UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA DE TELEINFORMÁTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE TELEINFORMÁTICA

MESTRADO ACADÊMICO EM ENGENHARIA DE TELEINFORMÁTICA

MARÍLIA MAGALHÃES MAIA

DISCRIMINANT INDEPENDENT VECTOR ANALYSIS

FORTALEZA

2026

MARÍLIA MAGALHÃES MAIA

DISCRIMINANT INDEPENDENT VECTOR ANALYSIS

Dissertação apresentada ao Curso de Mestrado Acadêmico em Engenharia de Teleinformática do Programa de Pós-Graduação em Engenharia de Teleinformática do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Engenharia. Área de Concentração: Sinais e Sistemas.

Orientador: Prof. Dr. Charles Casimiro Cavalcante.
Coorientador: Prof. Dr. Zois Boukouvalas.

FORTALEZA

2026

À minha família amada, ao meu noivo querido, aos meus bons amigos e aos professores incríveis que me apoiaram, incentivaram e acreditaram em mim.

AGRADECIMENTOS

Primeiramente, agradeço a Deus por me conceder esta oportunidade, por me fortalecer no caminho e iluminar cada passo dado ao longo desta jornada.

Agradeço aos meus pais, Marcos e Enilda, por me permitirem dedicar-me a esta etapa da minha vida, sendo sempre meu suporte, minha base e minha fonte diária de fé e oração. Às minhas irmãs, Mariana e Marcela, por me ouvirem nos momentos difíceis, com paciência, carinho e incentivo constante. Sem vocês, este caminho teria sido muito mais pesado.

Ao meu noivo, Luiz Emanoel, agradeço profundamente pelo apoio, pela compreensão, pela força e por estar ao meu lado em cada desafio. Enfrentamos juntos, e juntos também celebramos a conquista de nos tornarmos Mestres.

Aos professores Dr. Charles e Dr. Zois, deixo registrado meu sincero agradecimento pelo apoio, pela paciência e pelos ensinamentos valiosos. Minha eterna gratidão. Agradeço também aos colegas do SPIRAL, pelos momentos de descontração, pelas conversas animadoras e, especialmente, pelas dicas milagrosas que me ajudaram a superar muitos obstáculos.

Registro ainda meu agradecimento aos professores do DETI e do GTel pelos ensinamentos transmitidos, e à UFC, instituição honrosa que me ofereceu toda a estrutura necessária para chegar até aqui.

O presente trabalho foi realizado com apoio da Fundação Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP).

"A mente que se abre a uma nova ideia jamais voltará ao seu tamanho original." (Albert Einstein)

RESUMO

Com o avanço tecnológico e o crescimento acelerado da produção de dados, métodos de Separação Cega de Fontes (BSS) têm ganhado destaque devido à sua ampla aplicabilidade em diferentes áreas. Em cenários onde múltiplas modalidades de dados são misturadas e o objetivo é recuperar informações originais, técnicas capazes de explorar relações entre fontes tornam-se essenciais, especialmente em tarefas de representação ou classificação. O método tradicional para esse tipo de dados é a Análise Vetorial Independente (IVA); entretanto, sua natureza não discriminativa limita seu desempenho quando a separação de fontes está associada a objetivos classificatórios. Neste contexto, este trabalho apresenta o Discriminant Independent Vector Analysis (DIVA), uma extensão supervisionada do IVA construída pela incorporação do critério Discriminativo de Fisher (FLD) ao modelo IVA. O método resultante busca estimar fontes independentes que, além de satisfazerem independência estatística, maximizam a separabilidade entre classes, tornando-o adequado para problemas de classificação binária. O modelo proposto foi implementado a partir do IVA-G, amplamente consolidado na literatura. Para avaliar o desempenho do DIVA-G, utilizou-se um classificador Support Vector Machines (SVM) em abordagem semi-supervisionada, e o F1-score como métrica principal. Experimentos com dados sintéticos permitiram identificar características estatísticas que favorecem o método, demonstrando desempenho consistente e superior aos demais algoritmos derivados do IVA analisados. Em seguida, os conjuntos reais NSL-KDD e MediaEval2016 foram utilizados para investigar o comportamento do método em cenários complexos e ruidosos. Os resultados obtidos evidenciam desempenho satisfatório, estabilidade e confiabilidade, indicando que o DIVA-G é uma abordagem promissora para separação discriminativa de fontes e merece investigação futura mais aprofundada.

Palavras-chave: análise de vetores independentes; aprendizado multimodal; extração de características discriminativas; aprendizado de representação.

ABSTRACT

With the rapid advancement of technology and the accelerated growth in data production, Blind Source Separation (BSS) methods have gained increasing relevance due to their broad applicability across diverse domains. In scenarios where multiple data modalities are mixed and the objective is to recover the original underlying sources, techniques capable of exploiting relationships among them become essential, particularly in representation and classification tasks. The traditional method for handling such multimodal data is Independent Vector Analysis (IVA); however, its non-discriminative nature limits its performance when source separation is directly linked to classification objectives. In this context, this work introduces Discriminant Independent Vector Analysis (DIVA), a supervised extension of IVA constructed through the incorporation of Fisher's Linear Discriminant (FLD) criterion into the IVA framework. The resulting method aims to estimate independent sources that, in addition to satisfying statistical independence, maximize class separability, making it particularly suitable for binary classification problems. The proposed model was implemented based on IVA-G, a widely established formulation in the literature. To evaluate the performance of DIVA-G, a Support Vector Machine (SVM) classifier was employed in a semi-supervised setting, using the F1-score as the primary evaluation metric. Experiments with synthetic datasets enabled the identification of statistical characteristics that favor the method, demonstrating consistent and superior performance compared with other IVA-derived algorithms. Subsequently, the real-world datasets NSL-KDD and MediaEval2016 were used to investigate the behavior of the method in complex and noisy scenarios. The results indicate satisfactory performance, stability, and reliability, suggesting that DIVA-G is a promising approach for discriminative source separation and warrants further, more in-depth investigation.

Keywords: independent vector analysis; multimodal learning; discriminative feature extraction; representation learning.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUCTION

The acquisition of multiple signals related to the same phenomenon has become increasingly common, as each signal captures complementary aspects of the underlying process from different perspectives. When these signals are observed as mixed data and the objective is to recover the original underlying information, the task can be formulated as a Joint Blind Source Separation (JBSS) problem (Anderson et al., 2012; Kim et al., 2006). Such problems arise in diverse fields, including brain activity analysis, representation learning, and misinformation detection, where multiple correlated datasets must be jointly analyzed to uncover their latent sources.

Independent Vector Analysis (IVA) is a statistical method within the JBSS framework that enables the joint separation of multiple datasets by exploiting statistical dependencies among corresponding sources (Kim et al., 2006; Luo, 2023). Although highly effective for modeling multimodal and multiset data, conventional IVA algorithms are inherently unsupervised. Consequently, they treat all components uniformly, without leveraging discriminative information that could guide the separation process toward directions in the feature space that maximize inter-class separability while minimizing intra-class variability. This limitation restricts the applicability of IVA to classification-oriented tasks, where the goal extends beyond statistical independence to include the extraction of representations that are also discriminatively meaningful.

To overcome this limitation, the present work introduces the Discriminant Independent Vector Analysis (DIVA), which extends the traditional IVA cost function through the inclusion of a discriminative term. This term follows the same intention as the models proposed by Dhir and Lee (2011) and Mollaee and Moattar (2016), in which discriminative criteria are incorporated into ICA. The central idea is to integrate Fisher's Linear Discriminant (FLD) into the IVA optimization framework, enabling the simultaneous optimization of both statistical independence and class separability. By coupling the IVA formulation with a multivariate extension of FLD, the proposed model generalizes Fisher's original criterion—traditionally defined for univariate projections—to operate effectively across multiple modalities or feature spaces.

The multivariate FLD term integrated into DIVA is analytically derived as a differentiable function with respect to the demixing matrices, which allows its seamless incorporation into the IVA-G optimization scheme. The resulting model introduces a regularization parameter that balances the trade-off between independence and discriminative power, effectively bridging the gap between unsupervised source separation and supervised representation learning. Th-

rough this formulation, DIVA-G establishes a principled framework for extracting discriminative multimodal representations, offering enhanced robustness, interpretability, and performance in classification tasks involving complex and correlated datasets.

1.1 Contribution

The main contributions of this dissertation are organized as follows:

(a) Chapter 3 – Fisher's Linear Discriminant

Fisher's Linear Discriminant (FLD) is a classical discriminative technique that enhances the separation between two classes by maximizing the distance between their means while minimizing the variance within each class. It is also commonly referred to as Linear Discriminant Analysis (LDA). In this work, FLD is applied to multimodal data in conjunction with Independent Vector Analysis (IVA). To enable this integration, it was necessary to extend the traditional FLD formulation so that it could operate effectively on multivariate data.

The specific contributions of this chapter include:

- Proposing a new formulation of LDA suitable for multivariate datasets;
- Presenting an alternative derivation of the FLD criterion that explicitly depends on the means and variances of the mixed data.

(b) Chapter 4 – Discriminant Independent Vector Analysis

Motivated by the goal of enhancing IVA's efficiency in class identification tasks, a multimodal version of Fisher's Linear Discriminant was incorporated into the IVA framework, resulting in the development of the *Discriminant Independent Vector Analysis* (DIVA). The DIVA algorithm is formulated as a supervised machine learning method that jointly optimizes statistical independence and class separability.

The main contributions of this chapter are:

- Developing a novel IVA-based algorithm, termed *Discriminant Independent Vector Analysis (DIVA)*;
- Present the equations that describe DIVA.;
- Demonstrating the superior performance of DIVA in both source separation and classification tasks, using simulated and real-world datasets, when compared to widely used IVA algorithms.

1.2 Overview

Following the presentation of the context, motivation, and objectives in the introductory chapter, Chapter 2 reviews the foundations of the underlying method, including the equations and concepts necessary for understanding the proposed approach. Chapter 3 introduces the Fisher Discriminant, which serves as the discriminative component in the construction of the new method, and presents its formulation for multimodal data. In Chapter 4, the development of the Discriminant Independent Vector Analysis (DIVA) is detailed from both conceptual and mathematical perspectives. Chapter 5 describes the experiments conducted, covering the datasets used, the evaluation model, and the results obtained. Finally, Chapter 6 presents the concluding remarks and outlines directions for future work.

2 INDEPENDENT VECTOR ANALYSIS

In this chapter, a concise and objective review of IVA is presented, providing the necessary context and introducing the main concepts required for the development of this work, as described in Section 2.1. Subsequently, the cost function and its key mathematical formulations are discussed, forming the foundation of the algorithm presented in Section 2.3. Finally, it is acknowledged that, like any method, IVA has inherent limitations, which are addressed in Section 2.4.

2.1 IVA

The *Independent Vector Analysis* (IVA) was developed to extend the *Independent Component Analysis* (ICA) for application to correlated multimodal datasets. ICA is effective when a single dataset is considered, assuming statistical independence among the sources. However, when different channels measure the same phenomenon, multiple related datasets are obtained, and it was for such scenarios that IVA was developed (Kim et al., 2006; Bouveresse and Rutledge, 2016).

Intuitively, IVA seeks to "unmix" the sources across multiple sets of observations, exploiting both the independence between vectors and the internal dependencies within each multivariate source. IVA is a statistical technique belonging to the field of *Blind Source Separation* (BSS), whose goal is to recover unknown original signals, referred to as latent sources, from mixed observations without prior knowledge of the mixing process (Duarte et al., 2006; Anderson et al., 2012).

In its simplest formulation, the linear mixing model is represented as:

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}, \tag{2.1}$$

where $\boldsymbol{x}$ is the observation vector, $\boldsymbol{A}$ is the mixing matrix, and $\boldsymbol{s}$ represents the matrix that contains the source signals, this being the characteristic function of the ICA. In the context of multiple datasets, this model extends to:

$$\boldsymbol{s}^{[k]} = \boldsymbol{A}^{[k]}\boldsymbol{s}^{[k]}, \quad k = 1, 2, \ldots, K, \tag{2.2}$$

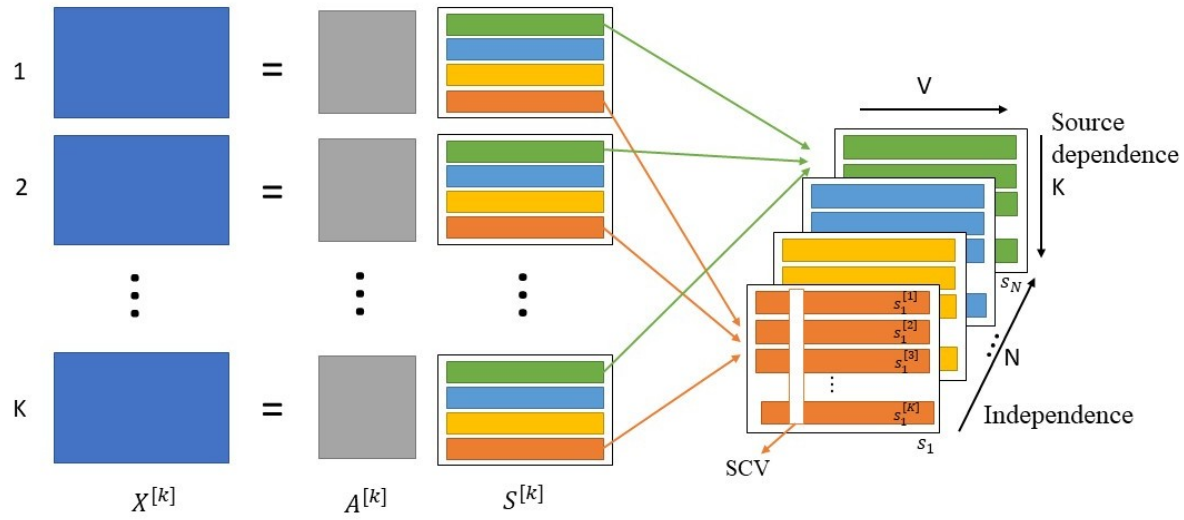
where each dataset k has its own mixing matrix $\boldsymbol{A}^{[k]}$, but the latent sources $\boldsymbol{s}^{[k]} = [\mathbf{s}_1^{[k]}, ..., \mathbf{s}_N^{[k]}]^T$ share a common statistical structure (Anderson et al., 2012; Jouan-Rimbaud et al.)

In the IVA model, each component $\mathbf{s}_n^{[k]}$ is independent from the others within a given modality, but exhibits statistical dependence across modalities. This dependence is modeled through the *Source Component Vector* (SCV):

$$\boldsymbol{s}_n = [\mathbf{s}_n^{[1]}, \mathbf{s}_n^{[2]}, \ldots, \mathbf{s}_n^{[K]}]^T, \tag{2.3}$$

which represents the same latent factor observed under different views or sensors. Figure 1 visually illustrates the formation structure of SCVs.

Figura 1 – Graphical representation of the SCV formulation in the multimodal Blind Source Separation problem.



Source: Prepared by the author.

It follows that each source component $s_n^{[k]}$ belongs to a SCV that aggregates the same latent factor across all modalities, i.e., $\mathbf{s}_n = [s_n^{[1]}, \ldots, s_n^{[K]}]^T$. The elements within a given SCV are statistically dependent, as they correspond to different observations of the same underlying source. In contrast, SCVs associated with different indices are mutually independent. That is, $\mathbf{s}_n$ is statistically independent of $\mathbf{s}_m$ for $n \neq m$. This implies that, although components sharing the same index across modalities are dependent, they remain independent from components associated with other latent factors.

The joint analysis of multichannel data enables the capture of latent patterns shared across different measurements of the same phenomenon. This process is performed by maximizing the likelihood of the observed data and minimizing the mutual information among the

estimated vectors, thereby defining a multivariate probability density function (PDF) for the latent sources (Boukouvalas, 2017).

2.2 IVA cost function

A dataset $\boldsymbol{X} = [\mathbf{x}^{[1]}, \ldots, \mathbf{x}^{[K]}]$, composed of $K \geq 2$ subsets or modalities set $\boldsymbol{x}^{[k]} \in \mathbb{R}^{M \times V}$, where M corresponds to the number of observed features and V to the number of samples, is formed from mixtures of independent sources. Let $\boldsymbol{s}^{[k]} \in \mathbb{R}^{N \times V}$ denote the latent source matrix for the k-th dataset, and $\boldsymbol{A}^{[k]} \in \mathbb{R}^{M \times N}$ the k-th mixing matrix:

$$\boldsymbol{x}^{[k]} = \boldsymbol{A}^{[k]} \boldsymbol{s}^{[k]}, \quad k = 1, \ldots, K. \tag{2.4}$$

If $M > N$, a dimensionality reduction step is applied so that $\boldsymbol{A}^{[k]}$ becomes square, $\boldsymbol{A}^{[k]} \in \mathbb{R}^{N \times N}$.

To estimate the independent sources, a separation matrix $W^{[k]} \in \mathbb{R}^{N \times N}$ is defined such that:

$$\boldsymbol{y}^{[k]} = \boldsymbol{W}^{[k]} \boldsymbol{x}^{[k]}. \tag{2.5}$$

The objective is to estimate all separation matrices $\boldsymbol{W} = [\boldsymbol{W}^{[1]}, \ldots, \boldsymbol{W}^{[K]}]$ by maximizing the independence among the vectors y_n, which is equivalent to minimizing the IVA cost function (Boukouvalas, 2017):

$$J_{\text{IVA}}(\mathscr{W}) = \sum_{n=1}^{N} H(\boldsymbol{y}_n) - \sum_{k=1}^{K} \log |\det(\boldsymbol{W}^{[k]})| - H(\boldsymbol{x}^{[1]}, \ldots, \boldsymbol{x}^{[K]}), \tag{2.6}$$

where $H(\cdot)$ denotes the differential entropy.

2.3 IVA algorithm

The process of transforming the method into an optimization algorithm is a fundamental step, as the objective is to determine an optimal demixing matrix capable of recovering the estimated sources as closely and feasibly as possible to the original independent sources. To this end, the gradient descent method is employed, which defines an iterative rule to update the matrix $\boldsymbol{W}$, making the current value an improved version of the previous one.

Initially, to determine the form of the update rule for $\boldsymbol{W}$ and to understand its iterative behavior, it is necessary to analyze the dependence of the objective function on the separation

matrices. Thus, the gradient of the cost function with respect to each separation matrix $\boldsymbol{W}^{[k]}$ is given by:

$$\frac{\partial J_{\mathrm{IVA}}(\boldsymbol{W})}{\partial \boldsymbol{W}^{[k]}} = E\{\boldsymbol{\theta}^{[k]}(\boldsymbol{x}^{[k]})^T\} - (\boldsymbol{W}^{[k]})^{-T}, \tag{2.7}$$

where the score function vector is defined as:

$$\boldsymbol{\theta}^{[k]} = -\left[\frac{\partial \log p_{\boldsymbol{s}_1}(\boldsymbol{y}_1)}{\partial \boldsymbol{y}_1^{[k]}}, \ldots, \frac{\partial \log p_{\boldsymbol{s}_N}(\boldsymbol{y}_N)}{\partial \boldsymbol{y}_N^{[k]}}\right]. \tag{2.8}$$

The iterative update follows a gradient descent scheme:

$$(\boldsymbol{W}^{[k]})^{\mathrm{new}} \leftarrow (\boldsymbol{W}^{[k]})^{\mathrm{old}} - \gamma \frac{\partial J_{\mathrm{IVA}}(\mathscr{W})}{\partial \boldsymbol{W}^{[k]}}, \tag{2.9}$$

where γ denotes the learning rate. This algorithm is known as *IVA-G* when a multivariate Gaussian density is assumed for the SCVs (Boukouvalas, 2017; Maia et al., 2025).

By understanding the iterative process that underlies the update step of the algorithm, it becomes possible to more clearly comprehend the overall functioning of IVA-G, the code corresponding to the original IVA algorithm, which is widely recognized and well-established in the literature (Anderson et al., 2012).

Algorithm 1 IVA-G Training Procedure

Require: Multimodal data sets $\{\boldsymbol{x}^{[k]}\}_{k=1}^{K}$, learning rate γ.

Preprocess each dataset: center and whiten $\boldsymbol{x}^{[k]}$.
Initialize separation matrices $\boldsymbol{W}^{[k]} = \boldsymbol{I}$ for all k.
for iteration $= 1$ to $T_{\max}$ **do**
  for each dataset $k = 1, \ldots, K$ **do**
    Compute estimated sources: $\boldsymbol{y}^{[k]} = \boldsymbol{W}^{[k]}\boldsymbol{x}^{[k]}$.
  end for
  Form source component vectors (SCVs): $\mathbf{y}_n = [y_n^{[1]}, y_n^{[2]}, \ldots, y_n^{[K]}]^\top$, for $n = 1, \ldots, N$.
  Compute the cost gradient for each dataset:

$$\nabla J_{\text{IVA}}(\mathbf{W}^{[k]}) = \mathbb{E}\left[\boldsymbol{\theta}^{[k]}(x^{[k]})^T\right] - (\mathbf{W}^{[k]})^{-\top},$$

  Update the separation matrices:

$$\mathbf{W}^{[k]} \leftarrow \mathbf{W}^{[k]} - \gamma \nabla J_{\text{IVA}}(\mathbf{W}^{[k]}).$$

  if convergence criterion satisfied **then**
    break
  end if
end for
Compute final separated sources $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$.
return Estimated independent vectors $\{\mathbf{y}^{[k]}\}_{k=1}^{K}$.

2.4 Limitation

Although IVA was developed to overcome the limitations of ICA, the method still faces several challenges. Its statistical modeling is complex, as the definition of an appropriate density function for multivariate sources is crucial for accurately capturing the true statistical behavior of the data. Furthermore, the high computational cost and sensitivity to initialization remain issues, since IVA is a non-convex problem. Small variations in initial conditions can lead the estimation process to local minima, affecting the stability and reproducibility of the results. There is also a limitation in separating highly dependent sources, as strong dependencies among different datasets can degrade separation quality (Boukouvalas, 2017; Damasceno et al., 2021).

These constraints have motivated the continuous development of semiparametric models and more robust estimation methods, such as those proposed by Boukouvalas (2017) and Damasceno et al. (2021), which explore more flexible density functions and efficient integration schemes to improve performance and stability. Nevertheless, an important limitation persists: the difficulty in discriminating between sources with very similar statistical patterns, which reduces

IVA's ability to properly distinguish components with closely correlated structures.

According to the ideas presented by Maia et al. (2024) and Anderson et al. (2012), IVA is not inherently discriminative, as it treats all components equally without favoring directions in the feature space that increase inter-class mean separation and reduce intra-class variability. In other words, IVA identifies statistically independent components, but not necessarily those that are relevant for class discrimination. IVA does not guarantee that the obtained components maximize separability between different sample classes or categories.

These limitations have motivated the development of supervised and discriminative extensions of ICA and IVA, which incorporate class separability criteria into their formulations. Among these, the *Discriminant Independent Component Analysis* (DICA) model stands out, as it integrates inter- and intra-class dispersion measures, inspired by the *Fisher Linear Discriminant* (FLD), into the component extraction process (Dhir and Lee, 2011). Such approaches represent an important conceptual advancement, as they aim to reconcile the statistical independence of sources with maximal class discrimination, thereby overcoming one of the main limitations of classical IVA in the context of supervised analysis.

2.5 Summary

In summary, this chapter presented the theoretical and mathematical foundations of Independent Vector Analysis (IVA), establishing the conceptual basis required to understand the method upon which the proposal of this dissertation is built. The modeling principles, cost function, and algorithmic mechanisms that enable the joint separation of sources across multiple modalities were reviewed, as well as the inherent limitations of the approach. This understanding is essential for moving on to the next chapter, where Fisher's Discriminant, a classical method for maximizing class separability, is introduced. When integrated into the IVA framework, this discriminative criterion serves as the foundation for the proposed method developed in this work.

3 FISHER'S DISCRIMINANT

This chapter is organized into two sections. The first presents the main concepts and mathematical formulations of Fisher's Discriminant in its traditional form. The second introduces an extension of this method for multimodal data, adapting it to the framework required for the development of the new approach proposed in this dissertation.

3.1 Fisher's Linear Discriminant

The *Fisher Linear Discriminant* (FLD), also known as *Linear Discriminant Analysis* (LDA), is a classical linear projection is a method widely used for discriminant feature extraction and dimensionality reduction in supervised classification tasks (Fisher, 1936; Lu et al., 2003; Mollaee and Moattar, 2016). Its formulation is based on finding an optimal linear subspace in which samples from different classes are maximally separated, while samples belonging to the same class exhibit minimal dispersion.

Mathematically, the FLD seeks a projection vector $\mathbf{w} \in \mathbb{R}^m$ that maximizes the ratio between the inter-class variance and the intra-class variance, known as the Fisher criterion:

$$F(\mathbf{w}) = \frac{\mathbf{w}^T \mathbf{S}_B \mathbf{w}}{\mathbf{w}^T \mathbf{S}_W \mathbf{w}} \tag{3.1}$$

where $\mathbf{S}_B$ and $\mathbf{S}_W$ denote the between-class and within-class scatter matrices, respectively, defined as:

$$\mathbf{S}_B = \frac{1}{V} \sum_{c=1}^{C} V_c (\mathbf{m}_c - \mathbf{m})(\mathbf{m}_c - \mathbf{m})^T, \tag{3.2}$$

$$\mathbf{S}_W = \frac{1}{V} \sum_{v=1}^{V} (x'_v - m_{c(v)})(x'_v - m_{c(v)})^T, \tag{3.3}$$

where $\mathbf{m}_c$ represents the mean vector of class c, C is the number of existing classes, V_c is the number of samples in class c, since V is the total number of samples, and $\mathbf{m}$ is the global mean (Dhir and Lee, 2011; Belhumeur et al., 1997)

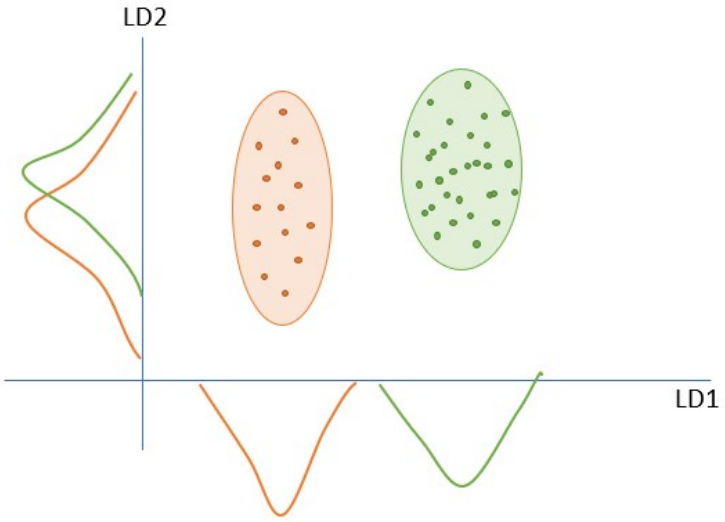
Maximizing $F(\mathbf{w})$ leads to the solution of the generalized eigenvalue problem:

$$\mathbf{S}_B \mathbf{w} = \lambda \mathbf{S}_W \mathbf{w}, \tag{3.4}$$

where the eigenvectors associated with the largest λ define the discriminant subspace, which contains the directions of maximum linear separability between classes.

Geometrically, FLD seeks the hyperplane that best separates the classes by maximizing the distance between their projected means while minimizing the intra-class variance (Daqi et al., 2014). This is achieved by determining the direction that minimizes the overlap between the Gaussian distributions that model each class. Figure 2 provides a visual representation of this behavior. In this figure, a dataset composed of two classes is shown, represented by the colors green and orange. By examining the Gaussian distributions resulting from the projections of the data onto each axis, it becomes evident that the separation between the classes is more pronounced along the LD1 direction than along LD2. Thus, the FLD indicates that projecting the data onto the LD1 direction is more advantageous for maximizing class separability. Furthermore, this direction also exhibits lower intra-class variance, suggesting that it is the most effective for clustering samples belonging to the same class.

Figura 2 – Representation of FLD behavior



Source: Prepared by the author.

However, the method presents known limitations. The main one is the *Small Sample Size* (SSS) problem, which occurs when the number of samples is smaller than the data dimensionality, making $\mathbf{S}_W$ singular and non-invertible.

To address this issue, several variants have been proposed:

– **Fisherfaces** (Belhumeur et al., 1997), which apply a PCA step prior to FLD to reduce dimensionality;

– **Direct LDA (D-LDA)** and **Fractional-step LDA (F-LDA)** (Ye, 2005), which preserve discriminant information contained in the null space of $\mathbf{S}_W\mathbf{S}_W$;

Additionally, Daqi et al. (2014) proposed the Integrated Fisher Linear Discriminants (IFLD), which introduce class-balancing mechanisms and iterative learning, improving the performance of FLD on imbalanced datasets.

An extension of the traditional FLD formulation becomes necessary when working with multimodal data. This generalization, which adapts the criterion to operate consistently across multiple modalities, is presented in the following section.

3.2 LDA multimodal

The multivariate formulation, employed in multimodal models such as the Discriminant IVA (DIVA) proposed in this work (Maia et al., 2024), generalizes the FLD to projection matrices $\mathbf{W}^{[k]} \in \mathbb{R}^{N\times N}$, applied to each data subset $\mathbf{x}^{[k]} \in \mathbb{R}^{M\times V}$.

The objective is to maximize the determinant of the ratio between the between-class and within-class scatter:

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \log F(\mathbf{W}^{[k]}) \tag{3.5}$$

$$F(\mathbf{W}^{[k]}) = \frac{|(\mathbf{W}^{[k]})^T \mathbf{S}_B^{[k]} \mathbf{W}^{[k]}|}{|(\mathbf{W}^{[k]})^T \mathbf{S}_W^{[k]} \mathbf{W}^{[k]}|} \tag{3.6}$$

The scatter matrices are defined as:

$$\mathbf{S}_B^{[k]} = \frac{1}{V}\sum_{c=1}^{C} V_c(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})^T, \tag{3.7}$$

$$\mathbf{S}_W^{[k]} = \frac{1}{V}\sum_{v=1}^{V} (x'^{[k]}_v - m^{[k]}_{c(v)})(x'^{[k]}_v - m^{[k]}_{c(v)})^T, \tag{3.8}$$

where $\mathbf{m}^{[k]}$ denotes the global mean of the samples, $\mathbf{m}_c^{[k]}$ is the mean of class c, and V_c represents the number of samples belonging to class c. $\mathbf{x}'^{[k]}_v$ is the vth column verctor of $\mathbf{x}^{[k]}$.

Here, the additive property of statistically independent features is used, where $\mathbf{w}_n^{[k]}$ is the n-th vector of the $\mathbf{W}^{[k]}$ matrix.

$$F(\mathbf{W}^{[k]}) = \prod_{n=1}^{N} F(\mathbf{w}_n^{[k]}). \tag{3.9}$$

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \sum_{n=1}^{N} \log \frac{{\mathbf{w}_n^{[k]}}^T \mathbf{S}_B \mathbf{w}_n^{[k]}}{{\mathbf{w}_n^{[k]}}^T \mathbf{S}_W \mathbf{w}_n^{[k]}}. \tag{3.10}$$

Taking each term of the fraction separately:

For the numerator, substituting the value of S_B and developing the expression, we get:

$${\mathbf{w}_n^{[k]}}^T \mathbf{S}_B \mathbf{w}_n^{[k]} = {\mathbf{w}_n^{[k]}}^T \left[\frac{1}{V} \sum_{c=1}^{C} V_(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})^T \right] \mathbf{w}_n^{[k]}. \tag{3.11}$$

Expanding further:

$$\sum_{c=1}^{C} V_c {\mathbf{w}_n^{[k]}}^T (\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]})^T \mathbf{w}_n^{[k]} = \sum_{c=1}^{C} V_c \left({\mathbf{w}_n^{[k]}}^T \mathbf{m}_c^{[k]} - {\mathbf{w}_n^{[k]}}^T \mathbf{m}^{[k]} \right) \left({\mathbf{m}_c^{[k]}}^T \mathbf{w}_n^{[k]} - {\mathbf{m}^{[k]}}^T \mathbf{w}_n^{[k]} \right). \tag{3.12}$$

Thus, the terms are defined as:

- ${\mathbf{w}_n^{[k]}}^T \mathbf{m}_c^{[k]} = \mu_{nc}^{[k]}$.
- ${\mathbf{w}_n^{[k]}}^T \mathbf{m}^{[k]} = \mu_n^{[k]}$.

Thus, the expression can be written as:

$${\mathbf{w}_n^{[k]}}^T \mathbf{S}_B \mathbf{w}_n^{[k]} = \frac{1}{V} \sum_{c=1}^{C} V_c (\mu_{nc}^{[k]} - \mu_n^{[k]})^2. \tag{3.13}$$

For the denominator, similarly,

$${\mathbf{w}_n^{[k]}}^T \mathbf{S}_W \mathbf{w}_n^{[k]} = {\mathbf{w}_n^{[k]}}^T \left[\frac{1}{V} \sum_{v=1}^{V} ({x'}_v^{[k]} - {m_{c(v)}}^{[k]})({x'}_v^{[k]} - {m_{c(v)}}^{[k]})^T \right] \mathbf{w}_n^{[k]}. \tag{3.14}$$

Expanding,

$${\mathbf{w}_n^{[k]}}^T \mathbf{S}_W \mathbf{w}_n^{[k]} = \frac{1}{V} \sum_{v=1}^{V} ({\mathbf{w}_n^{[k]}}^T {x'}_v^{[k]} - {\mathbf{w}_n^{[k]}}^T m_{c(v)}^{[k]})({x'}_v^{[k]} \mathbf{w}_n^{[k]} - {m_{c(v)}}^{[k]T} \mathbf{w}_n^{[k]}). \tag{3.15}$$

Similarly, define

- ${\mathbf{w}_n^{[k]}}^T {x'_v}^{[k]} = y_{nv}^{[k]}$
- ${\mathbf{w}_n^{[k]}}^T {m_{c(v)}}^{[k]} = \mu_{nc(v)}^{[k]}$

the resulting expression for the denominator is:

$$\mathbf{w}_n^{[k]T}\mathbf{S}_W\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{v=1}^{V}(y_{nv}^{[k]} - \mu_{nc(v)}^{[k]})^2. \tag{3.16}$$

We can bring the variance in as follows:

$$\sigma_{nc}^{[k]2} = \frac{1}{V^{[k]}}\sum_{v\in\text{Class } c}(y_{nv}^{[k]} - \mu_{nc}^{[k]})^2, \tag{3.17}$$

it follows that classes are grouped as:

$$\sum_{v=1}^{V}(y_{nv}^{[k]} - \mu_{nc(v)}^{[k]})^2 = \sum_{c=1}^{C}\sum_{v\in\text{Class } c}(y_{nv}^{[k]} - \mu_{nc}^{[k]})^2 = \sum_{c=1}^{C}V_c\sigma_{nc}^{[k]2}. \tag{3.18}$$

Thus, rewriting the denominator:

$$\mathbf{w}_n^{[k]T}\mathbf{S}_W\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{c=1}^{C}V_c\sigma_{nc}^{[k]2}. \tag{3.19}$$

The objective function can be expanded:

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \sum_{n=1}^{N}\log\frac{\sum_{c=1}^{C}V_c(\mu_{nc}^{[k]} - \mu_n^{[k]})^2}{\sum_{c'=1}^{C}V_{c'}(\sigma_{c'}^{[k]})^2}. \tag{3.20}$$

In this formulation, the numerator measures the dispersion between class means (between-class variance), while the denominator measures the dispersion within each class.

This logarithmic formulation is particularly useful in joint optimization contexts, as it allows the integration of the Fisher discriminant term into broader cost functions, such as the method proposed in this work, without compromising the algorithm's convergence.

3.3 Summary

In this chapter, the theoretical foundations of Fisher's Discriminant were reviewed, emphasizing its role as a separation method based on maximizing the distance between classes while minimizing intraclass variability. Subsequently, its extended formulation for multimodal data was presented, enabling the Fisher criterion to operate consistently across multiple modalities or simultaneous representations. This extension is particularly relevant, as it provides the structure

required to integrate the discriminant into the IVA framework, thereby reconciling statistical independence with class separability.

With this understanding established, we lay the conceptual foundation for the development of the new method proposed in this dissertation. The next chapter introduces the Discriminant Independent Vector Analysis (DIVA), detailing how the principles discussed here are incorporated into the IVA mechanism to construct a supervised method capable of maximizing both source separation and class discrimination.

4 DISCRIMINANT INDEPENDENT VECTOR ANALYSIS

4.1 DIVA

The *Discriminant Independent Vector Analysis* (DIVA) is a supervised extension of the *Independent Vector Analysis* (IVA) model, proposed with the aim of incorporating class information into the blind source separation process. While traditional IVA seeks to estimate statistically independent components from multiple correlated datasets, DIVA introduces a discriminant term into the cost function in order to maximize class separability and reduce within-class dispersion (Maia et al., 2024; Dhir and Lee, 2011).

From a conceptual perspective, DIVA combines the principle of statistical independence (inherent to IVA) with the principle of maximal class discrimination, inspired by the *Fisher Linear Discriminant* (FLD). Thus, in addition to recovering independent sources, the method favors directions in the latent space that are relevant for supervised classification tasks. These properties make DIVA particularly suitable for supervised multimodal scenarios, such as multimodal speech recognition, hyperspectral image fusion, and discriminative representation learning across multiple views (Maia et al., 2024; Boukouvalas, 2017).

On the other hand, DIVA inherits practical limitations from IVA, such as high computational cost and sensitivity to initialization, and introduces the additional challenge of selecting the regularization parameter λ, which controls the trade-off between statistical independence and class discriminability (Damasceno et al., 2021).

4.2 DIVA cost function

Consider a multimodal dataset $\mathbf{X}$, composed of K subsets, each containing V samples vectors with N feature and their corresponding class labels. Formally, we define $\mathbf{X} = \left[\mathbf{x}^{[1]}, \ldots, \mathbf{x}^{[K]}\right]$, in such a way that $\mathbf{x}^{[\mathbf{k}]^T} = \left[\mathbf{x}_1^{[k]}, \ldots, \mathbf{x}_V^{[k]}\right]^T$, where each vector $\mathbf{x}_v^{[k]} \in \mathbb{R}^{N\times 1}$ represents the N-dimensional feature vector associated with sample v in subset k. Each sample is assigned a class label C_v, forming the pair $(\mathbf{x}_v^{[k]}, C_v)$.

For each subset $\mathbf{x}^{[k]}$, let $\mathbf{x}_{(c)}^{[k]}$ denote the collection of vectors belonging to class c. It follows that $\mathbf{x}_{(c)}^{[k]} \subset \mathbf{x}^{[k]}$, and that the union of all class-specific subsets reconstructs the original dataset:

$\mathbf{x}^{[k]} = \mathbf{x}^{[k]}_{(1)} \cup \cdots \cup \mathbf{x}^{[k]}_{(C)}$.

The classes are mutually exclusive; therefore, each vector belongs to exactly one class. Let $V_c^{[k]}$ denote the number of vectors of class c in subset k, such that $V^{[k]} = V_1^{[k]} + \cdots + V_C^{[k]}$.

Having established the objective of estimating the independent sources $\mathbf{y}^{[k]}$ that generate each observation set $\mathbf{x}^{[k]}$, we seek source representations that also maximize the separability between classes. To this end, we adopt the linear mixing model $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$, where $\mathbf{W}^{[k]} \in \mathbb{R}^{N\times N}$ denotes the demixing matrix associated with modality k. By defining the concatenated matrices

$$\mathscr{W} = \left[\mathbf{W}^{[1]}, \ldots, \mathbf{W}^{[K]}\right], \qquad \mathbf{Y} = \left[\mathbf{y}^{[1]}, \ldots, \mathbf{y}^{[K]}\right],$$

we obtain the global relation $\mathbf{Y} = \mathscr{W}\mathbf{X}$, which represents the set of estimated independent sources subsequently used in the discriminative learning framework.

To achieve this objective, we combine two complementary methods: IVA, which is responsible for estimating statistically independent sources, and Fisher's criterion, which introduces a discriminatory restriction on the sources estimated by IVA to improve the separability of the class.

Recalling the discussion presented in Chapter 3, the Fisher criterion is defined as

$$F(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_B\,\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_W\,\mathbf{w}},$$

whereas the corresponding cost function adopted in this work is

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \sum_{n=1}^{N} \log\left(\frac{{\mathbf{w}_n^{[k]}}^{\mathrm{T}}\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]}}{{\mathbf{w}_n^{[k]}}^{\mathrm{T}}\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]}}\right),$$

where $\mathbf{S}_B$ is the between-class scatter matrix, dependent on the class means and the global mean of the dataset (see Eq. 3.7), and $\mathbf{S}_W$ is the within-class scatter matrix (see Eq. 3.8), which centers the observations of each class around their respective means.

When analyzing the influence of $\mathscr{W}$ on the Fisher objective, we differentiate the logarithmic cost directly. For each column vector $\mathbf{w}_n^{[k]}$ of $\mathbf{W}^{[k]}$, we have

$$\frac{\partial\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C)}{\partial\mathbf{w}_n^{[k]}} = \frac{\partial}{\partial\mathbf{w}_n^{[k]}} \sum_{n=1}^{N} \log\left(\frac{{\mathbf{w}_n^{[k]}}^{\mathrm{T}}\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]}}{{\mathbf{w}_n^{[k]}}^{\mathrm{T}}\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]}}\right). \tag{4.1}$$

Applying the derivative of the logarithm and the quotient rule for quadratic forms yields the compact expression

$$\frac{\partial \phi(\mathbf{W}^{[k]}, \mathbf{X}^{[k]}, C)}{\partial \mathbf{w}_n^{[k]}} = \frac{2\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]}}{\mathbf{w}_n^{[k]\mathrm{T}}\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]}} - \frac{2\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]}}{\mathbf{w}_n^{[k]\mathrm{T}}\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]}}. \tag{4.2}$$

To proceed, we substitute the classical definitions of $\mathbf{S}_B^{[k]}$ and $\mathbf{S}_W^{[k]}$. Let V be the total number of samples and V_c the number of samples in class c. The class means and global mean are defined as

$$\mathbf{m}_c^{[k]} = E[\mathbf{x}^{[k]} \mid c], \qquad \mathbf{m}^{[k]} = E[\mathbf{x}^{[k]}].$$

Then the between-class scatter matrix is

$$\mathbf{S}_B^{[k]} = \frac{1}{V}\sum_{c=1}^{C} V_c \left(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]}\right)\left(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]}\right)^{\mathrm{T}}. \tag{4.3}$$

Since $\mathbf{y}_n^{[k]} = \mathbf{w}_n^{[k]\mathrm{T}}\mathbf{x}^{[k]}$, we define the projected means as

$$\mu_{nc}^{[k]} = \mathbf{w}_n^{[k]\mathrm{T}}\mathbf{m}_c^{[k]}, \qquad \mu_n^{[k]} = \mathbf{w}_n^{[k]\mathrm{T}}\mathbf{m}^{[k]}.$$

Thus,

$$\mathbf{w}_n^{[k]\mathrm{T}}\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{c=1}^{C} V_c\left(\mu_{nc}^{[k]} - \mu_n^{[k]}\right)^2, \tag{4.4}$$

and

$$\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{c=1}^{C} V_c\left(\mu_{nc}^{[k]} - \mu_n^{[k]}\right)\left(\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]}\right). \tag{4.5}$$

Similarly, for the within-class scatter matrix,

$$\mathbf{S}_W^{[k]} = \frac{1}{V}\sum_{v=1}^{V}\left(\mathbf{x}_v^{[k]} - \mathbf{m}_{c(v)}^{[k]}\right)\left(\mathbf{x}_v^{[k]} - \mathbf{m}_{c(v)}^{[k]}\right)^{\mathrm{T}}, \tag{4.6}$$

leading to

$$\mathbf{w}_n^{[k]\mathrm{T}}\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{c=1}^{C}\sum_{v\in\mathscr{C}_c}\left(y_{n,v}^{[k]} - \mu_{nc}^{[k]}\right)^2, \tag{4.7}$$

and

$$\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]} = \frac{1}{V}\sum_{v=1}^{V}\left(y_{n,v}^{[k]} - \mu_{nc(v)}^{[k]}\right)\left(\mathbf{x}_v^{[k]} - \mathbf{m}_{c(v)}^{[k]}\right). \tag{4.8}$$

Equations (4.5) and (4.8) highlight why the gradient in (4.2) cannot be trivially simplified: both $\mathbf{S}_B^{[k]}\mathbf{w}_n^{[k]}$ and $\mathbf{S}_W^{[k]}\mathbf{w}_n^{[k]}$ mix quantities in the observation domain $\mathbf{X}$ (e.g., $\mathbf{m}_c^{[k]} - \mathbf{m}^{[k]}$, $\mathbf{x}_v^{[k]} - \mathbf{m}_{c(v)}^{[k]}$) with scalar terms in the projected domain $\mathbf{Y}$ (e.g., $\mu_{nc}^{[k]} - \mu_n^{[k]}$, $y_{n,v}^{[k]} - \mu_{nc(v)}^{[k]}$). Thus, an explicit gradient requires applying the chain rule to the scalar projected means and variances.

Substituting (4.4) and (4.7) into the cost yields

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \sum_{n=1}^{N} \log\left(\frac{\sum_{c=1}^{C} V_c (\mu_{nc}^{[k]} - \mu_n^{[k]})^2}{\sum_{c'=1}^{C} \sum_{v \in \mathscr{C}_{c'}} (y_{nv}^{[k]} - \mu_{nc'}^{[k]})^2} \right). \tag{4.9}$$

Using the logarithm decomposition,

$$\phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C) = \sum_{n=1}^{N} \log\left(\sum_{c=1}^{C} V_c (\mu_{nc}^{[k]} - \mu_n^{[k]})^2 \right) - \sum_{n=1}^{N} \log\left(\sum_{c'=1}^{C} \sum_{v \in \mathscr{C}_{c'}} (y_{nv}^{[k]} - \mu_{nc'}^{[k]})^2 \right). \tag{4.10}$$

Differentiating (4.10) with respect to $\mathbf{w}_n^{[k]}$ yields the practical gradient:

$$\frac{\partial \phi(\mathbf{W}^{[k]}, \mathbf{X}^{[k]}, C)}{\partial \mathbf{w}_n^{[k]}} = \sum_{c=1}^{C} \sum_{v \in \mathscr{C}_c} \left[\frac{2(\mu_{nc}^{[k]} - \mu_n^{[k]})}{\sum_{c'} V_{c'} (\mu_{nc'}^{[k]} - \mu_n^{[k]})^2} - \frac{2(y_{nv}^{[k]} - \mu_{nc}^{[k]})}{\sum_{c'} \sum_{v' \in \mathscr{C}_{c'}} (y_{nv'}^{[k]} - \mu_{nc'}^{[k]})^2} \right] \mathbf{x}_v^{[k]}. \tag{4.11}$$

Equation (4.11) is the final usable form of the gradient for the multimodal Fisher criterion: it expresses $\partial\phi/\partial\mathbf{w}_n$ as a linear combination of the samples $\mathbf{x}_v$, with weights determined by the projected class means and variances. Since DIVA is constructed by incorporating Fisher's discriminative criterion into the IVA framework, we first revisit the traditional IVA cost function in order to contextualize how this integration is carried out. The IVA cost function is given by

$$J_{\mathrm{IVA}}(\mathscr{W}) = \sum_{n=1}^{N} H(\mathbf{y}_n) - \sum_{k=1}^{K} \log|\det(\mathbf{W}^{[k]})| - H(\mathbf{x}^{[1]}, \ldots, \mathbf{x}^{[K]}),$$

whose gradient is

$$\frac{\partial J_{\mathrm{IVA}}(\mathscr{W})}{\partial \mathbf{W}^{[k]}} = E\{\theta^{[k]} (\mathbf{x}^{[k]})^T\} - (\mathbf{W}^{[k]})^{-T},$$

where

$$\theta^{[k]} = -\left[\frac{\partial \log p_{s_1}(\mathbf{y}_1)}{\partial \mathbf{y}_1^{[k]}}, \ldots, \frac{\partial \log p_{s_N}(\mathbf{y}_N)}{\partial \mathbf{y}_N^{[k]}} \right].$$

Based on this formulation, DIVA is defined by adding a multivariate discriminative term inspired by Fisher's criterion. Thus, the DIVA cost function becomes

$$J_{\mathrm{DIVA}}(\mathscr{W}) = J_{\mathrm{IVA}}(\mathscr{W}) - \lambda \sum_{k=1}^{K} \phi\left(\mathbf{W}^{[k]}, \mathbf{X}^{[k]}, C\right), \tag{4.12}$$

where $\lambda > 0$ is the regularization parameter that weights the contribution of the discriminant term.

The gradient-based optimization is obtained from the gradient of J_{DIVA} with respect to each separation matrix $\mathbf{W}^{[k]}$, given by:

$$\frac{\partial J_{\text{DIVA}}(\mathscr{W})}{\partial \mathbf{W}^{[k]}} = \frac{\partial J_{\text{IVA}}(\mathscr{W})}{\partial \mathbf{W}^{[k]}} - \lambda \frac{\partial \phi\left(\mathbf{W}^{[k]}, \mathbf{X}^{[k]}, C\right)}{\partial \mathbf{W}^{[k]}}. \tag{4.13}$$

The derivative of the discriminant term can be written as:

$$\frac{\partial J_{\text{DIVA}}(\mathscr{W})}{\partial \mathbf{W}^{[k]}} = E\{\theta^{[k]}(\mathbf{x}^{[k]})^T\} - (\mathbf{W}^{[k]})^{-T} - \lambda \sum_{c=1}^{C} \sum_{v \in \mathscr{C}_c} \left[A_{nc}^{[k]} - B_{nv}^{[k]}\right] \mathbf{x}_v^{[k]}. \tag{4.14}$$

where $\mathscr{C}_c$ denotes the index set of samples belonging to class c, and $\mathbf{x}_v^{[k]} \in \mathbb{R}^N$ is the feature vector of the v-th sample in modality k.

The terms $A_c^{[k]}$ and $B_{cv}^{[k]}$ are defined as:

$$A_{nc}^{[k]} = \frac{\left(\mu_{nc}^{[k]} - \mu_n^{[k]}\right)}{\sum_{c'=1}^{C} V_{nc'}^{[k]} \left(\mu_{nc'}^{[k]} - \mu_n^{[k]}\right)^2}, \tag{4.15}$$

$$B_{nv}^{[k]} = \frac{\left(y_{nv}^{[k]} - \mu_{nc}^{[k]}\right)}{\sum_{c'=1}^{C} V_{c'}^{[k]} \left(\sigma_{nc'}^{[k]}\right)^2}. \tag{4.16}$$

In addition to being essential for understanding the influence of the matrix $\mathbf{W}^{[\mathbf{k}]}$ on the DIVA cost function, the gradient plays a fundamental role in the automated optimization procedure used to estimate the optimal demixing matrix. Based on this formulation, it becomes possible to implement an iterative, gradient-based update scheme that ensures the convergence of the method. In the next section, we present a detailed description of the implementation of the DIVA algorithm and the code developed for this work.

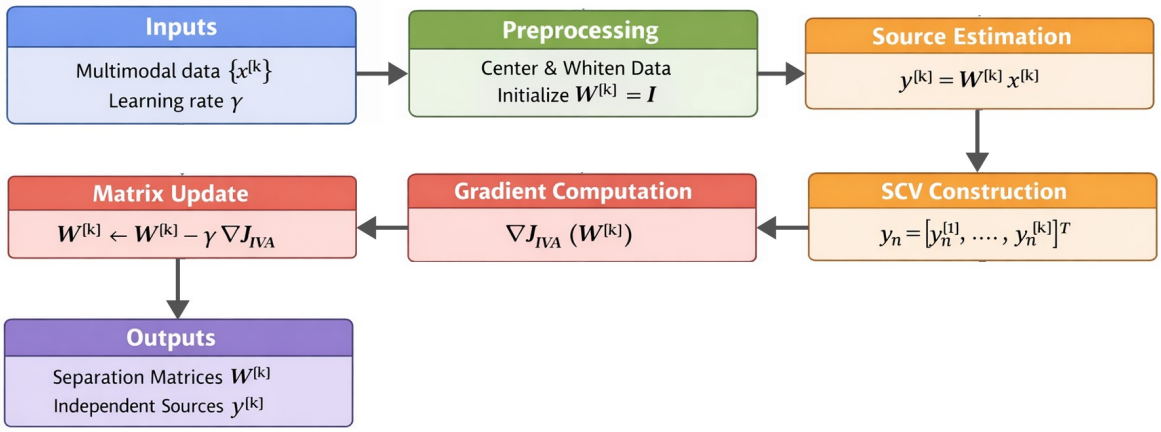
4.3 DIVA algorithm

Understanding that estimating the sources closest to the originals requires determining the optimal demixing matrix $\mathbf{W}$, we define an iterative update rule for $\mathbf{W}$ to progressively refine its estimation.

With the gradient defined, the gradient descent update rule becomes:

$$\mathbf{W}^{[k]^{new}} \leftarrow \mathbf{W}^{[k]^{old}} - \gamma \left(\frac{\partial J_{\text{IVA}}(\mathscr{W})}{\partial \mathbf{W}^{[k]}} - \lambda \frac{\partial \phi\left(\mathbf{W}^{[k]}, \mathbf{X}^{[k]}, C\right)}{\partial \mathbf{W}^{[k]}} \right), \tag{4.17}$$

Figura 3 – Block diagram of the IVA-G operation.



Source: Prepared by the author.

where $\gamma > 0$ denotes the learning rate. In practical implementations, it is recommended to apply a normalization or projection step after each iteration (e.g., row orthonormalization of $\mathbf{W}^{[k]}$) in order to improve numerical stability (Boukouvalas, 2017).

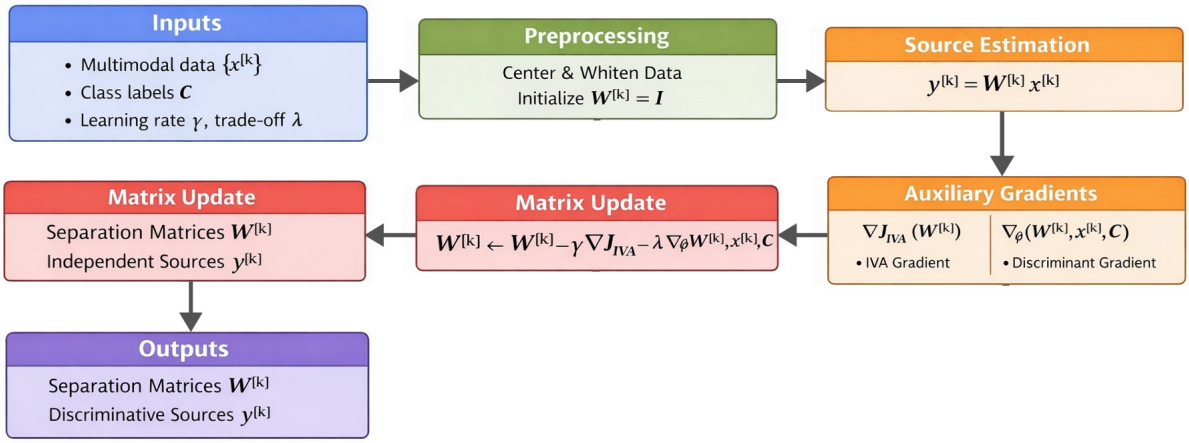
For the development of the DIVA algorithm, the implementation was constructed on top of the original IVA-G (Anderson et al., 2012; Anderson et al., 2010) framework, which has been extensively tested and widely employed as the basis for numerous extensions in the literature. Remaining faithful to its structure and fundamental properties, two auxiliary functions were incorporated: one responsible for computing the multivariate FLD cost function, as defined in Equation (4.9), and another dedicated to computing its corresponding gradient with respect to the demixing matrix, as shown in Equation (4.11).

The overall workflow of the proposed method, as well as its differences from the original IVA-G, are illustrated in Figure 3 and Figure 4. While Figure 3 presents the standard IVA-G pipeline, Figure 4 highlights the modifications introduced in DIVA-G, particularly the inclusion of discriminative information through auxiliary gradient terms and class labels.

These functions were implemented externally to the core IVA-G routine in order to modularize the code and reduce computational overhead, thereby enabling a more efficient processing pipeline. Their outputs are directly incorporated into the corresponding stages of the IVA-G algorithm—namely, the computation of the cost function and the update of the separation matrices—each weighted by the regularization parameter λ. As depicted in Figure 4, this modification directly affects the gradient computation stage, following the Equation (4.11).

With these modifications, DIVA-G preserves the same input–output structure as IVA-G, requiring a multimodal dataset with at least K=2 modalities and ensuring that all datasets

Figura 4 – Block diagram of the DIVA-G operation.



Source: Prepared by the author.

share compatible dimensions so that projection is feasible. This structural similarity can be observed by comparing the input and output blocks in both Figure 3 and Figure 4. However, in contrast to IVA-G, which is entirely unsupervised, DIVA-G explicitly requires class labels, as they are necessary for computing the discriminant terms, as indicated in the extended input stage of Figure 4.

As observed in Equation (4.14), all processing is carried out in the matrix domain. Consequently, both the dimensionality of the data and the number of modalities have a direct impact on execution time. The introduction of discriminant terms—including additional gradient and logarithmic computations—naturally increases the computational cost, although convergence remains practically attainable.

The output of the algorithm consists of the set of matrices $\mathbf{W}^{[k]}$ that produce sources that are simultaneously maximally independent and maximally separable across classes. The stopping criterion may be defined either by a maximum number of iterations or by requiring that successive updates of the matrices $\mathbf{W}^{[k]}$ fall below a predetermined threshold.

Accordingly, the following pseudocode illustrates the operation of the proposed algorithm:

Algorithm 2 DIVA-G Training Procedure

Require: Multimodal data $\mathbf{x}^{[k]}$, labels C, learning rate γ, trade-off parameter λ.

Preprocess: whiten each $\mathbf{x}^{[k]}$; initialize $\mathbf{W}^{[k]} = \mathbf{I}$.
for iteration $= 1$ to $T_{\max}$ **do**
  Compute sources: $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$.
  Evaluate IVA cost gradient: $\nabla J_{\text{IVA}}(\mathbf{W}^{[k]})$.
  Evaluate Fisher discriminant gradient: $\nabla \phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C)$.
  Update separation matrices:

$$\mathbf{W}^{[k]} \leftarrow \mathbf{W}^{[k]} - \gamma\big(\nabla J_{\text{IVA}}(\mathbf{W}^{[k]}) - \lambda \nabla \phi(\mathbf{W}^{[k]}, \mathbf{x}^{[k]}, C)\big).$$

  if convergence criterion met **then**
    break
  end if
end for
Extract discriminative sources $\mathbf{y}^{[k]}$ and concatenate across modalities.
return multimodal embeddings for downstream classification.

With these elements in place, the operation of DIVA-G can be understood as consisting of three main stages. The first stage corresponds to preprocessing, during which the datasets are formatted to meet the requirements of the algorithm. Next, as represented in the pseudocode of the algorithm, the iterative optimization stage takes place. This stage is responsible for estimating the demixing matrices $\mathbf{W}^{[k]}$ that best balance statistical independence and class separability.

Once these matrices have been obtained, the feature extraction stage begins, in which the discriminative independent sources are estimated through the linear transformation $\mathbf{y}^{[k]} = \mathbf{W}^{[k]}\mathbf{x}^{[k]}$. The resulting sources can then be used for different tasks; in the context of this work, they are employed in a classification procedure using an SVM classifier. The classifier receives $\mathbf{Y}_{\text{train}}$ and their corresponding labels for training and subsequently $\mathbf{Y}_{\text{test}}$ for class prediction. This procedure is described in detail in Section 5.2.

4.4 Limitation

Like any method, DIVA has intrinsic limitations inherent to its structure. One such limitation arises from the assumption that latent sources are linearly mixed and that class separability relies on Fisher's discriminant criterion. This dependency can restrict performance in scenarios where the relationships among sources are strongly nonlinear or when highly complex

interactions occur between modalities.

Furthermore, the requirement for labeled samples to drive the discriminative process can also be considered a limitation, particularly in situations involving severe class imbalance or when the cost of obtaining reliable labels is high.

In addition, the inclusion of the discriminative term, and consequently, the introduction of an additional gradient computation, may challenge numerical stability when dealing with extremely high-dimensional or highly multimodal datasets.

4.5 Convergence conditions

- **Identifiability:** intra-vector independence and limited cross-modal dependence. These conditions are similar to those presented in the IVA literature (Boukouvalas, 2017; Anderson et al., 2012).
- **Ergodicity / i.i.d.:** it is assumed that samples within each class and modality are sufficiently representative (i.i.d. or ergodic) to allow consistent estimates of $S_B^{[k]}$ and $S_W^{[k]}$.
- **Non-singularity / regularization:** the matrices $(W^{[k]})^\top S_{(\cdot)}^{[k]} W^{[k]}$ must remain non-singular during optimization, or alternatively be numerically regularized (e.g., by adding εI) to ensure stable matrix inversions.
- **PDF modeling:** the multivariate probability density function (PDF) of the SCVs must be properly modeled (parametric, semi-parametric, or non-parametric); modeling inaccuracies affect both the consistency and efficiency of the estimator (Damasceno et al., 2021).

4.6 Summary

This chapter presented the complete formulation of DIVA-G, from its motivation to its mathematical construction and computational implementation. By integrating the Fisher discriminant criterion into the IVA framework, the proposed method combines statistical independence with class separability, providing a new supervised approach for the analysis of multimodal data. Understanding these foundations is essential for correctly interpreting the behavior and performance of the algorithm. In the next chapter, these concepts are put into practice through experiments that evaluate DIVA-G across different datasets, enabling an assessment of its effectiveness and a comparison with related methods.

5 PERFORMANCE EVALUATION: ANOMALY AND MISINFORMATION DETECTION

The parameter λ is a regularization coefficient that controls the influence of the discriminant term in the optimization process, and its value has a direct impact on the final results. However, it was not possible to derive an analytical expression to determine its optimal value. Therefore, a heuristic exploration was performed.

For larger and more complex datasets, where the computational cost is considerably higher, λ was empirically assigned the following values: [0, 0.0001, 0.001, 0.01, 0.1, 1].

For smaller datasets, such as the synthetic ones, a Bayesian-like search was carried out within the interval $[0, 1]$ using a step size of 0.001. Empirical observations indicated that the optimal value for λ typically lies within this range, although it cannot be guaranteed that the value found corresponds to a global rather than a local optimum.

The experiments were conducted to evaluate performance both across different DIVA implementations, each using distinct datasets configured with the value of λ that yielded the best results, and within the same dataset, by comparing the optimally configured DIVA model against other IVA-derived algorithms. Among all tested approaches, DIVA was the only supervised method, whereas the others operated in a fully unsupervised manner.

This chapter is organized into four sections. The first presents the datasets used in the experiments, as well as the preprocessing procedures applied to each of them to ensure their suitability for the proposed algorithm. Section 5.2 describes the classification process performed using the sources estimated by DIVA-G. Section 5.3 presents and discusses the results obtained, enabling an evaluation of the method's performance in the classification scenario. Finally, Section 5.4 provides a summary of the main findings of the chapter.

5.1 Datasets

To evaluate the performance of DIVA-G, both synthetic and real datasets were employed. The synthetic datasets allow the analysis of different characteristics that can reveal the most suitable conditions for the proposed method, providing agility and flexibility in experimentation. In contrast, the real datasets demonstrate the applicability of the method to real-world scenarios, which naturally involve unpredictable variations.

5.1.1 Synthetic

The synthetic datasets were constructed with 1,000 samples, with 30% allocated for training and 70% for testing, while maintaining the desired balance between classes. For each modality k, n independent sources were generated, modeled from one-dimensional Gaussian distributions. Samples belonging to class 0 were drawn from a distribution with mean μ_0 and standard deviation σ_0, whereas samples from class 1 were generated from a distinct distribution with mean μ_1 and standard deviation σ_1. This procedure was independently repeated for each dimension n and each modality k, ensuring that the sources were statistically independent from one another. The means were sampled from the closed interval [0, 1], and the variances were drawn from the open interval (0, 1]. This process ensures that each class exhibits different regions of concentration in the feature space, introducing both partial overlap between classes and intermodal dependency, conditions that are fundamental for evaluating the discriminative performance of DIVA-G.

After generating the pure sources $\mathbf{s}^{[k]}$, each dataset was subjected to a linear transformation through a randomly generated mixing matrix $A^{[k]}$, resulting in the observed signals $\mathbf{x}^{[k]} = A^{[k]}\mathbf{s}^{[k]}$. Finally, the data were standardized using *z-score* normalization, with the parameters estimated from the training set and applied to the test set, thereby ensuring consistency among modalities and numerical stability for subsequent experiments.

The class proportions in the datasets were also varied to investigate the impact of class separability. Three groups of datasets were created, each defined by sources with the same Gaussian parameters but with sample imbalance varying according to the following proportions for the minority class: 1%, 5%, 10%, 20%, 30%, 40%, and 50%. The group named *Sint1* was composed of sources with Gaussian parameters [(0.5, 0.5), (0.9, 1.0)], where each ordered pair corresponds to, respectively, the mean and variance of the Gaussian matrix that generates one of the classes. The *Sint2* group, characterized by strongly overlapping distributions, used parameters [(0.1, 0.3), (0.1, 0.3)], while *Sint3* employed [(0.9, 0.5), (0.5, 0.9)], representing well-separated classes in the sample space.

5.1.2 Real data

Two real-world datasets were used to demonstrate the applicability of the method and to represent the method's behavior on data with a higher level of complexity.

5.1.2.1 NLS-KDD

The first real dataset used in this study was the *NSL-KDD*[1], which represents a revised and improved version of the original KDD'99 dataset, developed by the University of New Brunswick (UNB). The main objective of this revision was to eliminate redundancies, reduce class imbalance, and minimize biases that previously compromised the evaluation of intrusion detection models. As discussed by Khraisat et al. (2019), the NSL-KDD has become one of the most widely used benchmark datasets in the fields of Machine Learning and Cybersecurity, combining realistic network traffic, manageable complexity, and public availability, making it ideal for standardized comparison across intrusion detection approaches (IDS).

Unlike the KDD'99 dataset, which contained numerous duplicate instances and highly unbalanced class distributions, the NSL-KDD provides carefully balanced training and test subsets. This ensures fairer and more reproducible evaluations of machine learning algorithms (Zhang et al., 2019).

The *NSL-KDD* dataset contains 43 attributes per network connection: 41 extracted features and two label columns (*class* and *difficulty level*). The *class* column indicates whether a connection is normal or corresponds to a specific type of attack, while the *difficulty level* column quantifies the complexity of each sample, allowing more refined performance analyses of detection models.

The 41 traffic features are conventionally grouped into four functional categories, according to the nature of the collected information and their role in characterizing network behavior (Zhang et al., 2019), and this grouping is used to form the $k = 4$ subsets:

- **Basic features:** *duration*, *protocol_type*, *service*, *flag*, *src_bytes*, *dst_bytes*, *land*, *wrong_fragment*, *urgent*.
- **Content features:** *hot*, *num_failed_logins*, *logged_in*, *num_compromised*, *root_shell*, *su_attempted*, *num_root*, *num_file_creations*, *num_shells*, *num_access_files*, *num_outbound_cmds*, *is_host_login*, *is_guest_login*.
- **Time-based traffic features:** *count*, *srv_count*, *serror_rate*, *srv_serror_rate*, *rerror_rate*, *srv_rerror_rate*, *same_srv_rate*, *diff_srv_rate*, *srv_diff_host_rate*.
- **Host-based traffic features:** *dst_host_count*, *dst_host_srv_count*, *dst_host_same_srv_rate*, *dst_host_diff_srv_rate*, *dst_host_same_src_port_rate*, *dst_host_srv_diff_host_rate*, *dst_host_serror_rate*, *dst_host_srv_serror_rate*,

[1] <https://www.kaggle.com/datasets/hassan06/nslkdd>

dst_host_rerror_rate, *dst_host_srv_rerror_rate*.

Among the aforementioned variables, several have a categorical nature, such as *protocol_type*, *service*, and *flag*. To enable their use in machine learning models such as *DIVA-G*, an encoding step was applied to convert them into numerical representations. Specifically, a *one-hot encoding* procedure was used, transforming each category into a binary vector while avoiding the introduction of artificial hierarchies among categorical values (Khraisat et al., 2019).

Subsequently, to satisfy the dimensional uniformity requirement across the subsets used by *DIVA-G*, a dimensionality reduction step was performed using *Principal Component Analysis* (PCA). This technique projects the data into a lower-dimensional space while preserving the directions of highest variance and minimizing redundancy, resulting in a more compact and efficient representation.

The preprocessing described above was consistently applied to the three subsets provided in the *NSL-KDD* dataset and used in this work: *Small Training Set*, *Train+*, and *Test+*. After encoding and dimensionality reduction, all subsets exhibited equivalent dimensions, organized as follows: Small Training Set $(V = 1011, N = 9, K = 4)$, KDDTrain+ $(125.973, 9, 4)$, and KDDTest+ $(22.543, 9, 4)$.

5.1.2.2 MediaEval2016

The second dataset used in this study is the *MediaEval2016 Image Verification Corpus*[2] (Boididou et al., 2018), which has been widely employed in research on disinformation detection and multimodal content verification on social media. This corpus contains tweets and images related to high-impact events such as natural disasters, terrorist attacks, and political crises. Each tweet is labeled as either *fake* or *real*, where the *fake* label indicates that the multimedia content does not faithfully represent the described event, while the *real* label refers to posts that accurately correspond to the reported event.

The training set consists of 9,140 tweets associated with 352 images and 15 events, of which five include both real and fake tweets, and ten contain only fake ones. The test set comprises 796 tweets related to 92 images and 23 events, including seven with both real and fake tweets, fifteen with only fake tweets, and one with only real tweets. This unbalanced structure reflects the inherent heterogeneity and realism of social media information during major events, making the corpus particularly suitable for evaluating multimodal fusion and multivariate

[2] <https://github.com/MKLab-ITI/image-verification-corpus>

analysis approaches.

Text preprocessing involved the removal of emojis, URLs, user mentions, stopwords, and punctuation, as well as lemmatization, lowercasing, and the elimination of duplicate or retweeted messages. Only English-language tweets were retained to ensure greater linguistic consistency within the dataset (Damasceno et al., 2022; Damasceno et al., 2024). For the image modality, all images were resized to 224×224 pixels and normalized to standardize their scale and facilitate the extraction of visual features.

For textual representation, a dense word embedding model (*Word2Vec*) with 300 dimensions was employed. Each tweet was represented by the average of the word vectors composing it, thus capturing semantic relationships within the vector space. For visual representation, 4,096-dimensional feature vectors were extracted from the fully connected layer of a pre-trained convolutional model of type VGG-16. This fusion strategy is consistent with the multivariate paradigm underlying IVA models, which exploit both independence across modalities and internal dependencies within each modality (Kim et al., 2006; Luo, 2023).

After the feature extraction process, the tweets were divided into training and test sets. For each modality (text and image) observation matrices were constructed, where each row represents a feature vector and each column corresponds to a sample. Subsequently, PCA was applied separately to each modality to reduce dimensionality while preserving the components with the highest variability. After dimensionality reduction, each modality was represented with $N = 300$ features. The resulting matrices were then stacked to form a three-dimensional tensor, such that $\mathbf{X} \in \mathbb{R}^{N \times V \times K}$, where $K = 2$ denotes the number of modalities (text and image), and V corresponds to the number of samples, with $V_{\text{train}} = 9140$ for the training set and $V_{\text{test}} = 796$ for the test set. The same procedure was applied consistently to both training and test data.

5.2 Classification

After the preprocessing stage, the training data are fed into the DIVA-G algorithm, yielding as output a set of separation matrices $\mathscr{W} = \{\mathbf{W}^{[k]}\}$. These matrices represent the optimal transformations that project the original observations $\mathbf{X}^{[k]}$ into a new latent space, resulting in the estimated independent sources $\mathbf{Y}^{[k]} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}$. Thus, the latent representations for the training data are obtained as $\mathbf{Y}^{[k]}_{\text{train}} = \mathbf{W}^{[k]}\mathbf{X}^{[k]}_{\text{train}}$.

Once the independent sources have been extracted, a supervised classification stage is performed. The latent representations $\mathbf{Y}^{[k]}_{\text{train}}$, now decorrelated and discriminative, are used as

input feature vectors for training a classifier based on *Support Vector Machines* (SVM).

The SVM is employed due to its robustness in separating classes in high-dimensional spaces and its strong generalization ability even with limited training data. In this setting, the classifier learns the decision boundaries that best separate the classes within the latent space provided by DIVA, effectively mapping the estimated independent sources to their corresponding output categories.

During the inference phase, the matrices $\mathbf{W}^{[k]}$ obtained during training are applied to new input data $\mathbf{X}_{\text{test}}^{[k]}$ to generate their corresponding latent representations $\mathbf{Y}_{\text{test}}^{[k]}$, which are then passed to the SVM classifier to predict the class labels.

This processing pipeline allows the independent component analysis to be guided by discriminative criteria, simultaneously maximizing the statistical independence of the sources and the class separability within the latent subspace of interest.

5.3 Results

This section presents the results obtained using both the synthetic and real datasets introduced in the previous subsections, along with discussions and analyses of the outcomes.

5.3.1 Synthetic

First, using the groups of synthetic datasets with different distributions between classes, we sought to analyze whether the method is affected by class imbalance and how this relationship manifests. To this end, we began by examining the datasets in order to gain a deeper understanding of their characteristics.

The three groups of datasets created consist of distributions with different class proportions, as described in Section 5.1.1, and each group was constructed from pairs of Gaussian distributions representing the two classes.

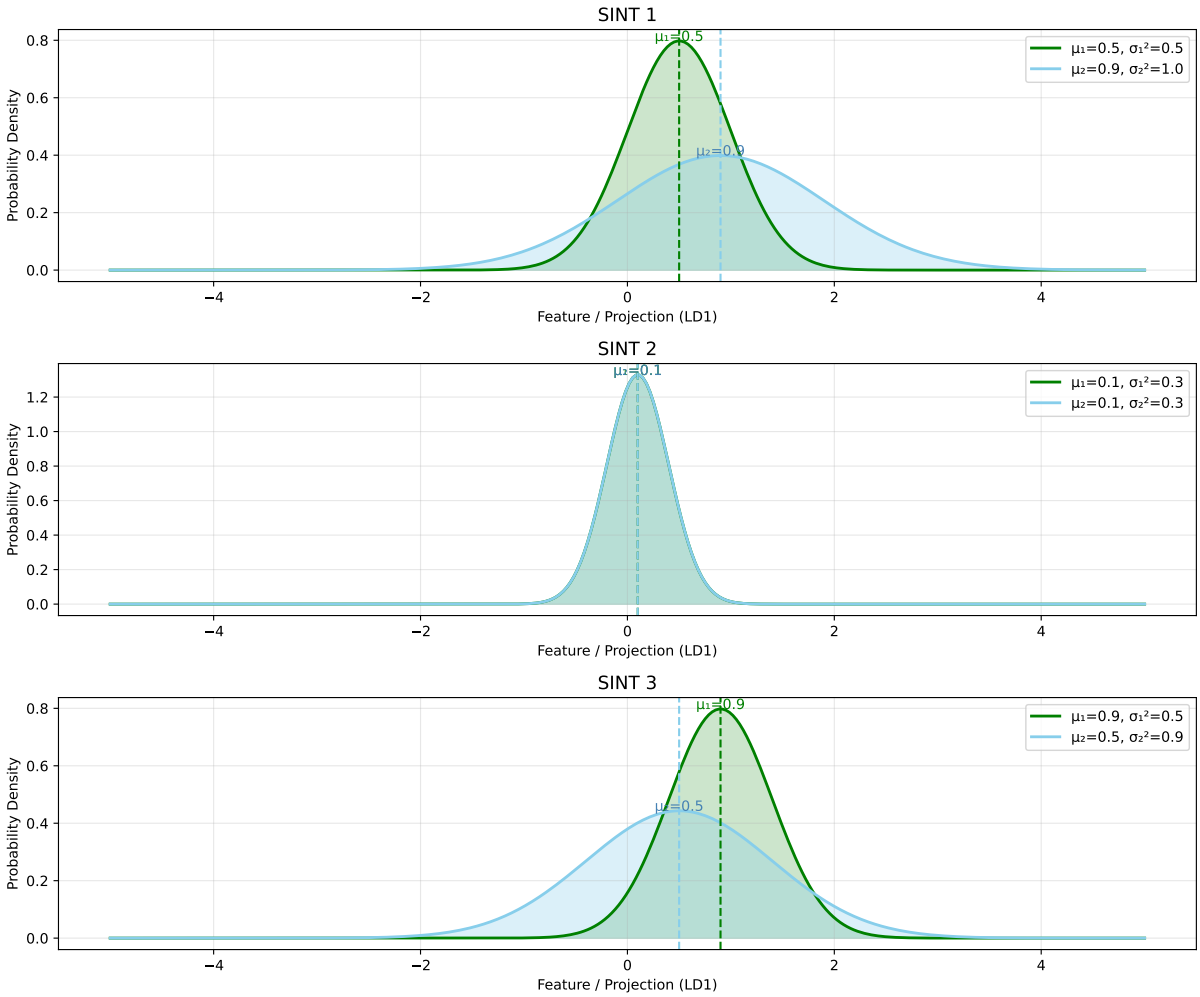
In the SINT1 group, the majority class (class 0) is modeled by a Gaussian distribution with mean 0.5 and variance 0.5, whereas the minority class has mean 0.9 and variance 1.0. The Gaussian pair is presented in the first plot of Figure 3, allowing visualization of both their overlap and the individual behavior of each class. It can be observed that the majority class exhibits lower variance, which results in greater concentration of values around the mean; theoretically, this makes the data from this class more homogeneous compared to those from the minority

class, whose higher variance produces a broader spread.

The SINT2 group consists of two fully overlapping Gaussian distributions, as illustrated in the central plot of Figure 3. Both classes have mean 0.1 and variance 0.3, implying that all data points, regardless of class, share identical probabilities for their values. Moreover, the reduced variance confers limited variability around the mean, making this dataset particularly challenging for class identification.

In the SINT3 group, the majority class is represented by a Gaussian distribution with mean 0.9 and variance 0.5, while the minority class has mean 0.5 and variance 0.9, as shown in the third plot of Figure 5. There is a structural similarity to SINT1, but with a reversal of the class means that shifts the region of overlap between the Gaussians. Additionally, the minority class exhibits lower variance compared to SINT1, resulting in reduced dispersion of its values. These modifications subtly alter the relationship between the distributions, and the relevance of these differences will be discussed in the analysis of the results for each dataset implementation.

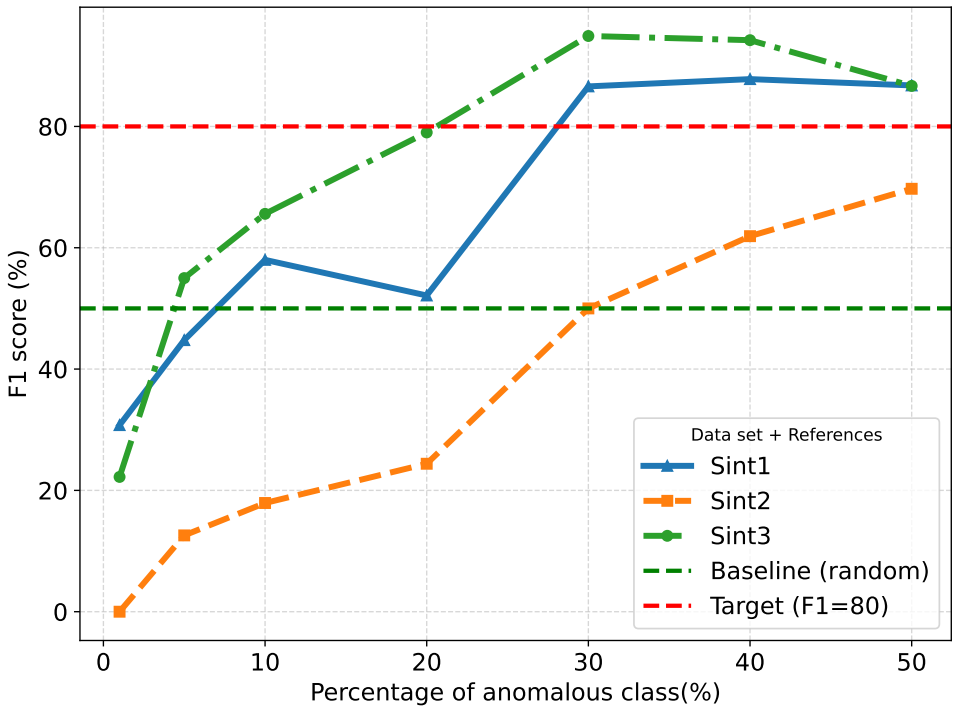
Figura 5 – Graphical representation of the Gaussian classes for the 3 groups of data sets, SINT1, SINT2 and SINT3.



Source: Prepared by the author.

The individual implementation of the SINT1, SINT2, and SINT3 datasets in DIVA-G enables the analysis of how the classification performance varies according to the class distributions in each scenario. Figure 4 presents the F1-score percentages of the minority class for each of the seven imbalance proportions evaluated across the three groups of synthetic data. These proportions correspond to the relative participation of class 0 in the datasets, following the percentages: 1%, 5%, 10%, 20%, 30%, 40%, and 50%.

Figura 6 – Performance evaluation of IVA proposed method of the F1-score for different synthetic datasets.



Source: Prepared by the author.

It is observed that, regardless of the statistical characteristics of each dataset, the performance of the method increases as the class imbalance decreases. However, when each group is examined individually, and compared with one another in light of their construction properties, additional insights can be drawn. The SINT3 group generally exhibits the best results, except at the 1% minority proportion, for which SINT1 slightly outperforms it. Recalling the structural similarity between SINT1 and SINT3, the proximity in their performance becomes evident; nonetheless, the reduced variance of the minority class in SINT3 plays a crucial role, indicating that lower variability in this class contributes to improved performance.

The SINT2 group, in contrast, yields the worst performance among the three. This outcome is directly attributable to the high degree of difficulty introduced by the complete overlap between the class distributions, as both classes share identical generating statistics. Thus, we can say that the greater the separation between the sources in the sample space, the better the classification performance tends to be. This behavior is consistent, as it is easier to identify the origin of a given sample when the sources have distinct statistical characteristics.

From a broader perspective, even in the most challenging scenario, the method achieves more than 50% correct identification when at least 30% of the samples belong to

the minority class. Consequently, for the remaining synthetic datasets used in this work, the proportion of 30% for class 0 and 70% for class 1 was adopted, enabling the assessment of the method under a still demanding yet sufficiently informative and high-performing setting.

With the aim of more precisely analyzing the relationship between the statistical characteristics of the datasets and the performance of DIVA-G, Table 1 presents the statistical construction properties of ten datasets, together with their corresponding F1-scores for the minority class.

Tabela 1 – Parameters and performance obtained for each synthetic dataset.

μ_0	σ_0^2	μ_1	σ_1^2	Dataset ID	F1–score
1.0	1.0	0.9	0.9	1	82.77
0.9	1.0	0.5	0.5	2	87.09
0.9	0.5	0.5	0.9	3	94.91
0.7	0.3	0.7	0.3	4	99.76
0.5	0.5	0.9	1.0	5	86.60
0.5	0.5	0.6	0.5	6	82.04
0.5	0.4	0.4	0.3	7	71.31
0.2	0.3	0.3	0.1	8	44.20
0.1	0.3	0.1	0.3	9	50.00
0.1	1.0	0.0	1.0	10	42.22

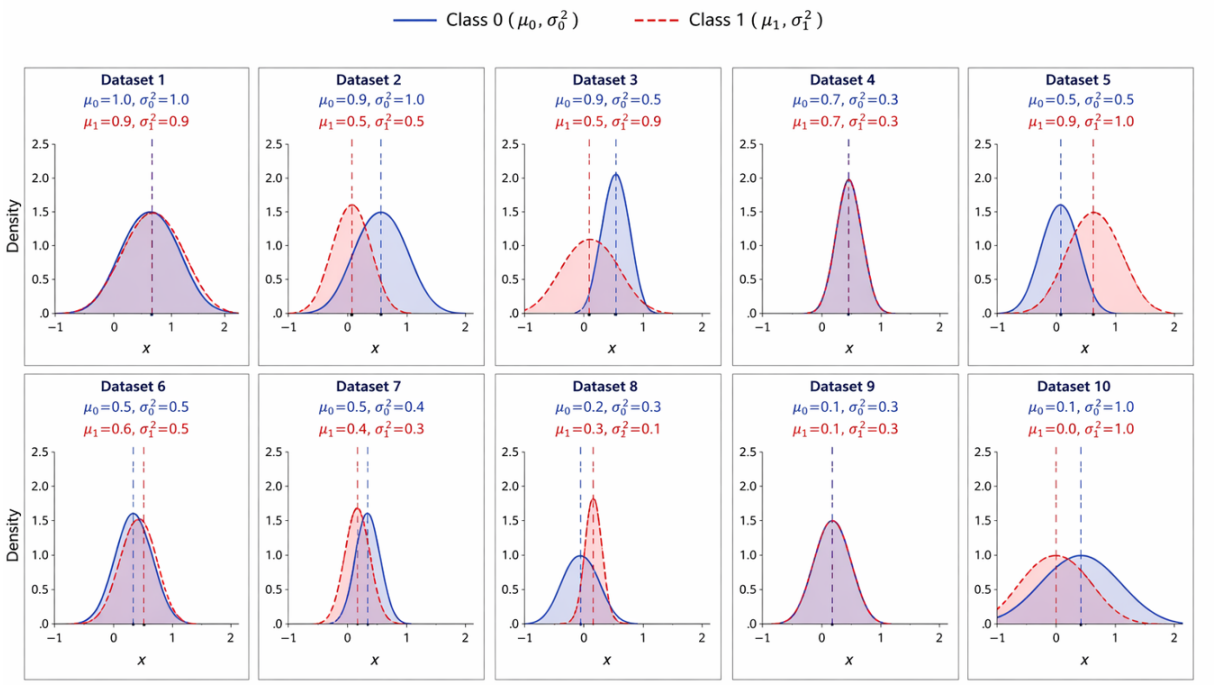
Source: Prepared by the author.

To enrich the discussion of the results obtained from the implementation of the ten datasets, Figure 7 presents visual representations of the pairs of Gaussian distributions corresponding to the class-generating distributions described in Table 1.

Based on the results obtained by the method, it is observed that the means and variances exert distinct influences on the performance of DIVA-G. When comparing datasets 2 and 5, one notes an inversion in the statistical characteristics defining the classes; however, this change has only a minimal impact on the results. In other words, even when the minority class, which represents only 30% of the dataset, exhibits a variance twice as large, the performance is only slightly affected.

Datasets 5 and 7 highlight another important aspect: although they preserve nearly the same distribution for class 0, the results indicate that having more widely separated class means is more advantageous than having lower variances. In dataset 5, despite the higher variance of class 1, the increased mean of this class enlarges the separation between distributions, yielding better performance. We can also compare datasets 2 and 3, which maintain identical

Figura 7 – Graphical representations of the relationships between the Gaussian processes forming the classes of the sets in Table 1.



Source: Prepared by the author.

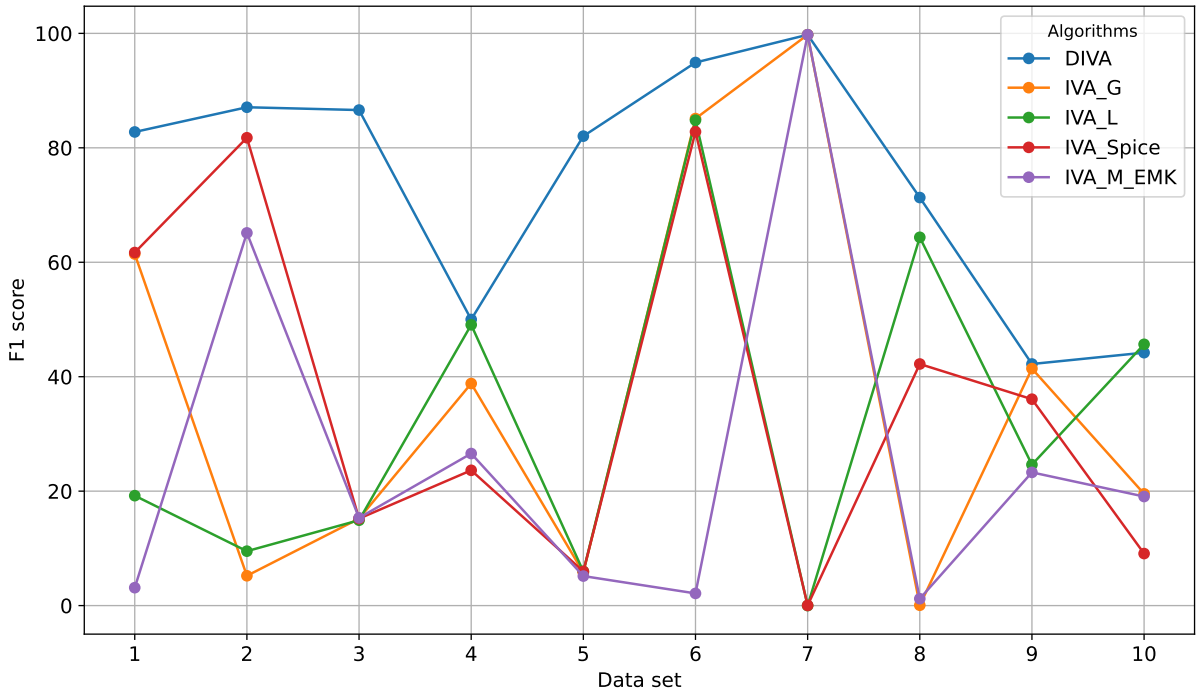
means but present nearly inverted variances. The results show that higher variances tend to slightly reduce performance, a behavior also observed when comparing datasets 10 and 8, where significant changes in variance lead to only minor differences in final performance.

In contrast, datasets 4 and 9 represent scenarios of complete overlap between the Gaussian distributions that define the classes; however, their results differ substantially. The key distinction between these datasets lies in the class means, as both share the same variance (0.3). Highlighting this aspect is important, as it reinforces the observation that higher class means are associated with improved performance, a trend consistently observed across the results.A more detailed analysis can be obtained by comparing datasets 2 and 5, where an inversion of class characteristics occurs. This comparison provides further evidence that configurations in which class 0 exhibits a higher mean than class 1 tend to yield superior performance. It is also important to note that DIVA-G is sensitive to the initialization process, which underscores the need for a careful and comprehensive interpretation of the results.

We now aim to assess whether the performance achieved by DIVA-G on the 10 datasets of Table 1 is indeed promising, or whether other methods from the same family, namely, algorithms derived from IVA, yield superior results. To this end, the plot presented in Figure 8 compiles the best performance obtained for each of the 10 datasets when implemented using five

IVA-based algorithms, including both DIVA-G and IVA-G.

Figura 8 – F1-score of the minority class of the sets in Table 1 when subjected to the classification process using the results of the IVA-G, IVA-L, IVA-Spice, IVA-A-EMK and DIVA-G algorithms,



Source: Prepared by the author.

The IVA-G algorithm adopts a multivariate Gaussian model, providing a simple and computationally efficient basis (Anderson et al., 2010). In contrast, IVA-L (based on Laplace) assumes a multivariate Laplacian distribution, which is more suitable for capturing sparse source structures (Kim et al., 2006). Building on this idea, IVA-SPICE incorporates priors that promote sparsity to increase robustness in scenarios with structured or low-energy components (Boukouvalas, 2023). Finally, IVA-M-EMK employs a flexible mixture modeling approach combined with entropy-maximizing kernels, allowing the algorithm to adapt to a wide range of source distributions (Damasceno et al., 2024). This perspective enables a more critical analysis of the results.

It is important to emphasize that the connections between points in the plot serve only to facilitate visualization; each point is independent, since the performance obtained on dataset 2 depends solely on that dataset and is unrelated to the performance on datasets 1 or 5.

Examining the overall behavior, one observes that DIVA-G, represented in blue, achieves the best performance in nearly all implementations, demonstrating its consistency and reliability. Nevertheless, several aspects deserve further attention to enrich the analysis.

Datasets 5 and 6, for example, concentrate the performance of all methods below the 20% threshold, with the exception of DIVA-G. This discrepancy highlights the substantial positive impact of incorporating label information into DIVA-G, which enables more accurate class identification and mitigates the limitations faced by purely unsupervised methods.

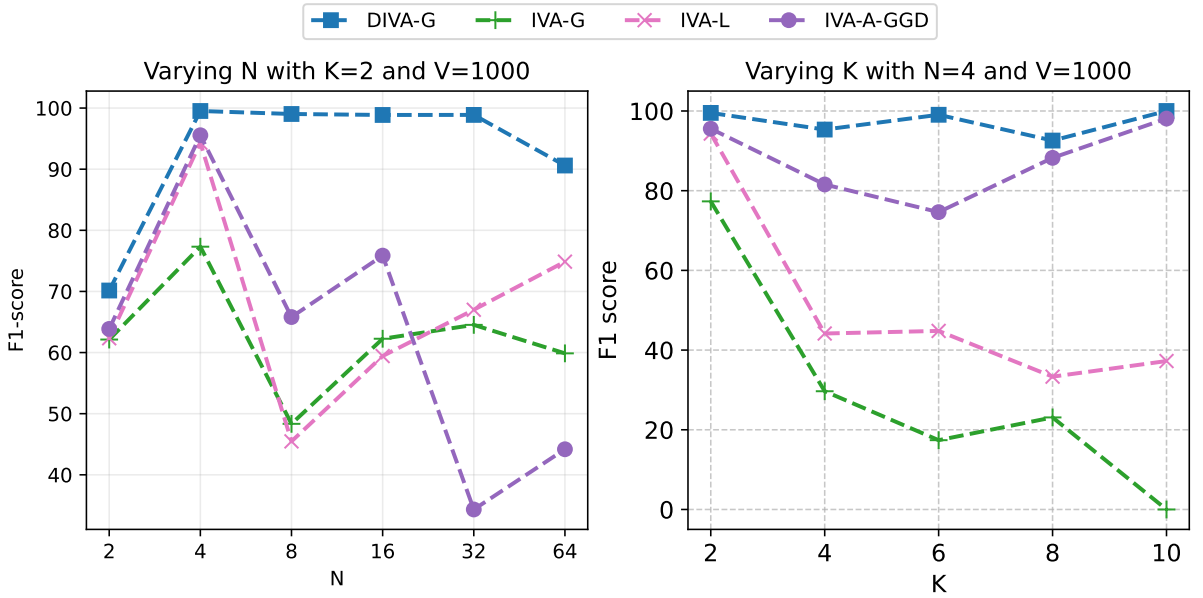
Dataset 4 yields the highest performance for DIVA-G, IVA-G, and IVA-E-EMK. This result suggests that, although the Gaussian distributions forming the classes are fully overlapped, the dataset is relatively easy to classify for most of the evaluated methods. In contrast, datasets 9 and 10, also presenting a high degree of overlap between distributions, produce performances below 50% for all algorithms. This indicates that classification difficulty is not determined solely by the statistical properties of the generating Gaussians; factors related to the intrinsic randomness of the data generation process also significantly influence the results.

Overall, the plot shows that despite the challenges imposed by the statistical characteristics of the datasets, the incorporation of the discriminative term, thus transforming the method into a supervised approach, enhances the performance of IVA-G, the base algorithm, enabling it to overcome limitations that affect the other IVA-derived methods.

Now, we aim to evaluate the performance of DIVA-G when the datasets share the same Gaussian parameters defining the classes, but vary in data dimensionality. Using the Gaussian parameters of dataset 4 from Table 1 and a minority-class proportion of 30%, we varied both the number of features (N) and the number of subsets (k). Accordingly, six datasets were created by fixing $k = 2$ and varying the number of features as $N = \{2, 4, 8, 16, 32, 64\}$, and five datasets were generated by fixing $N = 4$ and varying $k = \{2, 4, 6, 8, 10\}$. All datasets followed the same structure described in the *Synthetic Data* section.

These datasets were applied to four IVA-based algorithms: IVA-G, IVA-L, IVA-A-GGD, and the proposed DIVA-G, among which only DIVA-G is supervised, requiring labeled data during its training process. The results of these implementations are presented in Figure 9, which shows the F1-score values for the minority-class classification based on the outputs produced by the tested algorithms.

Figura 9 – Performance (F1-score %) obtained by varying the parameters of the datasets, number of sources, and number of subsets.



Source: Prepared by the author.

Analyzing first the graph on the left, which represents the variation in the number of features, it can be observed that for $N = 2$, all algorithms achieved similar results. However, as N increases, their performances begin to diverge. It is noteworthy that DIVA-G consistently maintains high accuracy levels, above 90% for all values of N greater than 2, exhibiting an almost steady behavior with minimal oscillations. In contrast, the other algorithms demonstrate a weaker ability to handle variations, showing multiple performance peaks and fluctuations throughout the graph.

Regarding the variation in the number of subsets, it can be observed that the results for $K = 2$ are comparable to those obtained for $N = 4$ in the feature-variation graph, representing the configuration that yields the most similar outcomes across the four algorithms. For $K = 2$, only IVA-G presents results below 80%, which is expected, as it corresponds to the first version of a multivariate IVA algorithm. The subsequent variants were developed precisely to overcome the limitations of IVA-G. Nevertheless, as the number of subsets increases, the performance of IVA-L drops sharply, approaching that of IVA-G, although still slightly superior. On the other hand, IVA-A-GGD is the algorithm that most closely approaches the performance of DIVA-G, deviating only for $K = 4$ and $K = 6$. It is worth emphasizing that DIVA-G maintains performance levels close to 100% even as the number of subsets increases, demonstrating its robustness and reliability across diverse scenarios.

5.3.2 Real data

All analyses presented so far in this work were conducted using synthetic data. However, to better understand the behavior of DIVA-G when dealing with the complexity of real-world data, we proceed with the results obtained from the implementation on the NSL-KDD dataset, followed by the experiments performed on the MediaEval2016 dataset.

5.3.2.1 NLS-KDD

The NSL-KDD dataset contains information related to network security and, after the preprocessing described in Section 5.1.2, it was treated as a multimodal dataset for binary classification, with the categories *Attack* and *Normal*. This dataset provides two distinct training sets: the *Small Training Set*, containing 1,011 samples, and *KDDTrain+*, with 125.973 samples. Given this structure, the classification process was carried out independently for each training set, applying four IVA-derived methods—IVA-G, IVA-L, IVA-A-GGD, and the proposed DIVA-G, with the goal of assessing and comparing their performance.

Table 2 reports the F1-scores obtained by each method in the classification task using both training sets.

Tabela 2 – Performance comparison of IVA algorithms on the NSL-KDD dataset (F1-score %).

Algorithm	**DIVA-G**	**IVA-G**	**IVA-L**	**IVA-A-GGD**
Small Training Set	**77.48**	70.76	76.60	37.26
KDDTrain+	**77.66**	42.36	53.74	46.35

Source: Prepared by the author.

It can be observed that the performance of the IVA-G and IVA-L algorithms drops sharply when using the *KDDTrain+* dataset, whereas the IVA-A-GGD shows a considerable improvement under the same condition. Although IVA-L outperforms both IVA-G and IVA-A-GGD for both training sets, it still fails to reach the performance achieved by DIVA-G.

Furthermore, even though the *KDDTrain+* dataset contains a significantly larger number of samples, the performance gain achieved by DIVA-G compared to the *Small Training Set* is marginal, indicating that the increase in data volume does not offset the additional computational cost. Another important point is that the *Small Training Set* does not include all attack configurations present in the test set and in *KDDTrain+*, which further reinforces the robustness of DIVA-G, as it is able to maintain similar performance even under more limited

training conditions.

5.3.2.2 MediaEval2016

Using the real MediaEval2016 dataset, whose characteristics were described in Section 5.1.2, we analyze the classification performance obtained from the sources estimated by DIVA-G. To this end, the training set, composed of 9,142 samples, was applied to DIVA-G using six different values of $\lambda = [0, 0.0001, 0.001,\ 0.01, 0.1, 1.0]$. All experiments were executed with the same initialization to ensure a fair comparison across configurations. It is noteworthy that DIVA-G with $\lambda = 0$ is equivalent to the IVA-G method. Table 3 presents the F1-score values obtained for each of the six λ settings.

Tabela 3 – Performance (F1-score %) obtained by applying the MediEval 2016 dataset with different λ values in DIVA-G.

λ values	**F1-score %**
0	49.87
0.0001	53.77
0.001	60.05
0.01	61.93
0.1	52.89
1	51.01

Source: Prepared by the author.

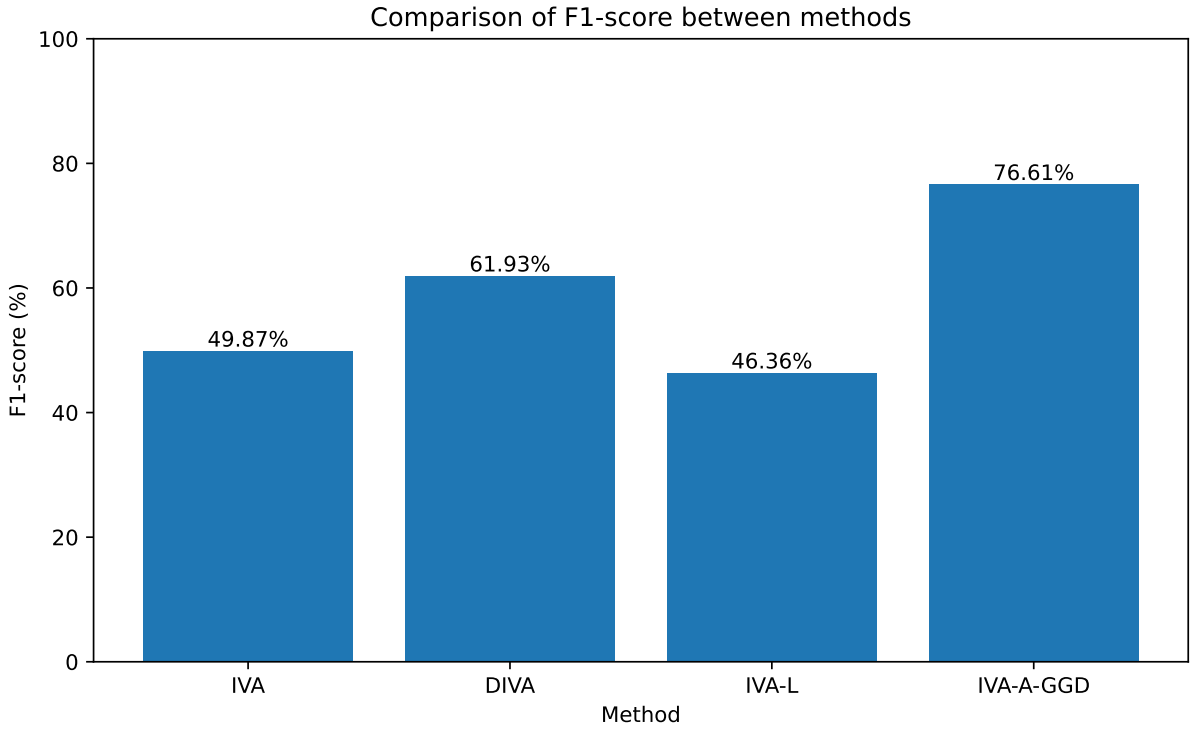
The use of the same initialization for all DIVA-G configurations tends to induce similar behavioral patterns across executions, which implies that a more favorable initialization might exist and could potentially lead to superior results. Furthermore, it is important to note that there is no guarantee that the tested values of λ are the most suitable. However, due to the high computational cost associated with the complexity of the MediaEval2016 dataset, it was not feasible to perform a more exhaustive search for the optimal λ within the scope of this work.

The analysis of Table 3 shows that $\lambda = 0.01$ produced the highest F1-score among the tested configurations. It can also be observed that, for all executions that included the discriminative term, that is, for all values of λ except $\lambda = 0$, the F1-score exceeded 50%. This reinforces the promising nature of incorporating the discriminative criterion into the IVA framework.

Based on the best performance achieved by DIVA-G on the MediaEval2016 dataset, the next step is to compare it with other IVA-derived algorithms in order to assess the extent to

which the proposed method offers advantages over existing approaches.

Figura 10 – Best performance (F1-score %) obtained by applying the MediaEval2016 dataset to DIVA-G, IVA-G, IVA-L, and IVA-A-GGD.



Source: Prepared by the author.

Based on the results presented in Figure 10, it is observed that DIVA-G outperforms both the original IVA-G method and IVA-L, indicating that the incorporation of the discriminative term effectively improves the identification of the minority class in the MediaEval2016 dataset. This performance gain reinforces the importance of leveraging label information within the IVA framework, thereby transforming the process into a supervised approach.

However, for the tested values of λ, DIVA-G did not reach the performance achieved by IVA-A-GGD, which obtained the highest F1-score among all evaluated methods. This result suggests that, although the discriminative term provides clear benefits, there remains room for improvement, whether through a more refined selection of the parameter λ, the use of more suitable initialization strategies, or additional adjustments to the discriminative formulation.

Overall, the results indicate that DIVA-G is both competitive and promising, particularly given its supervised nature and its relative simplicity compared to more sophisticated models such as IVA-A-GGD. Nevertheless, the pursuit of improved performance remains a challenge to be addressed in future research.

5.4 Summary

In this chapter, the experiments supporting the evaluation of the proposed method, DIVA-G, were presented. The detailed description of the datasets, preprocessing steps, and classification procedure enabled a comprehensive analysis of the algorithm's behavior under both controlled scenarios and real-world conditions, including structural variations, noise, and different levels of complexity. The results discussed throughout the chapter demonstrate not only the robustness and discriminative potential of the method but also highlight its advantages over other IVA-based approaches. Consequently, this chapter plays a central role by providing solid empirical evidence that supports the conclusions presented in the following chapter and reinforces the relevance of DIVA-G as a promising contribution to separation and classification problems in multimodal data.

6 CONCLUSIONS AND FUTURE DIRECTIONS

The recovery of latent information from mixed observations characterizes a typical Joint Blind Source Separation (JBSS) problem, within which Independent Vector Analysis (IVA) has been widely employed. However, classical IVA methods treat all components equivalently, disregarding potential discriminative information that could guide the separation process. This limitation restricts the applicability of IVA in classification scenarios, where maximizing class separability is as important as enforcing statistical independence.

To overcome this structural limitation, this work introduced the *Discriminant Independent Vector Analysis* (DIVA), which incorporates into the traditional IVA framework a discriminative term based on Fisher's Linear Discriminant (FLD) criterion. This additional term enables the simultaneous optimization of two complementary objectives: (i) enforcing independence among the estimated sources and (ii) maximizing class discrimination. The proposed formulation extends the FLD to the multivariate setting, allowing its direct integration into the iterative estimation process for the demixing matrices. Built upon the IVA-G algorithm, the resulting DIVA-G model establishes a conceptual link between blind source separation and supervised learning, providing a mathematically coherent framework for extracting more informative and discriminative representations.

The analysis using synthetic data was essential for assessing the behavior of the method under controlled conditions. The experiments revealed that DIVA-G tends to improve its performance as class imbalance decreases, with moderately balanced proportions—such as 30/70—already yielding F1-scores consistently above 50% for the minority class. Furthermore, variations in the means and variances of the class-generating distributions influence performance in distinct ways: larger differences in class means have a more substantial impact on separability than moderate reductions in variance. Even in scenarios with strong overlap between class distributions, DIVA-G demonstrated competitive performance.

When subjected to increased structural complexity—either by raising the number of sources or the number of subsets, DIVA-G maintained high and remarkably stable performance, consistently outperforming or matching the other IVA-based methods evaluated. These findings indicate that incorporating the discriminative term renders the method more robust and less sensitive to the generative characteristics of the data.

Validation on real-world datasets reinforced these conclusions. In the NSL-KDD dataset, DIVA-G achieved the highest performance among all evaluated algorithms, demons-

trating robustness to noise and high dimensionality. In the MediaEval2016 dataset, despite the intrinsic complexity and contextual variability of multimodal data, the method obtained the second-highest F1-score, surpassing unsupervised IVA variants and approaching the performance of more sophisticated specialized models.

Overall, the results demonstrate that DIVA-G represents a meaningful advancement toward supervised JBSS methods. By embedding discriminative information into the separation process, the model significantly expands the utility of IVA in classification tasks, offering greater stability, interpretability, and performance. These findings point to promising directions for further refinement of the method and for advancing the analysis of real-world multimodal applications.

6.1 Future directions

Although DIVA demonstrates promising results and successfully overcomes several limitations of traditional IVA, some restrictions remain open for further investigation. One of the main challenges concerns the efficient determination of the regularization parameter λ, whose value directly influences the balance between statistical independence and class separability, and consequently affects the overall performance of the method. Moreover, the performance of DIVA is sensitive to the initialization of the demixing matrices, suggesting the need for more robust or adaptive initialization strategies.

Additional opportunities for improvement include the exploration of alternative discriminative functions, refinements to the cost terms, and the incorporation of more sophisticated mechanisms for controlling intra-class variance. Such enhancements may further increase the stability and generalization capability of the method in complex or highly heterogeneous scenarios.

REFERÊNCIAS

ANDERSON, M.; ADALI, T.; LI, X.-L. Joint blind source separation with multivariate Gaussian model: algorithms and performance analysis. *IEEE Transactions on Signal Processing*, v. 60, n. 4, p. 1672–1683, 2012. Disponível em: https://doi.org/10.1109/TSP.2011.2181836. Acesso em: 5 jul. 2024.

ANDERSON, M.; LI, X.-L.; ADALI, T. Nonorthogonal independent vector analysis using multivariate Gaussian model. In: *International Conference on Latent Variable Analysis and Signal Separation*, 2010. *Lecture Notes in Computer Science*, v. 6365, p. 354–361. Berlin; Heidelberg: Springer, 2010. Disponível em: https://doi.org/10.1007/978-3-642-15995-4_44. Acesso em: 5 jul. 2024.

ANDERSON, M.; LI, X.-L.; ADALI, T. Complex-valued independent vector analysis: application to multivariate Gaussian model. *Signal Processing*, v. 92, n. 8, p. 1821–1831, 2012. Disponível em: https://doi.org/10.1016/j.sigpro.2011.09.034. Acesso em: 5 jul. 2024.

BELHUMEUR, P. N.; HESPANHA, J. P.; KRIEGMAN, D. J. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, p. 711–720, 1997. Disponível em: https://doi.org/10.1109/34.598228. Acesso em: 28 fev. 2025.

BOIDIDOU, C.; PAPADOPOULOS, S.; ZAMPOGLOU, M.; APOSTOLIDIS, L.; PAPADOPOULOU, O.; KOMPATSIARIS, Y. Detection and visualization of misleading content on Twitter. *International Journal of Multimedia Information Retrieval*, v. 7, n. 1, p. 71–86, 2018. Disponível em: https://doi.org/10.1007/s13735-017-0143-x. Acesso em: 15 fev. 2025.

BOUKOUVALAS, Z. Development of ICA and IVA algorithms with application to medical image analysis. 2017. Tese (Doutorado em Filosofia) — University of Maryland, Baltimore County, Baltimore, MD, 2017. Disponível em: https://arxiv.org/abs/1801.08600. Acesso em: 3 nov. 2024.

BOUKOUVALAS, Z. Data fusion on the space of sparse positive definite matrices: an application on misinformation detection. 2023. Tese (Doutorado) — American University, Washington, DC, 2023. Disponível em: https://doi.org/10.57912/24264514. Acesso em: 3 nov. 2024.

BOUVERESSE, D. J.-R.; RUTLEDGE, D. N. Independent components analysis: theory and applications. In: RUCKEBUSCH, C. (ed.). *Resolving Spectral Mixtures*. Amsterdam: Elsevier, 2016. v. 30, p. 225–277. (Data Handling in Science and Technology). Disponível em: https://doi.org/10.1016/B978-0-444-63638-6.00007-3. Acesso em: 8 abr. 2024.

DAMASCENO, L. P.; CAVALCANTE, C. C.; ADALI, T.; BOUKOUVALAS, Z. Independent vector analysis using semi-parametric density estimation via multivariate entropy maximization. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2021. Proceedings [...]. p. 3715–3719. Disponível em: https://doi.org/10.1109/ICASSP39728.2021.9414773. Acesso em: 13 out. 2024.

DAMASCENO, L. P.; REXHEPI, E.; SHAFER, A.; WHITEHOUSE, I.; JAPKOWICZ, N.; CAVALCANTE, C. C.; CORIZZO, R.; BOUKOUVALAS, Z. Exploiting sparsity and statistical dependence in multivariate data fusion: an application to misinformation detection for high-impact events. *Machine Learning*, v. 113, p. 2183–2205, 2024. Disponível em: https://doi.org/10.1007/s10994-023-06424-8. Acesso em: 13 out. 2024.

DAMASCENO, L. P.; SHAFER, A.; JAPKOWICZ, N.; CAVALCANTE, C. C.; BOUKOUVALAS, Z. Efficient multivariate data fusion for misinformation detection during high impact events. In: PONCELET, P.; IENCO, D. (ed.). *Discovery Science*. Cham: Springer Nature Switzerland, 2022. p. 253–268. Disponível em: https://link.springer.com/chapter/10.1007/978-3-031-20865-2_17. Acesso em: 13 out. 2024.

GAO, D.; DING, J.; ZHU, C. Integrated Fisher linear discriminants: an empirical study. *Pattern Recognition*, v. 47, n. 2, p. 789–805, 2014. Disponível em: https://doi.org/10.1016/j.patcog.2013.07.021. Acesso em: 10 dez. 2024.

DHIR, C. S.; LEE, S.-Y. Discriminant independent component analysis. *IEEE Transactions on Neural Networks*, v. 22, n. 6, p. 845–857, 2011. Disponível em: https://doi.org/10.1109/TNN.2011.2122266. Acesso em: 28 abr. 2024.

DUARTE, L. T.; SUYAMA, R.; ATTUX, R. R. F.; VON ZUBEN, F. J.; ROMANO, J. M. T. Blind source separation of post-nonlinear mixtures using evolutionary computation and order statistics. In: *Independent Component Analysis and Blind Signal Separation*. Berlin; Heidelberg: Springer, 2006. v. 3889, p. 66–73. (Lecture Notes in Computer Science). Disponível em: https://doi.org/10.1007/11737475_10. Acesso em: 8 out. 2024.

FISHER, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, v. 7, n. 2, p. 179–188, 1936. Disponível em: https://doi.org/10.1111/j.1469-1809.1936.tb02137.x. Acesso em: 28 abr. 2024.

KHRAISAT, A.; GONDAL, I.; VAMPLEW, P.; KAMRUZZAMAN, J. Survey of intrusion detection systems: techniques, datasets and challenges. *Cybersecurity*, v. 2, n. 1, p. 1–22, 2019. Disponível em: https://doi.org/10.1186/s42400-019-0038-7. Acesso em: 6 jul. 2025.

KIM, T.; LEE, I.; LEE, T.-W. Independent vector analysis: definition and algorithms. In: *Asilomar Conference on Signals, Systems and Computers*, 40., 2006. Proceedings [...]. p. 1393–1396. Disponível em: https://doi.org/10.1109/ACSSC.2006.354986. Acesso em: 28 abr. 2024.

LU, J.; PLATANIOTIS, K. N.; VENETSANOPOULOS, A. N. Face recognition using LDA-based algorithms. *IEEE Transactions on Neural Networks*, v. 14, n. 1, p. 195–200, 2003. Disponível em: https://doi.org/10.1109/TNN.2002.806647. Acesso em: 6 jul. 2025.

LUO, Z. Independent vector analysis: model, applications, challenges. *Pattern Recognition*, v. 138, p. 109376, 2023. Disponível em: https://doi.org/10.1016/j.patcog.2023.109376. Acesso em: 13 out. 2024.

MAIA, M.; DAMASCENO, L.; CORIZZO, R.; CAVALCANTE, C. C.; BOUKOUVALAS,

Z. Class-constrained independent vector analysis for discriminative multimodal representation learning. In: *IEEE International Conference on Data Mining Workshops (ICDMW)*, 2025. Proceedings [...]. p. 1547–1553. Piscataway: IEEE, 2025. Disponível em: https://doi.org/10.1109/ICDMW69685.2025.00187. Acesso em: 15 dez. 2025.

MOLLAEE, M.; MOATTAR, M. H. Face recognition based on modified discriminant independent component analysis. In: *International Conference on Computer and Knowledge Engineering (ICCKE)*, 6., 2016. Proceedings [...]. p. 60–65. Disponível em: https://doi.org/10.1109/ICCKE.2016.7802116. Acesso em: 28 abr. 2024.

YE, J.; LI, Q. A two-stage linear discriminant analysis via QR-decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 27, n. 6, p. 929–941, 2005. Disponível em: https://doi.org/10.1109/TPAMI.2005.121. Acesso em: 22 set. 2024.

ZHANG, C.; RUAN, F.; YIN, L.; CHEN, X.; ZHAI, L.; LIU, F. A deep learning approach for network intrusion detection based on NSL-KDD dataset. In: *IEEE International Conference on Anti-counterfeiting, Security, and Identification (ASID)*, 13., 2019. Proceedings [...]. p. 41–45. Disponível em: https://doi.org/10.1109/ICASID.2019.8925239. Acesso em: 21 fev. 2025.