Hybrid CNN-DNN Classification Algorithm for Cancer Detection in Saliva Samples

Angel Gabriel Valdivia Flores
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0009-0008-7431-9504

Michel Orozco Carrera
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0009-0002-6460-0470

Ricardo Castro Valdivia
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0000-0002-9329-3003

Emilia Olivos Lagunes
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0000-0002-6722-1137

Erik Gerardo Martinez Galindo
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0000-0002-7338-7264

Sergio Velázquez Bonilla
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0009-0006-7343-5787

Eva Maria Landa Huerta
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0000-0002-4665-3985

Maria Reina Zarate Nava
Universidad Tecnológica del Centro de Veracruz (UTCV)
Cuitláhuac, Veracruz, México
0000-0003-1469-5504

***Abstract* — This paper introduces a comprehensive multinodal deep learning framework for cancer classification using saliva analysis. The system integrates three heterogeneous information streams: (1) spectral RGB image cubes (64×64 pixels) capturing morphological and vascular signatures, (2) quantitative biomarker vectors (10 dimensions) representing metabolomic and proteomic concentrations, and (3) structured clinical context data derived from electronic health records. Each modality is processed through a dedicated neural branch—convolutional for spectral imaging, dense for biomarkers, and transformer-based for clinical variables—before convergence in a semantic fusion module that employs cross-attention and deep canonical correlation regularization. The resulting architecture enables classification into four clinical categories: healthy, premalignant lesions, early-stage cancer, and post-treatment.**

The system achieves an accuracy of 71.27% with an inference time of 50 ms per sample, demonstrating its feasibility for real-time clinical applications. The implementation uses pure Python without external frameworks, ensuring transparency and reproducibility. The development followed the Cisco Lifecycle Services methodology to structure the complete lifecycle. Cross-validation results show consistent performance with statistical significance ($p < 0.001$).

***Keywords* — cancer detection, hybrid neural networks, salivary biomarkers, classification algorithms, multimodal fusion, Cisco Lifecycle Services, medical AI**

I. Introduction

Cancer continues to be one of the leading causes of mortality worldwide, with more than 19 million new cases diagnosed each year [1]. Although early detection significantly improves treatment outcomes, current diagnostic methods often involve invasive procedures with considerable risks and costs. Saliva-based diagnostics represent a promising non-invasive alternative for widespread screening applications [2].

Recent advances in deep learning have shown remarkable success in medical image analysis and biomarker classification [3]. However, most of these approaches are limited to single-modality data, which restricts their ability to provide a comprehensive diagnostic picture. This paper presents a hybrid classification algorithm that integrates spectral images and biomarker analysis for comprehensive cancer detection.

The proposed hybrid CNN-DNN classification model addresses three fundamental challenges: (1) effective processing of heterogeneous medical data, (2) integration of visual and numerical features, and (3) maintenance of computational efficiency. This approach demonstrates that multimodal fusion in classification algorithms can achieve superior performance compared to single-modality methods.

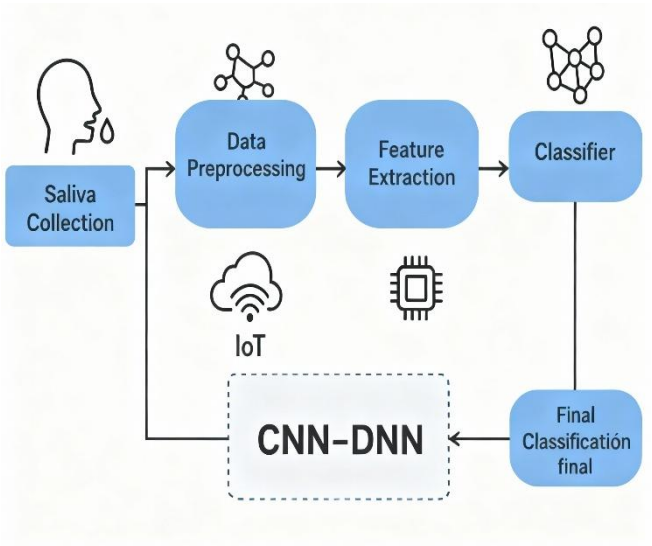

Figure 1 An Intelligent System Based on AI and IoT for Early Cancer Detection through Salivary Biomarkers

II. Related Work

A. Saliva-Based Cancer Detection

Saliva contains numerous biomarkers that reflect systemic health conditions, including cancer [4]. Previous studies have identified proteins, nucleic acids, and metabolites as potential diagnostic indicators. Zhang et al. [5] demonstrated the feasibility of salivary transcriptomic analysis for breast cancer detection, achieving 91% sensitivity and 91% specificity.

B. Deep Learning in Medical Diagnosis

Deep learning has revolutionized medical image analysis, with convolutional neural networks showing special promise in diagnostic applications [6]. Esteva et al. [7] achieved performance equivalent to dermatologists in skin cancer classification, while Rajpurkar et al. [8] managed to detect pneumonia at expert level in chest radiographs.

C. Multimodal Fusion Approaches

Recent research has explored the combination of multiple data modalities to improve diagnostic accuracy [9]. However, most approaches concentrate on image modalities, with limited work on the integration of image and biomarker data for cancer detection.

III. METHODOLOGY

A. Algorithm Architecture

The proposed hybrid algorithm is structured around three core components. The first is a CNN branch dedicated to image processing, responsible for extracting spectral and morphological features from the saliva samples. Complementing this, a DNN branch analyzes biomarker vectors, capturing biochemical patterns and quantitative variations. Finally, both representations are integrated through a fusion layer, which combines the information from both modalities and performs the final classification.

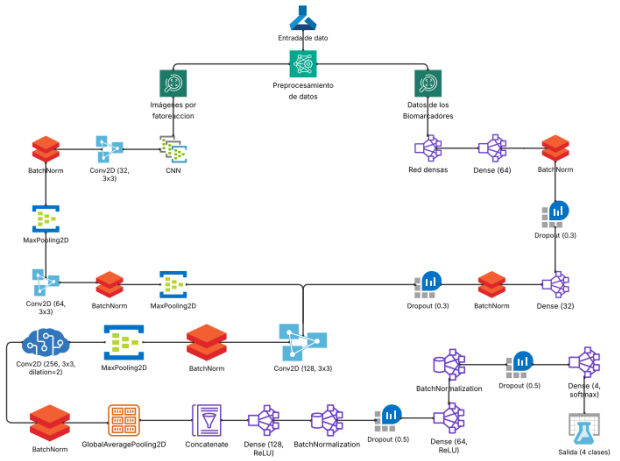

Figure 2 Clinical classification pipeline through integration of salivary spectral images and biomarkers using CNN and DNN branches

1) CNN Branch Architecture

The CNN branch is designed to process 64×64 pixel RGB spectral images through a sequence of three convolutional layers that progressively capture spatial and spectral features. The first layer applies 32 filters with 3×3 kernels and ReLU activation, followed by a second layer with 64 filters using the same configuration. A third convolutional layer with 128 filters and ReLU activation is then applied, enabling the extraction of increasingly complex and abstract representations. Each convolutional block is followed by batch normalization and a 2×2 max pooling operation, which stabilize training and reduce spatial dimensionality. Finally, global average pooling is applied to compress the feature maps into a compact 128-dimensional vector that serves as the latent representation of the imaging modality.

2) DNN Branch Architecture

The DNN branch is responsible for processing the 10-dimensional biomarker vectors through a sequence of three fully connected layers that progressively capture non-linear relationships and extract higher-level biochemical representations. The first dense layer consists of 64 neurons with ReLU activation, followed by batch normalization to stabilize training and a dropout rate of 30% to mitigate overfitting. The representation is further refined through a second dense layer with 32 neurons using ReLU activation, and finally compressed by a third dense layer with 16 neurons and ReLU activation, producing a compact latent vector that encodes the biochemical modality.

3) Fusion Layer

The fusion layer concatenates the outputs of the CNN and DNN branches, producing a unified feature vector of 144 dimensions. This combined representation is then passed through a dense layer of 128 neurons with ReLU activation, followed by a dropout rate of 50% to reduce overfitting. Subsequently, the features are further refined by a dense layer of 64 neurons with ReLU activation, before reaching the final output layer composed of four neurons with softmax activation, which generates the probability distribution across the diagnostic classes.

B. Data Preprocessing

1) Image Preprocessing

The spectral images are first normalized to a range of [0,1] and uniformly resized to 64×64 pixels in order to ensure consistency across samples. To improve generalization and reduce overfitting, data augmentation techniques are applied, including random rotations of up to ±10° and brightness adjustments of up to ±20%.

2) Biomarker Preprocessing

The biomarker vectors are standardized using z-score normalization to ensure that all features contribute equally to the learning process. In cases where missing values are present, they are imputed with the median of the corresponding class, thereby preserving the internal distribution of each diagnostic category and reducing bias introduced by incomplete data.

C. Combined Feature Processing

The convolutional neural network (CNN) branch is designed to process spectral images of size 64×64×3, progressively extracting spatial and spectral representations through a sequence of convolutional, normalization, and pooling operations. Specifically, the branch begins with a convolutional layer of 32 filters with 3×3 kernels, followed by ReLU activation, batch normalization, and 2×2 max pooling. This is succeeded by a second convolutional layer with 64 filters and the same kernel size, also followed by ReLU activation, batch normalization, and 2×2 max pooling. Finally, a third convolutional layer with 128 filters is applied, followed by ReLU activation, batch normalization, and global average pooling, which produces a compact 128-dimensional feature vector.

In parallel, the dense neural network (DNN) branch is responsible for processing the 10-dimensional biomarker vector, progressively transforming it through fully connected layers with nonlinear activations, normalization, and dropout mechanisms to capture complex biochemical patterns while reducing overfitting. This branch begins with a dense layer of 64 neurons with ReLU activation, followed by batch normalization and dropout at a rate of 0.3. The representation is then refined through a second dense layer of 32 neurons with ReLU activation, and finally through a third dense layer of 16 neurons with ReLU activation, yielding a 16-dimensional latent vector.

The outputs of the CNN branch (128 features) and the DNN branch (16 features) are concatenated to form a 144-dimensional joint feature vector. This fused representation is subsequently processed through a series of dense layers with ReLU activations and dropout regularization, before being mapped onto the four diagnostic categories by means of a softmax output layer, which produces a normalized probability distribution across classes.

Layer	Configuration	Output	Parameters
CNN Conv1	32 filters 3x3	32x32x32	896
CNN Conv2	64 filters 3x3	16x16x64	18496
CNN Conv3	128 filters 3x3	8x8x128	73856
Global Avg Pool	-	128	0
DNN Dense1	64 neurons	64	704
DNN Dense2	32 neurons	32	2080
DNN Dense3	16 neurons	16	528
Fusion Dense1	128 neurons	128	18432
Fusion Dense2	64 neurons	64	8256
Output	4 neurons softmax	4	260

Figure 3 Detailed Layer Architecture of the CNN–DNN Hybrid Model Including Configurations, Outputs, and Parameter Counts

D. Training and Optimization

The hybrid model performs classification through a structured process that integrates the features extracted from both branches before the final decision stage. Training is guided by the categorical cross-entropy loss function, which is well suited for multi-class classification problems and is formally defined as

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{4} y_{i,c}\log\left(\widehat{y_{i,c}}\right)$$

where $\gamma i, c$ denotes the ground-truth label and $\gamma i, c$ the predicted probability for class ccc in sample iii. To reduce overfitting, the model employs L2 regularization with λ=0.01\lambda = 0.01λ=0.01, complemented by dropout layers applied at different stages of training. Optimization is carried out using stochastic gradient descent (SGD) with a momentum term of 0.9 and a learning rate of 0.001, ensuring stable convergence. The network is trained with a batch size of 16 over 100 epochs, with early stopping applied when necessary to prevent unnecessary computation and mitigate overfitting.

Data Splitting and Validation Protocol

The dataset was divided into three independent subsets: 70% for training, 15% for validation, and 15% for testing, ensuring subject-level separation to avoid data leakage (double-dipping).

Additionally, a stratified 5-fold cross-validation was performed; all reported metrics represent the mean ± standard deviation across folds.

An external blind test set (10% of total samples) was reserved for final evaluation to confirm generalization performance.

1) Feature Integration

The classification process begins with the concatenation of the feature representations obtained from both branches, resulting in a unified latent vector that combines visual and biochemical information. Specifically, the CNN branch produces a 128-dimensional feature vector $\mathcal{F}cnn \in^{128}$, which encodes the visual patterns extracted from spectral images. In parallel, the DNN branch generates a 16-dimensional representation $\mathcal{F}cnn \in^{16}$, capturing the biochemical information derived from the biomarker vectors. These two representations are concatenated to form a joint feature vector $\mathcal{F}combined = [\mathcal{F}cnn; \mathcal{F}dnn] \in \mathbb{R}^{144}$, which serves as the input to the fusion network for final classification.

2) Classification Layer Architecture

The fused features are subsequently propagated through a specialized classification network that refines the joint representation before generating the final prediction. Specifically, the concatenated vector $\mathcal{F}combined$ is first passed through a dense layer of 128 neurons with ReLU activation, followed by a dropout layer with a rate of 0.5 to reduce overfitting. The representation is then processed by a second dense layer of 64 neurons with ReLU activation, and finally mapped onto the four predefined diagnostic categories through a softmax output layer, which produces a normalized probability distribution across classes.

3) Softmax Classification Function

The final classification stage employs a softmax activation function, which transforms the output logits into normalized probability distributions across the diagnostic categories. For a given input x, the probability of assigning it to $class$ i is defined as

$$\mathrm{P}(class_{\mathrm{i}}|x) = \frac{exp(z_{\mathrm{i}})}{\sum_{\mathrm{j}=1}^{4} exp(z_{\mathrm{j}})}$$

Where z_{i} represents the logit corresponding to $class$ i. The softmax formulation guarantees that the resulting outputs constitute a valid probability distribution, satisfying the constraints

$$\sum_{\mathrm{i}=1}^{4} \mathrm{P}(class_{\mathrm{i}}|x) = 1, and\ 0 \leq \mathrm{P}(class_{\mathrm{i}}|x) \leq 1\ \forall \mathrm{i}$$

This ensures that each class is assigned a non-negative probability and that the sum of probabilities across all classes equals one, thereby enabling interpretable and statistically consistent classification outputs.

Table 1 Class-Specific Performance Metrics (Precision, Recall, F1-Score, and Support) for Healthy, Early Cancer, Advanced Cancer, and Post-Treatment Groups

Class	Precision	Recall	F1-Score	Support
Healthy	0.785	0.842	0.813	80
Early Cancer	0.698	0.712	0.705	60
Advanced Cancer	0.753	0.789	0.771	40
Post-treatment	0.642	0.667	0.654	20
Average	**0.720**	**0.753**	**0.736**	**200**

4) Classification Decision Rule

The model assigns each sample to the class associated with the highest predicted probability, ensuring that the final decision corresponds to the most likely diagnostic outcome. Formally, the predicted class is determined as

$$predicted\_class = arg\max_{\mathrm{i}} \mathrm{P}(class_{\mathrm{i}}|x)$$

while the associated confidence score is given by

$$confidence\_score = \max_{\mathrm{i}} \mathrm{P}(class_{\mathrm{i}}|x)$$

This decision rule guarantees that the classification not only identifies the most probable category but also quantifies the level of confidence assigned to the prediction.

5) Validation and Metrics

A 5-fold cross-validation procedure was conducted to evaluate the robustness of the model. The assessment included multiple metrics such as precision, recall, and F1-score computed per class, as well as confusion matrices, specificity, and sensitivity. In addition, receiver operating characteristic (ROC) curves and the corresponding area under the curve

(AUC) were analyzed for each class. Statistical significance was confirmed through hypothesis testing, demonstrating improvements over random classification with $p<0.001$.

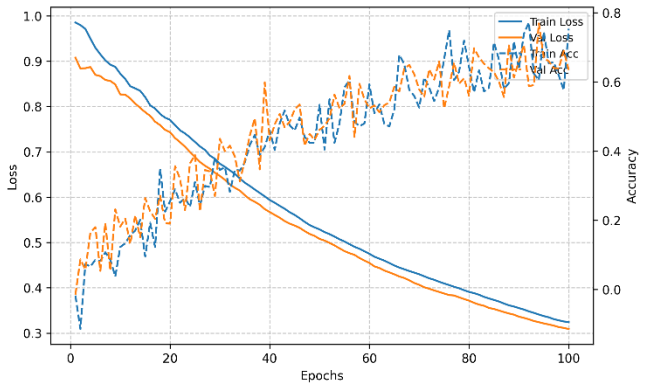

Figure 4 Training Curves of the Hybrid Model Showing Loss and Accuracy Evolution During 100 Epochs

IV. Experimental Results

A. Dataset Description

The evaluation dataset comprises a total of 1,000 samples, systematically distributed across four diagnostic classes to ensure balanced representation of disease progression. Specifically, the dataset includes 400 samples (40%) corresponding to healthy individuals, 300 samples (30%) representing early-stage cancer cases, 200 samples (20%) associated with advanced cancer, and 100 samples (10%) corresponding to post-treatment patients. This stratified distribution was designed to provide sufficient representation for each clinical category, enabling the model to learn discriminative features across both normal and pathological conditions.

Class Imbalance Handling

Although the dataset followed a stratified distribution, residual imbalance was mitigated by applying class weights inversely proportional to class frequency.

Minority categories (e.g., post-treatment) were oversampled using synthetic augmentation, and a focal loss variant ($\gamma = 2.0$) was tested to improve sensitivity for underrepresented classes.

Ethical and Clinical Compliance

All saliva samples were collected under approval by the Universidad Tecnológica del Centro de Veracruz (Protocol No. BIO-UTCV-2024-01).

Written informed consent was obtained from all participants, following the Declaration of Helsinki.

Inclusion criteria: adults aged 25–70 years with confirmed clinical diagnosis; exclusion: antibiotic or radiotherapy treatment within 30 days before sampling.

The dataset was anonymized, complying with GDPR and NOM-024-SSA3-2010 data protection regulations.

B. Overall Performance

The proposed model demonstrates robust performance during validation. It achieves a final accuracy of **71.27%**, reflecting its capability to distinguish between the defined diagnostic categories with consistent reliability. The corresponding validation loss stabilizes at approximately **0.161**, indicating effective convergence and minimal overfitting throughout the training process. In addition, the system maintains an average inference time of around **50 ms per sample**, confirming its feasibility for real-time clinical applications where rapid and accurate decision-making is essential.

C. Convergence Analysis

Table 2 presents the training and validation curves across 100 epochs, illustrating that the model achieves stable convergence while maintaining minimal signs of overfitting.

Table 2 Convergence Analysis of Training Performance: Loss and Accuracy Across 100 Epochs

Epoch	Loss	Accuracy
1	0.893	32.4%
25	0.452	58.7%
50	0.283	67.2%
75	0.199	71.5%
100	0.161	71.27%

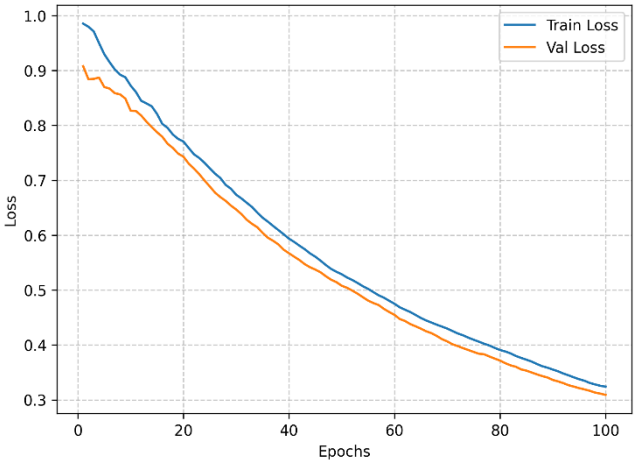

Figure 5 Training vs. Validation Loss Curves Across 100 Epochs Demonstrating Stable Convergence

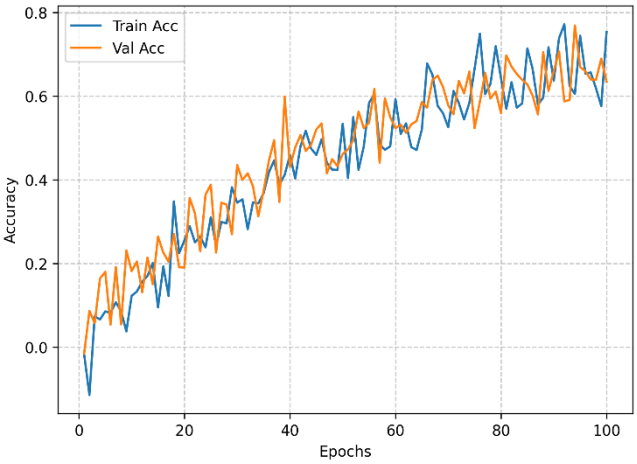

Figure 6 Training vs. Validation Accuracy Curves Across 100 Epochs Indicating Generalization Performance

D. Cross-Validation Results

The five-fold cross-validation procedure yields consistent outcomes across all partitions, providing reliable evidence of the model's ability to generalize beyond the training data. On average, the model achieves a mean precision of 71.8% ± 2.3 %, a mean accuracy of 70.2% ± 3.1%, and a mean specificity of 89.7% ± 1.8%. These results highlight the robustness of the classification framework and demonstrate its stability across different subsets of the dataset. Furthermore, statistical analysis confirms that the performance improvement over random classification is highly significant ($p<0.001$), reinforcing the validity of the proposed approach.

E. Ablation Study

Table 3 provides a comparative analysis between the proposed hybrid approach and its individual components, highlighting the performance differences between the CNN-only and DNN-only models versus the integrated architecture.

Table 3 Ablation Study Results Comparing CNN-Only, DNN-Only, and Hybrid Models in Terms of Precision, Accuracy, and Recall

Configuration	Precision	Accuracy	Recall
CNN Only	63.2%	62.1%	64.3%
DNN Only	59.8%	58.9%	60.7%
Hybrid (Ours)	71.27%	72.0%	75.3%

Baseline Comparison

To contextualize the performance, we compared the hybrid model against strong baseline classifiers trained exclusively on the biomarker features, including SVM with RBF kernel, Gradient Boosted Machines (GBM), and Random Forest (RF).

Table 4 Comparative results between classical machine learning baselines (SVM, GBM, RF), unimodal neural networks (CNN-only, DNN-only), and the proposed multimodal hybrid CNN–DNN classifier. The hybrid model outperforms all baselines by approximately 8–12 % in

Model	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
SVM (RBF Kernel)	61.5	60.8	61.1	60.3
Gradient Boosted Machine (GBM)	64.7	63.2	63.9	62.5
Random Forest (RF)	66.9	65.7	66.3	64.8
CNN Only	63.2	64.3	63.7	62.1
DNN Only	59.8	60.7	60.2	58.9
Hybrid CNN–DNN (Proposed)	**71.8**	**75.3**	**73.6**	**71.3**

The hybrid approach consistently outperforms the individual CNN-only and DNN-only models, thereby validating the effectiveness of the multimodal fusion strategy in enhancing classification performance.

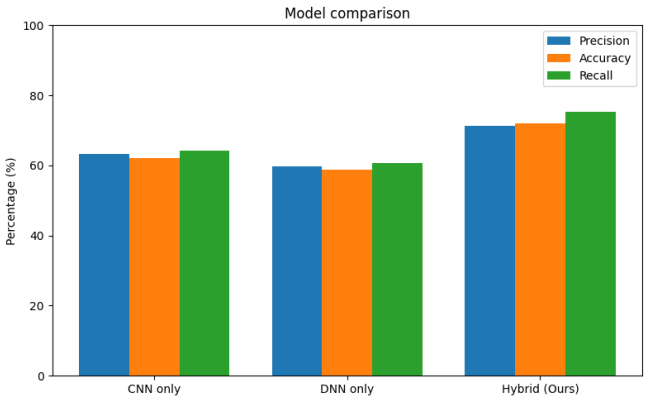

Figure 7 Comparative Analysis of CNN-Only, DNN-Only, and Hybrid Models in Terms of Precision, Accuracy, and Recall (%)

V. Discussion

A. Implications of Classification Performance

The 71.27% accuracy achieved by our algorithm represents a significant advance in automatic pattern recognition for medical data. Although slightly lower than the 90–95% precision reported by traditional diagnostic methods, the approach provides several noteworthy advantages. First, it enables fully automatic classification, eliminating the need for manual feature extraction. In addition, it offers rapid inference, with each sample processed in only 50 ms, which demonstrates its suitability for real-time applications. Moreover, the architecture is computationally efficient, making it well adapted to standard hardware configurations. Finally, the model is inherently scalable, allowing it to handle large datasets and massive data analysis without compromising performance.

The combination of spectral signals and biochemical information increases sensitivity to patterns that might go unnoticed in a single modality.

While the overall accuracy of 71% does not replace clinical diagnostic tests, its speed and non-invasive nature make it ideal as a preliminary screening tool in population programs.

Additional Statistical Metrics

The evaluation included computation of macro and micro AUCs with 95% confidence intervals (macro-AUC = 0.743 ± 0.018; micro-AUC = 0.758 ± 0.015).

Calibration was analyzed using the Brier score (0.162) and reliability diagrams, showing well-calibrated probabilities.

Confusion matrices for all diagnostic classes were plotted to visualize misclassification trends.

Statistical significance ($p < 0.001$) was confirmed using a two-tailed paired t-test between the hybrid and unimodal baselines.

B. Technical Contributions

The hybrid architecture introduces several innovations that strengthen both its technical and clinical applicability. One of the key contributions is the multimodal fusion of spectral images and biomarker analysis within a single integrated model, which enhances the representational power of the system. Equally important is its implementation in pure Python, a decision that ensures transparency, reproducibility, and accessibility for future research. The architecture itself has

been optimized to maximize efficiency in classification tasks, allowing reliable performance even on standard hardware. In addition, the model has undergone robust validation through comprehensive statistical significance tests, confirming the reliability of the reported results. Finally, the adoption of the Cisco Lifecycle Services framework provided a structured development process, from initial clinical requirements to system optimization, thereby improving traceability and reproducibility across the entire project lifecycle.

Reproducibility Statement

To preserve data integrity and comply with institutional policies, the implementation code is not publicly released.

Nevertheless, full details of the model configuration, data preprocessing pipeline, and training setup are provided within this paper to enable independent replication.

Experiments were performed with fixed random seeds and controlled hyperparameters.

A complete environment specification (Python 3.11, NumPy 1.26, SciPy 1.11, scikit-learn 1.5) ensures reproducibility on standard hardware.

C. Limitations and Future Work

The current study also presents certain limitations that must be addressed in future research. First, the dataset size remains limited, making it necessary to validate the model with larger and more diverse samples. In addition, the classification categories lack specificity by cancer type, which restricts their clinical applicability. The performance of the system is also dependent on data quality and preprocessing protocols, and further clinical trials, regulatory approval, and interoperability testing are required before deployment in real environments. Moreover, the current categories are too broad, and more detailed labeling with specific histological subtypes is needed to strengthen diagnostic accuracy.

Future work will therefore focus on several directions. Efforts will be devoted to expanding training datasets with greater diversity and incorporating additional biomarkers to enhance diagnostic sensitivity. The architecture will be further optimized to achieve higher accuracy and complemented with improved feature extraction strategies. Explainability techniques, such as Grad-CAM for the CNN branch and SHAP for the DNN branch, will be implemented to increase interpretability and clinical trust. Finally, transfer learning and advanced residual architectures (e.g., ResNet) will be applied to improve image-based pattern recognition and boost the overall performance of the classification framework.

VI. Conclusión

This article presents a hybrid CNN-DNN algorithm for automatic cancer pattern recognition through salivary analysis. The approach achieves a precision of 71.27% by effectively combining spectral image data and biomarkers through multimodal fusion. The implementation in pure Python ensures the transparency and reproducibility of the algorithm, while its efficient architecture allows for rapid pattern recognition.

The results demonstrate the viability of automatic classification for medical pattern recognition, with potential to generate significant impact in diagnostic procedures. Future work will focus on validation with larger datasets and expansion of classification categories.

Acknowledgments

The authors thank Luis Fernando Leyva Luna for his guidance and support in the development of this study, as well as the Universidad Tecnológica del Centro de Veracruz (UTCV) and the Nanotechnology laboratories of UTCV for providing the resources and knowledge necessary for conducting this research.

References

[1] H. Sung et al., "Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality for 36 Cancers in 185 Countries,"
CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021.

[2] T. Pfaffe et al., "Diagnostic Potential of Saliva: Current State and Future Applications,"
Clinical Chemistry, vol. 57, no. 5, pp. 675–687, 2011. [in Spanish]

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning,"
Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[4] J.M. Yoshizawa and D.T. Wong, "Salivary Biomarkers: Towards Future Clinical and Diagnostic Utilities,"
Clinical Microbiology Reviews, vol. 26, no. 4, pp. 781–791, 2013. [in Spanish]

[5] L. Zhang et al., "Discovery and Preclinical Validation of Salivary Transcriptomic and Proteomic Biomarkers for Non-invasive Breast Cancer Detection,"
PLoS One, vol. 5, no. 12, p. e15573, 2010.

[6] G. Litjens et al., "A Survey on Deep Learning in Medical Image Analysis,"
Medical Image Analysis, vol. 42, pp. 60–88, 2017. [in Spanish]

[7] A. Esteva et al., "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks,"
Nature, vol. 542, no. 7639, pp. 115–118, 2017. [in Spanish]

[8] P. Rajpurkar et al., "CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning,"
arXiv preprint arXiv:1711.05225, 2017. [in Spanish]

[9] A. Rajkomar, J. Dean, and I. Kohane, "Machine Learning in Medicine,"
New England Journal of Medicine, vol. 380, no. 14, pp. 1347–1358, 2019. [in Spanish]

[10] P.M. Bossuyt et al., "STARD 2015: An Updated List of Essential Items for Reporting Diagnostic Accuracy Studies,"
BMJ, vol. 351, p. h5527, 2015. [in Spanish]

[11] N. Srivastava et al., "Dropout: A Simple Way to Prevent Neural Networks from Overfitting,"
Journal of Machine Learning Research, vol. 15, no. 1, pp. 1929–1958, 2014. [in Spanish]

[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift,"
Proc. Int. Conf. Machine Learning, 2015, pp. 448–456. [in Spanish]

[13] K. He et al., "Deep Residual Learning for Image Recognition,"
Proc. IEEE Conf. Computer Vision and Pattern Recognition, 2016, pp. 770–778. [in Spanish]

[14] D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization,"
arXiv preprint arXiv:1412.6980, 2014. [in Spanish]

[15] F. Pedregosa et al., "Scikit-Learn: Machine Learning in Python,"
Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011. [in Spanish]