] S. Szabó, I. J. Holb, V. É. Abriha-Molnár, G. Szatmári, S. K. Singh, and D. Abriha, "Classification assessment tool: a program to measure the uncertainty of classification models in terms of class-level metrics," *Appl. Soft Comput.*, vol. 55, p. 111468, 2024.
[55] F. Movahedi, R. Padman, and J. F. Antaki, "Limitations of receiver operating characteristic curve on imbalanced data: assist device mortality risk scores," *J. Thoracic Cardiovascular Surgery*, vol. 165, no. 4, pp. 1433–1442, 2023.
[56] N. Al-Azzam and I. Shatnawi, "Comparing supervised and semi-supervised machine learning models on diagnosing breast cancer," *Ann. Med. Surgery*, vol. 62, pp. 53–64, 2021.
[57] L. Cances, E. Labbé, and T. Pellegrini, "Comparison of semi-supervised deep learning algorithms for audio classification," *EURASIP J. Audio Speech Music Proces.*, vol. 2022, no. 1, p. 23, 2022.
[58] R. Ghorbani and R. Ghousi, "Comparing different resampling methods in predicting students' performance using machine learning techniques," *IEEE Access*, vol. 8, pp. 67899–67911, 2020.