Distilling Epigenetics: Adaptive Regression Uncovers Spitzoid Tumor Biomarkers

Ilán Carretero*, Rocío del Amor*†, Germán Casabó-Vallés§, Silvia Perez-Deben‡, Andrés Mosquera-Zamudio‡, Eva García§, Carlos Monteagudo‡, Valery Naranjo*†
*Instituto Universitario de Investigación e Innovación en Tecnología Centarada en el Ser Humano, HUMAN-tech, Universitat Politècnica de València, Valencia, Spain
†Artikode Intelligence S.L., Valencia, Spain
‡INCLIVA, Universitat de València, Valencia, Spain
§EpiDisease, Valencia, Spain

***Abstract*—Accurate diagnosis of spitzoid tumors is crucial, as inappropriate treatment can have severe clinical consequences. Epigenetic biomarker signatures compatible with low-cost molecular techniques offer a promising strategy to enhance diagnostic accuracy and deepen our understanding of the disease. In this work, we introduce a novel framework for establishing a reliable and interpretable epigenetic signature derived from whole-genome bisulfite sequencing data and validated by pyrosequencing. Our approach combines advanced multivariate statistical techniques with two innovative methods. One efficiently selects an epigenetic biomarker signature from preliminary candidates, while the other adaptively addresses missing values commonly encountered in molecular assays. Together, these methods yield robust performance, high interpretability, and consistency, suggesting a promising pathway toward a clinically applicable diagnostic tool for spitzoid tumors.**

***Index Terms*—Epigenetic biomarkers, spitzoid tumors, multi-omics integration, logistic regression.**

I. INTRODUCTION

Spitzoid tumors (ST) are melanocytic neoplasms characterized by large spindle, epithelioid cellular morphology and unpredictable clinical behavior [1]. These tumors are classified into three categories: the benign form, called Spitz Nevus (SN); an intermediate category with uncertain malignant potential, known as Atypical Spitz Tumor (AST) or Spitz melanocytoma; and the malignant category, referred to as Spitz Melanoma (SM) [1]. An accurate diagnosis is crucial to prevent misclassification, leading to significant clinical consequences and inappropriate treatment [2]. Diagnostic protocols integrate histopathological examination, clinical data, and molecular analysis; however, discrepancies between morphological features and clinical outcomes, coupled with significant interobserver variability among pathologists, continue to pose challenges [3].

Although multiple studies have revealed useful genetic findings for diagnosing spitzoid tumors [4], few have examined the potential of epigenetic alterations in this context [5]. The investigation of epigenetic modifications, such as DNA methylation (DNAm) and microRNA (miRNA), has been well-documented in various cancers, including melanoma [6]. DNAm profiles are specific to each cancer type and have proven to be more stable and robust biomarkers than gene expression, thereby offering greater reliability for early diagnosis [7]. Moreover, miRNAs are valuable for analysis because many exhibit deregulation during melanoma progression [8]. Consequently, epigenetic biomarkers may provide a powerful approach for accurately stratifying spitzoid tumors, ultimately enhancing diagnostic accuracy.

Techniques such as whole-genome bisulfite sequencing (WGBS) provide single-base resolution analysis of DNA methylation by focusing on CpG sites (regions where cytosine precedes guanine) that play a critical role in gene regulation. These methods generate detailed epigenomic maps that reveal subtle changes in tumor biology. However, the datasets resulting from these techniques are highly dimensional requiring robust feature selection and interpretable algorithms to identify key methylation alterations and evaluate their diagnostic impact. To validate these computational insights, targeted methods such as pyrosequencing are commonly employed [9]. This technique precisely quantifies methylation at individual CpG sites and offers a cost-effective solution for clinical practice. Nevertheless, these approaches can occasionally yield missing values due to unspecific amplification bands or suboptimal primer hybridization, underscoring the need to establish validation frameworks that effectively manage missing values in real-world applications.

In this work, we introduce the first end-to-end framework that seamlessly integrates multi-omic data and machine learning (ML) to identify and clinically validate epigenetic biomarkers for accurately classifying spitzoid tumors. We first analyze DNA methylation and miRNA data from formalin-fixed paraffin-embedded samples to identify candidate biomarkers using the Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) [10] approach. These candidates are then refined through the proposed Nested Iterative Logistic Distillation (NILD) process. To ensure robustness in real-world scenarios, where epigenetic variable extraction may be incomplete, we utilize our Missingness-Adaptive Logistic Regression Tree (MALRT) approach that dynamically adapts to missing data while maintaining strong discriminative power and interpretability. This comprehensive framework advances the discovery of epigenetic biomarker signatures providing explainable methodolo-

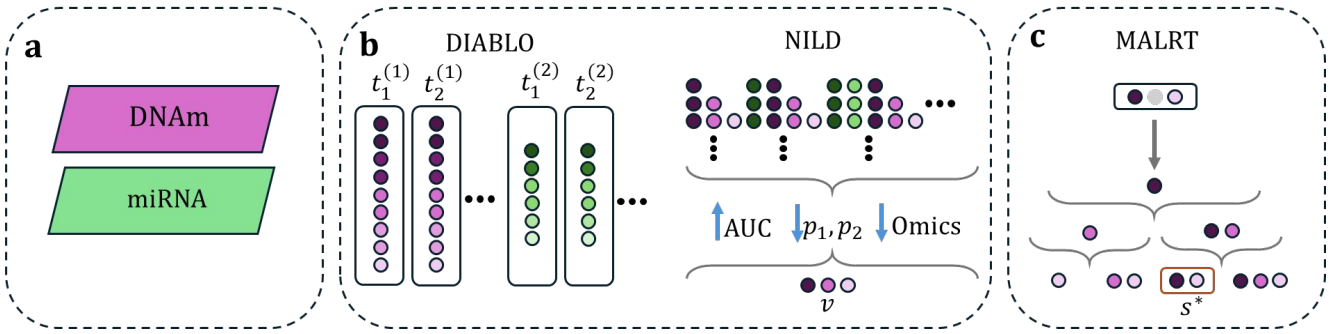

Fig. 1. **Method overview.** The developed framework comprises three main components: (a) the preprocessing of DNAm and miRNA data, (b) the selection of an epigenetic biomarker signature for classifying spizoid tumors, and (c) the application of this signature to samples in which the selected epigenetic biomarkers were obtained via pyrosequencing. Notation is detailed throughout the methodology section, and the term "Omics" denotes various omic datasets.

gies that may contribute to developing a clinically viable tool for enhancing the diagnosis of spitzoid tumors, with potential applications in other pathologies.

II. METHODOLOGY

An overview of the proposed framework is illustrated in Fig. 1. The following subsections provide detailed descriptions of its individual components.

*a) **Preprocessing miRNAs and DNA Methylation**:* Both miRNA and DNAm data are preprocessed following a similar workflow. Initially, raw FASTQ files undergo quality control (QC) with FastQC and multiQC, after which adapters and low-quality bases (Phred < 20) are removed using BBDuk, and a second QC confirms data integrity. For miRNAs, the clean reads are aligned to the human genome (GRCh38) using Bowtie v1.3.1, quantified with Rsubread and miRBase (v22), normalized via the trimmed mean M-values method from edgeR, and low-expression miRNAs are filtered out. For DNAm, bisulfite-treated reads are aligned with Bismark, and the resulting BAM files are deduplicated with Bismark and sorted with Samtools. These BAM files are then processed with methylKit to calculate base-level methylation percentages, and the data are filtered to remove CpG sites with fewer than 10 reads or with coverage above the 99.9th percentile, followed by normalization across samples for subsequent analysis.

*b) **Selecting Epigenetic Biomarkers**:* The epigenetic biomarkers with the greatest discriminatory power are selected in two phases. First, the DIABLO method is used for a preliminary selection, which is then refined using nested iterative logistic distillation.

DIABLO extends sparse generalized canonical correlation analysis to a supervised framework by incorporating class labels into the integration of multi-omics data. Let $X^{(1)} \in \mathbb{R}^{n\times p_1}$ and $X^{(2)} \in \mathbb{R}^{n\times p_2}$ represent the DNAm and miRNA data, respectively, each containing n samples and p_1 or p_2 variables. For the DNAm data, denote by $a_q^{(1)} \in \mathbb{R}^{p_1}$ the coefficient vector associated with the q-th latent component and define the corresponding latent component as $t_q^{(1)} = X^{(1)}a_q^{(1)}$, with $q \in \{1, 2, \ldots, Q\}$, where Q is the maximum number of latent components extracted from the DNAm data. Similarly, for the miRNA data, denote by $a_r^{(2)} \in \mathbb{R}^{p_2}$ the coefficient vector associated with the r-th latent component and define the latent component as $t_r^{(2)} = X^{(2)}a_r^{(2)}$, with $r \in \{1, 2, \ldots, R\}$, where R is the maximum number of latent components extracted from the miRNA data. DIABLO replaces one omics dataset with a dummy indicator matrix Y that encodes the class membership (e.g., SN or SM) and seeks these sparse coefficient vectors such that the latent components are optimized to maximize the covariance between the datasets while correlating with Y under an L_1 constraint. Furthermore, DIABLO provides a loading coefficient for each original variable, clearly indicating its importance in discriminating between classes.

To obtain a compact, optimal, and highly interpretable epigenetic signature for the classification of spitzoid tumors, we propose the NILD methodology. Initially, separated logistic regression models were fitted for DNAm and miRNA data following an additive iterative procedure: the DNAm variables $D = \{d_1, d_2, \ldots, d_{n_D}\}$ and the miRNA variables $M = \{m_1, m_2, \ldots, m_{n_M}\}$, ordered in descending order by their DIABLO loading coefficients, were sequentially incorporated into the models, thereby progressively refining performance as each additional top-ranked variable was added. Subsequently, an iterative and nested logistic regression was performed to integrate both data types, defined as

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \sum_{i=1}^{k} \beta_i\, d_i + \sum_{j=1}^{\ell} \gamma_j\, m_j,$$

where p represents the probability of the positive class, and the indices $k = 1, \ldots, n_D$ and $\ell = 1, \ldots, n_M$ denote the number of top variables selected from each omics dataset. All models were trained and validated using stratified 5-fold cross-validation, and the optimal model was chosen based on achieving the highest AUC while utilizing the fewest variables and omics techniques. Ultimately, NILD refines the initial DIABLO selection to yield a final logistic regression model based on an optimal subset of v variables.

*c) **Validating Epigenetic Biomarkers**:* To address the challenge of missing epigenetic measurements in pyrosequencing, which may preclude the direct application of the NILD-derived logistic regression model, we propose the missingness-adaptive logistic regression tree. MALRT organizes a collection of logistic regression models into a tree structure, where each node corresponds to a distinct, non-empty subset s of the v optimal variables. Thanks to the variable distillation performed by NILD, the combinatorial challenge of fitting all $2^v - 1$ possible subsets becomes more feasible. For a new sample, let $s^* \subseteq \{1, \ldots, v\}$ denote the indices of the variables with available measurements. MALRT then selects the pre-fitted logistic regression model using the greatest number of available variables. In this selected model, the log-odds are given by

$$\log\left(\frac{p}{1-p}\right) = \beta_0^{(s^*)} + \sum_{i \in s^*} \beta_i^{(s^*)} x_i.$$

This framework efficiently accommodates missing data while preserving the interpretability and predictive performance of logistic regression.

III. EXPERIMENTAL SETTING

*a) **Dataset**:* The Pathology Department at the *Hospital Clínico Universitario de Valencia*, Spain, conducted a comprehensive study of STs from 1985 to 2023. The team meticulously searched the laboratory archives using the institution's software, and each case was evaluated by experienced pathologists specialized in spitzoid tumors. A total of 46 tumor samples were included in the study, which were grouped into two batches: a first batch of 22 SN and 11 SM, and a second batch of 8 SN and 5 SM. The first batch was used for training and the second for testing. For each sample, we acquired DNAm and miRNA data. After preprocessing, DNA methylation was evaluated at 295,873 CpG sites, and expression levels were measured for 289 miRNAs. The selected epigenetic signature was confirmed by pyrosequencing on 5 SN and 7 SM samples, constituting an independent external validation cohort completely distinct from the previously analyzed samples. Data can be shared upon request and by prior agreement with INCLIVA and Universitat de València.

*b) **Processing**:* The DIABLO method was applied to the training epigenetic dataset. The analysis was limited to selecting 10 to 30 methylation positions and 1 to 5 miRNAs, since exceeding these numbers is impractical for RT-qPCR. In addition, only the first latent variable, which captures the maximum shared discriminative variation in both omics data types, was used. Experimental evidence demonstrated that this single component perfectly distinguished SN from SM in the training set, resulting in a more parsimonious model. Furthermore, the training data were standardized using the mean and standard deviation, and this normalization was subsequently applied to both the test set and the pyrosequencing validation set. The preprocessing of miRNAs and methylation positions, along with the DIABLO analysis, was performed in R 4.2.0 and bash. The main modules included Subread 2.0.3, Bismark 0.24.0, methylKit 1.20.0, and mixOmics 6.23.2.

*c) **Classification algorithms**:* The NILD and MALRT procedures, along with comparisons to other machine learning (ML) methods, were implemented in Python. Since grid search hyperparameter optimization was applied to all models, a Docker image with Python 3.8 was used to ensure reproducibility. The main modules included numpy 1.22.2, pandas 1.5.2, scikit-learn 0.24.2, and xgboost 1.7.1.

IV. RESULTS

*a) **Epigenetic Biomarker Evaluation**:* The results obtained using DIABLO are presented in Table I. These findings reveal that 20 methylation positions and only 1 miRNA were selected from the complete set of DNAm positions and miRNAs. The coefficients associated with each methylation position underscore their importance in shaping the latent variable to which they contribute.

Given the biomarkers initially selected by DIABLO, the NILD method can be applied at the multi-omics level. The results are presented in Fig. 2. As observed, the AUC improves with the selection of a greater number of methylation positions. Moreover, adding the miRNA does not significantly enhance the performance of the logistic regression models fitted using different methylation positions. As a result, the most efficient model, achieving the best performance, consists of 18 methylation positions. The coefficients of the fitted logistic regression are shown in (1).

The classification metrics were analyzed on the WGBS test set for the best logistic regression model. Alternative models, including k-nearest neighbors (KNN), random forest (RF), and extreme gradient boosting (XGB), were also fitted using the 18 selected methylation positions. Table II shows that logistic regression achieved the highest performance in terms of accuracy (ACC), sensitivity (SEN), specificity (SPE), positive predictive value (PPV), negative predictive value

TABLE I
MULTI-OMICS BIOMARKERS SELECTED BY THE DIABLO METHOD.

ID	DNAm position	Coefficient	miRNA	Coefficient
A	chr21:42976620	-0.56	hsa-miR-324-5p	1.00
B	chr3:77806280	-0.40		
C	chr15:91397225	-0.30		
D	chr13:23584928	0.28		
E	chr17:53494610	-0.27		
F	chr12:91381168	0.25		
G	chr15:91417325	-0.24		
H	chr4:87089695	0.21		
I	chr3:189791255	0.19		
J	chr13:75454496	0.13		
K	chr22:32021158	-0.11		
L	chr7:81411051	0.11		
M	chr1:107688094	0.10		
N	KI270442.1:77402	-0.10		
O	chr15:93698686	-0.10		
P	chr3:151933366	0.08		
Q	chr8:48721155	-0.08		
R	chr1:87169347	-0.05		
S	chr12:88144889	0.03		
T	chr2:177317964	0.02		

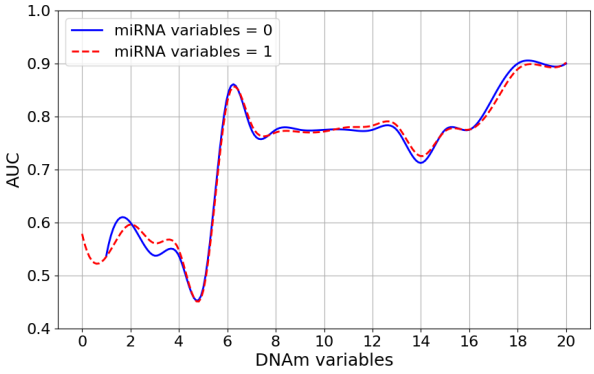

Fig. 2. Multi-omics nested iterative logistic distillation results

(NPV), F1-score (F1) and area under the receiver operating curve (AUC), demonstrating its strong generalization ability on WGBS-derived methylation data.

$$
\begin{aligned}
P_{melanoma}(\mathbf{ID_{DNAm}}) &= \frac{1}{1+e^{-(\beta_0+\boldsymbol{\beta}\cdot\mathbf{ID_{DNAm}})}} \\
\beta_0 &= -0.7 \\
\boldsymbol{\beta}\cdot\mathbf{ID_{DNAm\_pos}} &= -0.89A - 0.86B - 1.06C \\
&= +0.36D - 1.40E + 1.73F \\
&= -1.37G + 0.74H + 1.65I \\
&= +1.49J - 0.65K + 1.13L \\
&= +0.99M - 0.31N - 0.64O \\
&= +0.12P - 1.25Q - 0.19R
\end{aligned}
\tag{1}
$$

TABLE II
LOGISTIC REGRESSION RESULTS ON THE WGBS TEST SET AND COMPARISON WITH ML MODELS.

MODEL	ACC	SEN	SPE	PPV	NPV	F1	AUC
KNN	0.77	0.40	**1.00**	**1.00**	0.72	0.57	0.70
RF	0.77	0.40	**1.00**	**1.00**	0.73	0.58	0.72
XGB	0.77	0.60	0.88	0,75	0.78	0.67	0.74
LR	**0.92**	**0.80**	**1.00**	**1.00**	**0.89**	**0.89**	**0.90**

*b) **Validation Results**:* For the external validation cohort, pyrosequencing was employed to extract values from the 18 key methylation positions identified as crucial for distinguishing between SN and SM. Due to issues with the designed primer pairs, methylation values for positions with ID I, K, L, M, N, Q, and R could not be obtained for any sample. Table III reports the number of methylation positions successfully extracted for each sample.

Table IV presents the performance of the MALRT methodology, trained on WGBS data, when applied to an external validation cohort obtained via pyrosequencing (PYRO). In addition, the logistic regression models derived from the PYRO samples were also evaluated on the WGBS test set. For comparison with alternative ML approaches, these models were trained using the available variables from each pyrosequencing validation sample, yielding results for both WGBS and PYRO data. The performance metrics reveal that logistic

TABLE III
CPG SITES EXTRACTED PER SAMPLE FOR THE INDEPENDENT VALIDATION COHORT VIA PYROSEQUENCING

DNAm positions	Samples	DNAm positions IDs
11	8	A, B, C, D, E, F G, H, J, O, P
10	2	A, C, D, E, F, G, H, J, O, P
9	1	A, C, D, E, F, G, H, J, O
2	1	F, G

regression (through MALRT) not only achieves superior ACC, F1 and AUC, but also exhibits the smallest relative percentage performance degradation (ΔPct) when transitioning from the WGBS test set to the pyrosequencing validation cohort.

TABLE IV
RESULTS FOR AVAILABLE VARIABLES IN PYROSEQUENCING SAMPLES IN WGBS, PYRO, AND THE PERCENTAGE CHANGE IN METRICS USING THE MALRT METHODOLOGY AND ML MODELS.

MODEL	ACC	SEN	SPE	PPV	NPV	F1	AUC
KNN_{WGBS}	0.79	0.47	**1.00**	**1.00**	0.75	0.62	0.73
KNN_{PYRO}	0.58	0.29	**1.00**	**1.00**	0.50	0.44	0.64
$KNN_{\Delta Pct}$	$\downarrow$ 26.6%	$\downarrow$ 38.3%	0%	0%	$\downarrow$ 33.3%	$\downarrow$ 29.0%	$\downarrow$ 12.3%
RF_{WGBS}	0.78	0.57	0.92	0.83	0.77	0.67	0.74
RF_{PYRO}	0.58	0.29	**1.00**	**1.00**	0.50	0.44	0.64
$RF_{\Delta Pct}$	$\downarrow$ 25.6%	$\downarrow$ 49.1%	$\uparrow$ 8.7%	$\uparrow$ 20.5%	$\downarrow$ 35.1%	$\downarrow$ 34.3%	$\downarrow$ 13.5%
XGB_{WGBS}	0.78	0.60	0.90	0.79	0.78	0.68	0.75
XGB_{PYRO}	0.50	**0.71**	0.20	0.56	0.33	0.63	0.46
$XGB_{\Delta Pct}$	$\downarrow$ 35.9%	$\uparrow$ **18.3%**	$\downarrow$ 77.8%	$\downarrow$ 29.1%	$\downarrow$ 57.7%	$\downarrow$ 7.4%	$\downarrow$ 38.7%
LR_{WGBS}	**0.81**	**0.80**	0.81	0.74	**0.87**	**0.77**	**0.81**
LR_{PYRO}	**0.75**	0.57	**1.00**	**1.00**	**0.63**	**0.73**	**0.79**
$LR_{\Delta Pct}$	$\downarrow$ **7.4%**	$\downarrow$ 28.8%	$\uparrow$ **23.5%**	$\uparrow$ **35.1%**	$\downarrow$ **27.6%**	$\downarrow$ **5.2%**	$\downarrow$ **2.5%**

*c) **Biomarker Coefficient Analysis**:* Both the optimal logistic regression model and the MALRT are inherently interpretable, allowing for a transparent understanding of how each feature influences the decision-making process. However, logistic regression models fitted with different variables may yield varying interpretations for each variable. Fig. 3 compares the coefficients obtained from the optimal logistic regression model and those derived using the MALRT method. For MALRT, the reported coefficients represent the mean value across all fitted models and their corresponding standard deviations.

Analyzing the coefficient estimates offers valuable insights. First, the vast majority of coefficients retain their sign whether they are derived from the optimal logistic regression model or from the MALRT submodels. This consistency is crucial, as it confirms that the qualitative interpretation of each methylation position's contribution remains intact. In logistic regression, a positive coefficient implies that a one-unit increase in the predictor is associated with an increase in the log odds of the outcome (i.e., the natural logarithm of the odds of being classified as SM rather than SN), whereas a negative coefficient indicates the opposite.

Moreover, although the MALRT method generally produces coefficients with higher magnitudes, suggesting a stronger effect for individual variables, the relative proportional relationships among the coefficients are preserved. This maintenance of proportionality ensures that the inferred relative importance

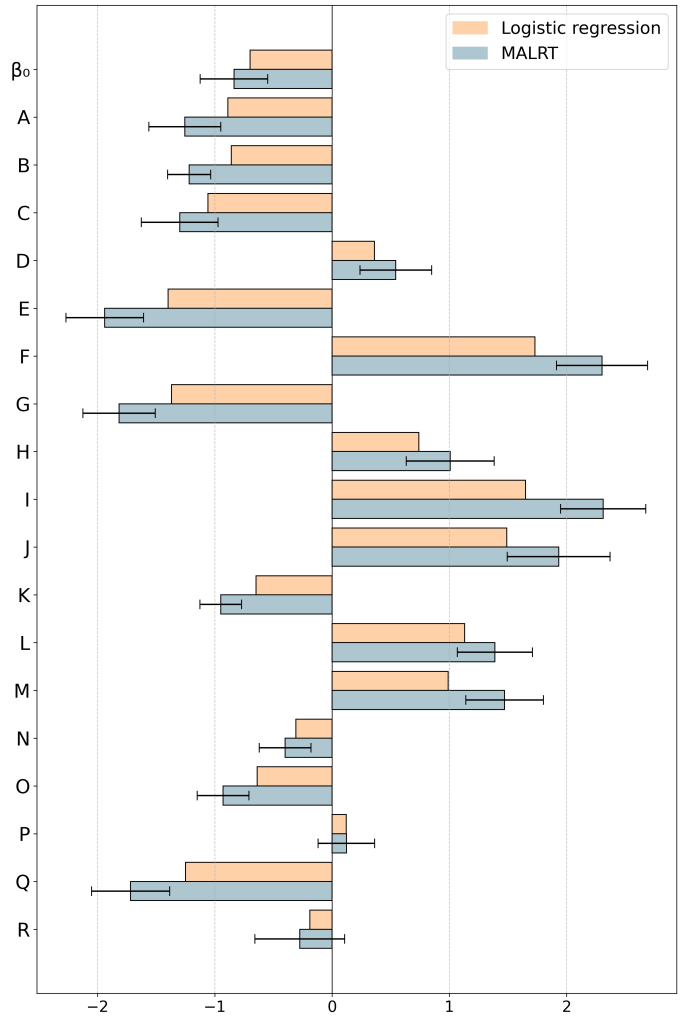

Fig. 3. Comparison of intercept and coefficient values from the epigenetic biomarker signature and MALRT approach.

of each variable in classification remains consistent.

Finally, the variability in the coefficients obtained from the MALRT is relatively low, indicating that these estimates are stable across different submodels. This stability, together with the preserved sign and proportional relationships, highlights the robustness of the MALRT methodology for managing missing data while maintaining both robust performance and interpretability.

V. CONCLUSION

Developing an epigenetic biomarker signature to classify spitzoid tumors using cost-effective molecular techniques holds significant clinical value. In this work, we propose an end-to-end methodology that combines epigenomic multi-omics data with interpretable bioinformatics and machine learning techniques. Our approach yields a robust epigenetic signature with direct clinical applicability while providing an explainable framework that clarifies the contribution of each individual biomarker to patient diagnosis. Using DIABLO along with our NILD and MALRT methods, we derived an epigenetic signature from WGBS data and validated it with pyrosequencing data. Our findings reveal that miRNAs play only a minor role in classifying spitzoid tumors compared to DNA methylation, and that MALRT preserves both performance and interpretability with minimal degradation. The primary limitations of this study are the small sample size and the occurrence of missing methylation values in the pyrosequencing data, challenges inherent to the tumors under investigation and the molecular techniques employed. Overall, this work paves the way for further analysis of the discriminative power of DNAm and the potential of epigenetic biomarkers to stratify atypical spitzoid tumors. Moreover, the proposed framework could be readily extended to other pathologies amplifying its impact on precision medicine and epigenetics.

ACKNOWLEDGEMENT

The authors thank the Central Unit for Research in Medicine (UCIM) from University of Valencia for small RNA library preparation and sequencing in their Illumina NextSeqTM platform and Epigenomics Unit from Health Research Institute La Fe for pyrosequencing analyses in their PyroMark Q24 System version 2.0.6 (Qiagen). This work has received funding from the Spanish Ministry of Economy and Competitiveness through project PID2022-140189OB-C21 (ASSIST) and has also been supported by the Generalitat Valenciana via projects INNEST/2021/321 (SAMUEL), CIPROM/2022/20 (PROMETEO) and SA2024-26 (IMAGEN).

REFERENCES

[1] R. L. Barnhill, "The spitzoid lesion: rethinking spitz tumors, atypical variants,'spitzoid melanoma'and risk assessment," *Modern pathology*, vol. 19, pp. S21–S33, 2006.

[2] D. C. Orchard, J. P. Dowling, and J. W. Kelly, "Spitz naevi misdiagnosed histologically as melanoma: prevalence and clinical profile," *Australasian journal of dermatology*, vol. 38, no. 1, pp. 12–14, 1997.

[3] B. Bhoyrul, G. Brent, F. Elliott, J. McLorinan, A. Wilson, H. Peach, B. Mathew, and A. Mitra, "Pathological review of primary cutaneous malignant melanoma by a specialist skin cancer multidisciplinary team improves patient care in the uk," *Journal of Clinical Pathology*, vol. 72, no. 7, pp. 482–486, 2019.

[4] V. L. Quan, B. Zhang, Y. Zhang, L. S. Mohan, K. Shi, A. Wagner, L. Kruse, T. Taxter, N. Beaubier, K. White, *et al.*, "Integrating next-generation sequencing with morphology improves prognostic and biologic classification of spitz neoplasms," *Journal of Investigative Dermatology*, vol. 140, no. 8, pp. 1599–1608, 2020.

[5] R. Del Amor, A. Colomer, C. Monteagudo, M. J. Garzón, J. L. García-Giménez, and V. Naranjo, "A deep embedded framework for spitzoid neoplasm classification using dna methylation data," in *2021 29th European Signal Processing Conference (EUSIPCO)*, pp. 1271–1275, IEEE, 2021.

[6] G. Micevic, N. Theodosakis, and M. Bosenberg, "Aberrant dna methylation in melanoma: biomarker and therapeutic opportunities," *Clinical epigenetics*, vol. 9, pp. 1–15, 2017.

[7] S. Saghafinia, M. Mina, N. Riggi, D. Hanahan, and G. Ciriello, "Pan-cancer landscape of aberrant dna methylation across human tumors," *Cell reports*, vol. 25, no. 4, pp. 1066–1080, 2018.

[8] A. Thyagarajan, K. Y. Tsai, and R. P. Sahu, "Microrna heterogeneity in melanoma progression," in *Seminars in cancer biology*, vol. 59, pp. 208–220, Elsevier, 2019.

[9] J. Tost and I. G. Gut, "Dna methylation analysis by pyrosequencing," *Nature protocols*, vol. 2, no. 9, pp. 2265–2275, 2007.

[10] A. Singh, C. P. Shannon, B. Gautier, F. Rohart, M. Vacher, S. J. Tebbutt, and K.-A. Lê Cao, "Diablo: an integrative approach for identifying key molecular drivers from multi-omics assays," *Bioinformatics*, vol. 35, no. 17, pp. 3055–3062, 2019.