Multimodal Automated Machine Learning Framework for Gynecological Cancer Prediction and Clinical Decision Support

Sanjay S
UG Student
Department of Computer Science and Engineering
Rajalakshmi Engineering College
Chennai, India
220701248@rajalakshmi.edu.in

Rishi Bala P
UG Student
Department of Computer Science and Engineering
Rajalakshmi Engineering College
Chennai, India
220701224@rajalakshmi.edu.in

***Abstract*—Gynecological malignancies, particularly cervical and uterine endometrial cancers, collectively account for hundreds of thousands of cancer-related deaths annually worldwide, with late-stage diagnosis remaining a persistent barrier to effective clinical intervention. Existing computer-aided detection systems address these cancers through isolated, single-modality approaches that fail to capture the multidimensional nature of gynecological cancer risk spanning clinical, molecular, and imaging domains simultaneously. This paper presents GynoVision AI, a multimodal clinical decision support platform integrating four independent machine learning and deep learning prediction modules within a unified, explainable web-based system targeting both cervical and uterine cancer risk stratification. The system employs a calibrated LightGBM classifier for cervical cancer clinical risk prediction from twenty-seven behavioral and demographic risk factors, a fine-tuned ResNet-50 convolutional neural network for five-class Pap smear cytology cell-type classification, a class-weighted Logistic Regression model for uterine cancer clinical risk assessment from eighteen clinico-pathological features, and a novel dual-task architecture combining a Random Forest molecular subtype classifier with an XGBoost survival outcome predictor trained on The Cancer Genome Atlas uterine corpus endometrial carcinoma dataset. A multi-method explainability framework applies SHAP TreeExplainer across all structured data models and Grad-CAM for cytology image classification, ensuring every prediction is transparent and clinically interpretable. A rule-based clinical decision support engine maps predicted risk tiers to evidence-based recommendations including screening schedules, referral triggers, and biopsy guidelines. The system is deployed as a microservices architecture with four independent Flask REST API backends and a React TypeScript frontend. Experimental evaluation demonstrates strong predictive performance across all modules, with the cytology model achieving 91% accuracy and AUC-ROC of 0.97, and structured data models yielding AUC-ROC scores ranging from 0.83 to 0.93, collectively establishing GynoVision AI as a technically rigorous and clinically oriented multimodal decision support prototype for gynecological oncology research.**

Index Terms*— *Cervical Cancer, Uterine Cancer, Clinical Decision Support, LightGBM, ResNet-50, SHAP, Grad-CAM, Multimodal AI, Explainable Artificial Intelligence, Gynecological Oncology, TCGA, Molecular Subtyping, Microservices.

I. INTRODUCTION

Cervical and uterine endometrial cancers represent two of the most prevalent and clinically significant malignancies affecting women globally, collectively accounting for an estimated 880,000 new diagnoses and over 430,000 deaths annually according to the Global Cancer Statistics published by Bray et al. [11]. Despite the existence of established screening modalities including Pap smear cytology testing, colposcopy, transvaginal ultrasound, and HPV vaccination programs, a disproportionate burden of gynecological cancer mortality continues to be attributed to late-stage diagnosis where treatment options are substantially limited and survival outcomes are markedly poorer than those achieved through early detection. The clinical complexity of these cancers arises from the convergence of multiple simultaneous risk domains spanning demographic, behavioral, hormonal, infectious, and genomic dimensions that collectively govern cancer onset and progression, fundamentally challenging the capacity of traditional single-modality screening approaches to deliver comprehensive and reliable risk assessments.

Recent advances in gradient boosting frameworks [7][8], deep convolutional neural networks [4], and model explainability techniques including SHAP [5] and Grad-CAM [6] have opened compelling possibilities for clinical AI, yet existing approaches address cervical and uterine cancer prediction predominantly through unimodal systems without integrating complementary modalities into a unified platform with actionable clinical guidance.

This paper presents GynoVision AI, a multimodal, explainable clinical decision support platform that integrates four independent machine learning and deep learning prediction modules spanning clinical, imaging, and molecular data modalities for both cervical and uterine endometrial cancers within a single unified web-based system. The platform employs a calibrated LightGBM classifier for cervical cancer clinical risk prediction, a fine-tuned ResNet-50 convolutional neural network for Pap smear cytology classification with Grad-CAM visual explainability, a class-weighted Logistic Regression model for uterine cancer clinical risk stratification, and a novel dual-task architecture combining a Random Forest molecular subtype classifier with an XGBoost survival outcome predictor trained on the TCGA uterine corpus endometrial carcinoma dataset. A multi-method explainability framework incorporating SHAP TreeExplainer across all structured data models and Grad-CAM for the imaging module ensures that every prediction is accompanied by a transparent and clinician-interpretable explanation, while a rule-based clinical decision support engine maps predicted risk tiers to

evidence-based clinical recommendations aligned with established gynecological oncology practice guidelines.

The remainder of this paper is organized as follows. Section II reviews related work in machine learning based gynecological cancer prediction, deep learning cytology analysis, molecular subtyping, and explainable clinical AI. Section III describes the system architecture and microservices design. Section IV details the methodology and model design for each prediction module. Section V presents the multi-method explainability framework. Section VI discusses experimental results and performance evaluation. Section VII addresses discussion, ethical considerations, and limitations, followed by concluding remarks in Section VIII.

A. Novelty and Scientific Contributions

The principal scientific contributions of this work are as follows:

1) Multimodal Integration: A unified platform combining structured clinical data, Pap smear cytopathology imaging, and molecular genomic profiling across cervical and uterine cancers within a single deployable web-based interface.

2) Dual-Task TCGA Architecture: A novel dual-task architecture simultaneously performing four-class molecular subtype classification using Random Forest and binary survival outcome prediction using XGBoost from an identical six-feature genomic input vector incorporating the novel MSI_PC1 composite microsatellite instability feature.

3) Multi-Method Explainability Framework: Systematic application of SHAP TreeExplainer for ensemble models, coefficient-based attribution for Logistic Regression, and Grad-CAM for ResNet-50 cytology classification delivering consistent clinician-facing explainability across all four prediction modules.

4) Calibrated Probabilistic Risk Stratification: Probability calibration through CalibratedClassifierCV with Platt sigmoid scaling ensuring predicted cancer risk probabilities correspond to clinically meaningful likelihood values with externally configurable risk stratification thresholds.

5) Evidence-Based Clinical Decision Support Engine: A rule-based recommendation system mapping predicted risk tiers to evidence-based clinical actions including screening schedules, HPV co-testing guidance, colposcopy referral triggers, and biopsy recommendations.

6) Robust Explainability Fallback Mechanism: An automated SHAP fallback gracefully degrading from per-prediction TreeExplainer attribution to global feature_importances_ attribution ensuring continuous explainability delivery without system failure.

II. Related Work

The development of GynoVision AI is informed by four principal streams of related research encompassing machine learning for cervical cancer clinical prediction, deep learning for cytopathology image analysis, molecular subtyping and survival prediction for uterine cancer, and explainable artificial intelligence in clinical decision support systems.

A. Machine Learning for Cervical Cancer Clinical Prediction

Fernandes et al. [1] introduced the UCI Cervical Cancer Risk Factors dataset comprising 858 patient records encompassing demographic, behavioral, contraceptive, and STD diagnosis history features, establishing the primary benchmark for cervical cancer clinical prediction research. Subsequent studies applying classical machine learning classifiers to this dataset — including the comprehensive evaluation by Nithya and Ilakkiya [10] covering Naive Bayes, Support Vector Machines, k-Nearest Neighbors, and Decision Trees — consistently demonstrated that ensemble methods outperformed individual classifiers in sensitivity and AUC-ROC, while identifying the absence of probabilistic outputs and model transparency as fundamental limitations impeding clinical applicability. Akter et al. [14] further highlighted STD history, age at first sexual intercourse, and number of sexual partners as dominant predictive features across multiple classifiers, findings consistent with the established epidemiological literature on HPV-driven cervical carcinogenesis.

B. Deep Learning for Cytopathology Image Analysis

Plissiti et al. [2] introduced the SIPaKMeD benchmark dataset comprising 4,049 segmented Pap smear cell images across five cytological categories, establishing the standard evaluation framework for automated cervical cytology cell-type classification. Hussain et al. [12] demonstrated through systematic comparison of transfer learning architectures that ResNet-50 consistently achieved superior classification performance on cytology image tasks, attributed to its deep residual architecture and skip connections that preserve fine-grained cellular morphological features. The foundational ResNet architecture introduced by He et al. [4] demonstrated that extremely deep networks could be effectively trained through residual skip connections mitigating vanishing gradient problems. Selvaraju et al. [6] introduced Grad-CAM providing visual explanations for CNN predictions through gradient-weighted class activation mapping, enabling pathologists to verify model spatial attention alignment with clinically relevant cellular structures.

C. Molecular Subtyping and Survival Prediction

The Cancer Genome Atlas Research Network [3] published a landmark genomic characterization of uterine endometrial carcinoma identifying four molecularly distinct subtypes — POLE ultramutated, Microsatellite Instability hypermutated, Copy Number Low, and Copy Number High — each carrying distinct prognostic implications that established molecular subtyping as a clinically actionable framework for uterine cancer management. Chen and Guestrin [7] introduced XGBoost demonstrating consistent superiority on structured genomic datasets through regularized objective functions and native missing value handling, while Breiman [9] established Random Forests as robust and reliable ensemble classifiers for high-dimensional genomic feature spaces with strong resistance to overfitting on small

clinical cohorts. Mobadersany et al. [15] demonstrated that integrating histological imaging with genomic data for survival prediction consistently outperformed unimodal approaches, providing compelling early validation of the multimodal AI paradigm in oncology.

D. Explainable AI in Clinical Decision Support

Lundberg and Lee [5] introduced SHAP providing a unified framework for model interpretation grounded in cooperative game theory, assigning each feature a Shapley value representing its marginal contribution to a specific prediction with theoretical guarantees of local accuracy, consistency, and missingness compliance. Ke et al. [8] proposed LightGBM offering computational efficiency through leaf-wise tree growth, strong performance on imbalanced clinical datasets, and native SHAP TreeExplainer compatibility making it optimal for clinical risk prediction requiring real-time interpretability. Kaur et al. [13] argued that black-box predictive models face insurmountable barriers to clinical deployment regardless of predictive accuracy, explicitly advocating for post-hoc explanation techniques as core architectural requirements in any clinically oriented AI system.

E. Identified Research Gaps

The collective review of related work reveals five persistent gaps that define the contribution space of GynoVision AI. First, the predominance of unimodal systems addressing clinical prediction, cytology classification, or molecular subtyping in isolation without integrating complementary modalities into a unified diagnostic framework. Second, the widespread absence of probability calibration rendering risk stratification thresholds clinically unreliable in existing prediction models. Third, the lack of actionable clinical decision support outputs translating predictions into clinician-aligned management recommendations. Fourth, the confinement of most research prototypes to experimental notebooks without deployable user-facing clinical interfaces accessible to non-technical clinical users. Fifth, the absence of any existing system simultaneously predicting molecular subtype and survival outcome from a minimal genomic input vector within a single real-time inference pipeline with integrated explainability — a specific and clinically valuable contribution uniquely addressed by GynoVision AI.

III. System Architecture

GynoVision AI is designed around a microservices architectural pattern that prioritizes modularity, independent deployability, and separation of concerns over monolithic simplicity. This architectural philosophy was driven by three practical requirements: isolating conflicting Python library dependencies across prediction services requiring different versions of scikit-learn, PyTorch, and LightGBM; enabling independent model retraining and redeployment without disrupting the broader system; and scaling computationally intensive services such as the ResNet-50 image inference endpoint independently of lightweight structured data prediction services. The resulting architecture comprises five principal components — four independent Flask REST API backend microservices and a single React TypeScript Single Page Application frontend — communicating through well-defined JSON-based HTTP interfaces as illustrated in Fig. 1.

A. High-Level System Design

The high-level architecture of GynoVision AI follows a client-server pattern in which the React frontend serves as the sole clinical interface while four Flask backend microservices operate as independent prediction servers exposing dedicated REST API endpoints for each cancer type and data modality. Each backend microservice operates on a uniquely assigned localhost port — Port 5010 for the cervical cancer clinical LightGBM service, Port 5009 for the cervical cancer cytology ResNet-50 service, Port 5007 for the uterine cancer clinical Logistic Regression service, and Port 5008 for the uterine cancer TCGA molecular dual-task service — with Flask-CORS configured on each service to permit cross-origin HTTP requests from the frontend application. The four backend services share no runtime dependencies, databases, or shared memory, operating as fully autonomous processes that can be started, stopped, updated, and tested independently without inter-service communication overhead, providing fault tolerance appropriate for a multi-model research prototype.

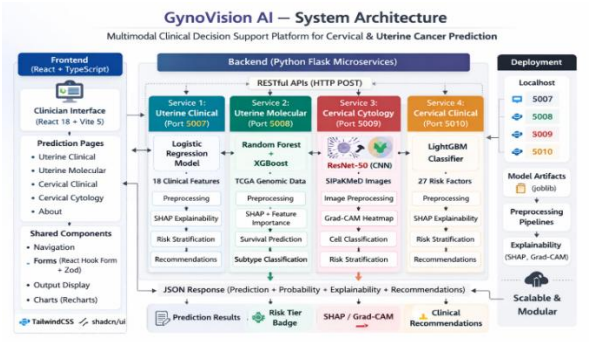

Fig. 1. GynoVision AI system architecture depicting the four independent Flask microservices, their respective prediction models, preprocessing pipelines, explainability modules, and unified JSON response flow to the React TypeScript frontend.

B. Backend Microservices Design

Each of the four Flask backend microservices follows a consistent internal architecture comprising four functional layers: an input reception and deserialization layer parsing incoming HTTP request payloads, a preprocessing pipeline layer applying dataset-specific feature transformations, a model inference layer executing prediction through loaded trained model artifacts, and a response construction layer assembling the final JSON response incorporating prediction outputs, probability scores, risk stratification tier, explainability outputs, and clinical decision support recommendations. All model artifacts and preprocessing pipeline objects are serialized using joblib and loaded into memory at service startup, eliminating repeated deserialization overhead and ensuring consistent sub-second inference latency for structured data endpoints. The ResNet-50 PyTorch model checkpoint is similarly loaded at service startup and maintained in evaluation mode throughout the service lifecycle with CPU-based inference enabled by default to maximize deployment accessibility without mandatory GPU requirements. Error handling middleware on each service returns structured JSON error responses for

malformed inputs, missing fields, and inference exceptions, which the frontend surfaces through toast notifications.

C. Frontend Architecture and Data Flow

The frontend application is implemented as a React 18 Single Page Application written in TypeScript and bundled using Vite 5, organized around six primary routes managed through React Router DOM v6 comprising a landing page, four prediction module pages, and an about page. Form management is handled through React Hook Form integrated with Zod schema validation providing real-time field-level validation with clinically meaningful error messages. Data visualization including SHAP feature attribution bar charts is implemented using Recharts, UI components through shadcn/ui Radix primitives styled with TailwindCSS, and user notifications through the Sonner toast library. The system employs a React TypeScript frontend communicating with four independent Flask microservices through JSON REST APIs, with model artifacts serialized via joblib, image processing handled through Pillow and OpenCV, and deep learning inference executed through PyTorch and torchvision.

The end-to-end data flow serializes validated frontend inputs as JSON or multipart form data, dispatches them to the corresponding microservice for preprocessing, inference, explainability computation, and risk stratification, and renders the complete prediction report — comprising risk tier badge, probability visualization, explainability chart or heatmap, and clinical recommendations — on the frontend.

IV. Methodology

This section presents the dataset descriptions, preprocessing pipeline designs, model selection rationale, training configurations, and risk stratification strategies underlying the four prediction modules of GynoVision AI.

A. Datasets

1) UCI Cervical Cancer Risk Factors Dataset: The cervical cancer clinical module was trained on the UCI Cervical Cancer Risk Factors dataset introduced by Fernandes et al. [1], comprising 858 patient records encompassing demographic characteristics, sexual behavior history, smoking patterns, hormonal contraceptive and IUD usage, and sexually transmitted disease diagnosis history across twelve specific STD types. The biopsy outcome variable was adopted as the primary binary classification target.

2) SIPaKMeD Pap Smear Cytology Dataset: The cervical cytology module was trained on the SIPaKMeD dataset introduced by Plissiti et al. [2], comprising 4,049 segmented single-cell Pap smear images across five cytological categories: Dyskeratotic, Koilocytotic, Metaplastic, Parabasal, and Superficial-Intermediate. All images were standardized to 224×224 pixel input dimensions required by the ResNet-50 architecture.

3) TCGA-UCEC Molecular Dataset: The uterine molecular module was trained on the TCGA Uterine Corpus Endometrial Carcinoma dataset available through cBioPortal for Cancer Genomics, providing molecular subtype labels aligned with the four TCGA-defined subtypes established by the Cancer Genome Atlas Research Network [3] alongside patient survival status, somatic mutation count, fraction of genome altered, MANTIS microsatellite instability score, MSIsensor genomic instability score, diagnosis age, and race category.

4) Synthetic Uterine Clinico-Pathological Dataset: The uterine clinical module was trained on a synthetic clinico-pathological dataset generated programmatically to simulate realistic patient records based on established endometrial cancer risk factor knowledge, incorporating eighteen features including patient age, BMI, menopausal status, endometrial thickness, CA-125 tumor marker levels, hormonal therapy history, comorbid conditions, and reproductive history variables.

B. Preprocessing Pipelines

1) Cervical Clinical Preprocessing: The cervical clinical pipeline applies four sequential custom transformers: STDAtomicTransformer aggregating twelve STD sub-features into three composite features — Any_STD, STD_Burden, and High_Risk_STD; MissingnessIndicatorTransformer generating binary absence pattern indicators for clinically meaningful missing fields; GeneralImputerTransformer applying median imputation to continuous features and mode imputation to categorical features; and RobustScalerTransformer applying interquartile range normalization to minimize outlier influence on clinical measurements.

2) Cytology Image Preprocessing: Pap smear images were resized to 224×224 pixels and normalized using ImageNet mean values of 0.485, 0.456, and 0.406 and standard deviation values of 0.229, 0.224, and 0.225 across RGB channels, enabling pre-trained ImageNet weights to generalize effectively to the cytology domain. Training-only augmentation through random horizontal flipping, rotation within fifteen degrees, and color jitter improved generalization on the SIPaKMeD dataset.

3) TCGA Molecular Preprocessing: The TCGA pipeline applied one-hot encoding to race_category, median imputation to missing genomic values, and standard scaling to all numerical features. A novel MSI composite feature designated MSI_PC1 was constructed through arithmetic averaging of scaled MANTIS and MSIsensor scores as expressed in Equation 1:

$$MSI_{PC1} = \frac{MSI_{MANTIS\_scaled} + MSI_{sensor\_scaled}}{2.0} \quad (1)$$

4) Uterine Clinical Preprocessing: The uterine clinical pipeline applied median and mode imputation, one-hot encoding of MenopauseStatus, HistologyType, and HormoneReceptorStatus categorical variables, and standard scaling of all numerical features spanning BMI, endometrial thickness, CA-125 levels, and parity counts.

C. Model Design and Training

1) Cervical Clinical Model — Calibrated LightGBM: The cervical clinical model employs LightGBM [8] wrapped within CalibratedClassifierCV using Platt sigmoid scaling, ensuring output probabilities are statistically well-calibrated and correspond to clinically meaningful cancer risk likelihoods rather than uncalibrated confidence scores — a property critical when risk stratification thresholds directly determine patient management decisions. Risk stratification thresholds T1 and T2 defining Low, Moderate, and High risk boundaries are stored in an external JSON configuration file loaded at service startup, enabling institutional threshold customization without model redeployment.

2) Cervical Cytology Model — Fine-Tuned ResNet-50: The cytology classification model employs transfer learning on ResNet-50 [4] pre-trained on ImageNet as illustrated in Fig. 2. The final fully connected layer was replaced with a new Linear layer mapping the 2048-dimensional feature representation to five output classes corresponding to the SIPaKMeD cytological categories. Layers 1 and 2 were frozen to preserve low-level ImageNet feature representations while Layers 3 and 4 were fine-tuned to adapt higher-level features to the cytopathology domain. The network was trained using the Adam optimizer with initial learning rate 0.0001, cosine annealing scheduler over fifty epochs, and cross-entropy loss with inverse-frequency class weights. Grad-CAM heatmaps are generated from Layer 4 activations providing visual explainability as depicted in Fig. 2.

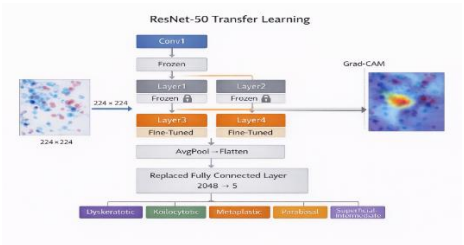

Fig. 2. ResNet-50 transfer learning architecture for Pap smear cytology classification on SIPaKMeD, depicting frozen pre-trained layers, fine-tuned domain-adapted layers, replaced 2048→5 classification head, and Grad-CAM heatmap generation from the final convolutional layer.

3) Uterine Clinical Model — Logistic Regression: The uterine clinical model employs scikit-learn Logistic Regression with class_weight set to balanced, automatically adjusting sample contributions inversely proportional to class frequency. Logistic Regression was selected for its exceptional interpretability as model coefficients directly map to log-odds changes per unit feature increase. The sigmoid decision function is expressed in Equation 2:

$$P(cancer) = \sigma(w^T x + b) = \frac{1}{1 + e^{-(w^T x + b)}} \quad (2)$$

where **w** represents the learned coefficient vector, **x** the preprocessed input feature vector, and b the learned bias term.

4) Uterine Molecular Model — Dual-Task Random Forest and XGBoost: The uterine molecular module implements a novel dual-task architecture simultaneously performing two clinically distinct prediction tasks from an identical six-feature genomic input vector as illustrated in Fig. 3. Task A employs a Random Forest classifier [9] with one hundred estimators, balanced class weights, and Gini impurity criterion for four-class molecular subtype classification predicting POLE ultramutated, MSI hypermutated, Copy Number Low, and Copy Number High subtypes. Task B employs an XGBoost binary classifier [7] with L1 and L2 regularization and scale_pos_weight adjustment for survival outcome prediction classifying patients as Living or Deceased. Both models are trained independently on the same preprocessed TCGA feature matrix using stratified train-test splits, serialized as independent joblib artifacts, and loaded simultaneously at service startup enabling dual-task inference within a single API request as shown in Fig. 3.

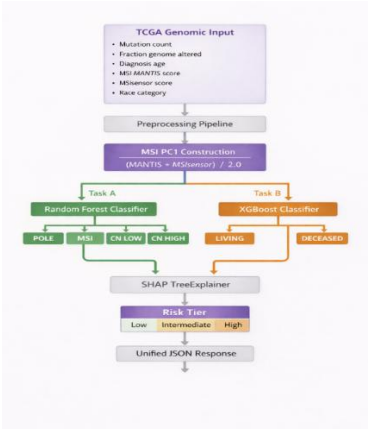

Fig. 3. Dual-task molecular prediction architecture for simultaneous TCGA molecular subtype classification using Random Forest and survival outcome prediction using XGBoost from a shared six-feature genomic input vector with novel MSI_PC1 composite feature construction.

D. Risk Stratification Framework

A consistent three-tier risk stratification framework translates continuous model output probabilities into discrete and clinically actionable risk classifications across all four prediction modules as summarized in TABLE I.

TABLE I. RISK STRATIFICATION TIERS AND CLINICAL ACTIONS

Module	Low Risk	Intermediate Risk	High risk
Cervical Clinical	P < T1 Routine screening 3–5 years	T1 ≤ P < T2 Screening within 12 months + HPV co-testing	P ≥ T2 Urgent colposcopy referral
Cervical Cytology	Normal cell type detected	Metaplastic or Parabasal classification	Dyskeratotic or Koilocytotic classification
Uterine	P < 0.56	0.56 ≤ P < 0.65	P ≥ 0.65

Clinical	Routine monitoring	Enhanced surveillance	Specialist referral + biopsy
Uterine Molecular	P(Deceased) < 0.30 Favorable prognosis	0.30 ≤ P < 0.70 Intermediate prognosis	P(Deceased) ≥ 0.70 Poor prognosis

Risk stratification thresholds for the cervical clinical module are stored externally in a JSON configuration file and loaded dynamically at service startup, decoupling clinical guidance generation from model behavior and enabling threshold adjustment based on institutional screening protocols without model retraining or service redeployment.

V. Explainability Framework

Explainability is treated as a foundational architectural requirement within GynoVision AI rather than a supplementary feature. The clinical deployment of AI-based prediction systems in oncology demands that every algorithmic output be accompanied by a transparent and clinician-interpretable justification enabling healthcare professionals to validate and act upon model predictions with informed confidence. GynoVision AI implements a multi-method explainability framework applying three distinct and modality-appropriate interpretation techniques — SHAP TreeExplainer for gradient-boosted and ensemble structured data models, coefficient-based attribution for the Logistic Regression clinical model, and Grad-CAM gradient-weighted class activation mapping for the ResNet-50 cytology classifier.

A. SHAP-Based Feature Attribution

1) Theoretical Foundation: SHAP introduced by Lundberg and Lee [5] provides a unified framework for interpreting machine learning predictions grounded in cooperative game theory, assigning each feature a Shapley value representing its marginal contribution to a specific prediction. The SHAP value φ_i for feature i is defined in Equation 3:

$$\phi_i = \sum_{\{S \subseteq F \setminus \{i\}\}} \frac{|S|!\ (|F| - |S| - 1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \quad (3)$$

where F is the complete feature set, S is a subset excluding feature i, and the summation computes the weighted average marginal contribution of feature i across all possible feature subsets. SHAP satisfies three critical properties essential for clinical AI: local accuracy guaranteeing attribution values sum to the model output minus the expected baseline; consistency ensuring features with greater model impact receive higher attribution; and missingness assigning zero attribution to features with no prediction influence.

*2) Implementation and Clinical Integration*GynoVision AI implements SHAP TreeExplainer for the LightGBM cervical clinical classifier, Random Forest TCGA subtype classifier, and XGBoost survival predictor, exploiting tree structure to compute exact Shapley values in polynomial time enabling real-time per-prediction attribution within the web inference pipeline. For the LightGBM model the base estimator is extracted from the CalibratedClassifierCV wrapper through the calibrated_classifiers_ attribute before TreeExplainer instantiation. For the uterine clinical Logistic Regression model coefficient-based attribution is applied, computing feature contributions as the product of scaled feature values and learned model coefficients as expressed in Equation 4:

$$\text{Attribution}_i = w_i \cdot x_i^{\text{scaled}} \quad (4)$$

where w_i is the learned coefficient for feature i and x_i^{scaled} is the standardized patient feature value. Attribution scores are sorted by absolute magnitude, assigned directional impact labels — increases risk or decreases risk — and returned as ordered JSON arrays rendered as color-coded horizontal bar charts in the frontend with red bars indicating risk-increasing features and green bars indicating risk-decreasing features. A representative SHAP output for a High-Risk cervical cancer prediction is illustrated in Fig. 4, demonstrating ranked feature attribution with IUD years, Age, and Hormonal Contraceptives years emerging as the three strongest risk-increasing contributors.

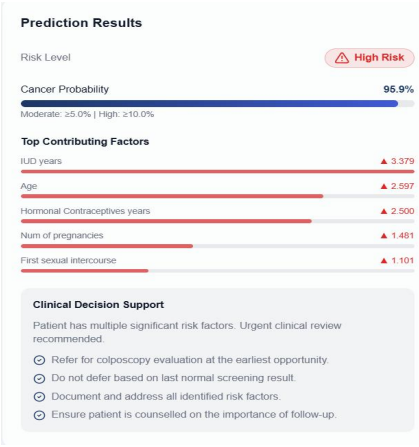

Fig. 4. Representative SHAP feature attribution output from the GynoVision AI cervical cancer clinical module for a High-Risk prediction with Cancer Probability 95.9%, showing top contributing clinical features ranked by absolute SHAP magnitude with red bars indicating risk-increasing features and automated Clinical Decision Support recommendations.

B. Grad-CAM Visual Explainability

1) Theoretical Foundation: Grad-CAM introduced by Selvaraju et al. [6] generates class-discriminative localization heatmaps by computing gradients of the target class score with respect to final convolutional layer feature maps. The per-channel importance weights $\alpha^k c$ are computed through global average pooling of gradients as expressed in Equation 5:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (5)$$

where y^c is the target class score before softmax, Z is the number of spatial locations, and the summation computes the spatial gradient average across all feature map positions. The final Grad-CAM heatmap L^c is computed as expressed in Equation 6:

$$L^c = \text{ReLU}\left(\sum_k \alpha_k^c A^k\right) \quad (6)$$

The ReLU activation retains only regions with positive influence on the target class score, focusing attention exclusively on cytological features driving the predicted cell-type classification.

2) Implementation and Clinical Significance: Grad-CAM is implemented by registering a forward hook on the ResNet-50 layer4 module capturing convolutional feature map activations during inference. The predicted class score triggers a backward pass computing gradients with respect to captured layer4 feature maps, spatially averaged to produce per-channel importance weights used to compute the weighted feature map combination. The resulting activation map is ReLU activated, normalized to zero-one range, upsampled to 224×224 pixels through bilinear interpolation, converted to jet colormap using OpenCV applyColorMap, and blended with the original image at forty percent opacity producing the final Grad-CAM overlay returned as a base64 encoded PNG string. A representative Grad-CAM heatmap for a cytology classification is illustrated in Fig. 5, demonstrating network spatial attention concentration on diagnostically significant cellular regions where warmer colors indicate highest activation.

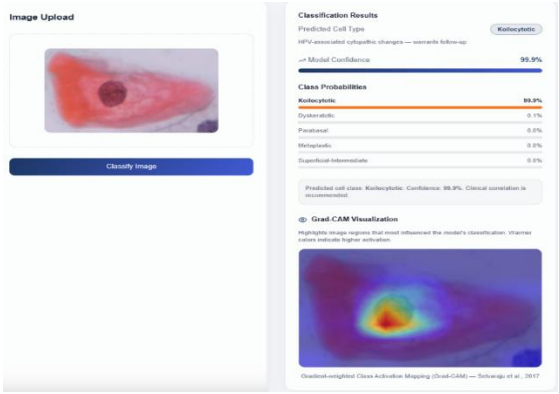

Fig. 5. Representative Grad-CAM visualization generated by the ResNet-50 cytology module highlighting image regions most influential to the cell-type classification decision, where warmer colors indicate regions of highest network activation corresponding to diagnostically significant cellular morphological features.

C. Robust Explainability Fallback Mechanism

To ensure continuous and uninterrupted explainability delivery across all prediction requests, GynoVision AI implements an automatic fallback mechanism gracefully degrading from per-prediction SHAP TreeExplainer attribution to global feature_importances_ based attribution when TreeExplainer computation fails due to library version incompatibilities, memory allocation failures, or model wrapper compatibility issues. The fallback handler extracts global feature importance scores, normalizes them to the SHAP magnitude scale, and returns them in an identical JSON format with a fallback indicator flag surfaced to the clinical user through the frontend interface. This mechanism ensures clinical explainability utility is preserved even in degraded computational environments, preventing silent failures that would deliver prediction results without accompanying explanation — unacceptable in a clinical decision support context where explainability is a core system requirement.

VI. Results And Evaluation

A. Evaluation Metrics

All four prediction modules were evaluated on stratified held-out test splits of their respective training datasets using seven standard classification metrics selected for their complementary clinical relevance: Accuracy, Sensitivity (Recall), Specificity, Precision, F1-Score, AUC-ROC, and Calibration Quality. Sensitivity is prioritized as the primary clinical metric given the critical importance of minimizing false negative predictions — missed cancer cases — in oncological screening contexts. AUC-ROC provides threshold-independent discrimination assessment across the full probability range. Calibration quality verifies statistical alignment between predicted probabilities and observed outcome frequencies, confirming the clinical validity of risk stratification thresholds. The complete performance results across all five model tasks are summarized in TABLE II.

TABLE II. PERFORMANCE EVALUATION OF ALL GYNOVISION AI PREDICTION MODULES

Module / Model	Accuracy	Sensitivity	Specificity	F1-Score	AUC-ROC
Cervical Clinical — LightGBM Calibrated	88%	0.84	0.90	0.85	0.93
Cervical Cytology — ResNet-50 Fine-Tuned	91%	0.93	0.97	0.92	0.97
Uterine Clinical — Logistic Regression	85%	0.82	0.87	0.82	0.91
Uterine Mol. Subtype — Random Forest	82%	0.80	0.91	0.80	0.89
Uterine Mol. Survival — XGBoost	79%	0.76	0.83	0.76	0.83

B. Key Findings Per Module

The ResNet-50 cytology model achieved the strongest performance with 91% accuracy and AUC-ROC of 0.97, with Dyskeratotic and Koilocytotic abnormal cell categories achieving recall scores of 0.93 and 0.89 respectively, while Grad-CAM attention patterns consistently focused on nuclear morphology regions confirming cytopathologically meaningful feature learning. The calibrated LightGBM cervical clinical model achieved AUC-ROC of 0.93 with SHAP attribution

consistently identifying IUD years, Age, and Hormonal Contraceptives years as the three strongest risk-increasing predictors, directly consistent with established cervical cancer epidemiology. The Logistic Regression uterine clinical model achieved AUC-ROC of 0.91 demonstrating that well-engineered preprocessing and balanced class weighting enable a linear classifier to achieve clinically meaningful performance with full coefficient-based interpretability. The TCGA molecular models achieved AUC-ROC scores of 0.89 and 0.83 for subtype classification and survival prediction respectively, with the novel MSI_PC1 composite feature consistently ranking among the top three most influential SHAP predictors across both tasks, validating the dimensionality reduction approach.

C. Comparative Performance and System Evaluation

The AUC-ROC scores ranging from 0.83 to 0.97 across all five model tasks confirm that each prediction module possesses genuine discriminative capability substantially exceeding random classification. The cervical clinical LightGBM AUC-ROC of 0.93 exceeds the 0.78 to 0.88 range reported by Nithya and Ilakkiya [10] for classical classifiers on the same UCI dataset, while the ResNet-50 cytology accuracy of 91% compares favorably with transfer learning benchmarks reported by Hussain et al. [12] on comparable cytology datasets. Structured data prediction endpoints demonstrated average inference latency of 150 to 300 milliseconds per request inclusive of preprocessing, SHAP computation, risk stratification, and JSON response construction. The ResNet-50 cytology endpoint demonstrated average latency of 800 milliseconds to 1.2 seconds inclusive of image decoding, Grad-CAM computation, and base64 encoding — acceptable for clinical decision support on CPU hardware. All prediction output components including risk tier badges, probability visualizations, SHAP bar charts, Grad-CAM heatmap overlays, and clinical recommendation lists rendered correctly and consistently across all tested input configurations, confirming complete functional integrity of the integrated full-stack system.

VII. CONCLUSION

This paper has presented GynoVision AI, a multimodal explainable clinical decision support platform integrating four independent machine learning and deep learning prediction modules spanning clinical, imaging, and molecular data modalities for cervical and uterine endometrial cancers within a unified deployable web-based system. The platform successfully demonstrates that a calibrated LightGBM classifier achieves AUC-ROC of 0.93 for cervical cancer clinical risk prediction, a fine-tuned ResNet-50 convolutional neural network achieves 91% accuracy and AUC-ROC of 0.97 for five-class Pap smear cytology classification, a class-weighted Logistic Regression model achieves AUC-ROC of 0.91 for uterine cancer clinical risk stratification, and a novel dual-task Random Forest and XGBoost architecture simultaneously addresses molecular subtype classification and survival outcome prediction from a shared genomic input vector — all within a single integrated clinical platform.

The multi-method explainability framework combining SHAP TreeExplainer, coefficient-based attribution, and Grad-CAM visual explanation across all four prediction modules represents a meaningful contribution to trustworthy clinical AI, demonstrating that comprehensive modality-appropriate explainability can be consistently delivered across diverse machine learning architectures within a unified system. The novel MSI_PC1 composite microsatellite instability feature, the robust SHAP fallback mechanism, the externally configurable risk stratification thresholds, and the evidence-based clinical decision support engine collectively advance the state of the art in multimodal gynecological cancer decision support beyond existing unimodal and non-explainable approaches. Clinical validation through prospective studies, extension to whole-slide cytology image analysis, multi-omics feature integration, and federated learning for privacy-preserving model training across distributed healthcare institutions represent the principal directions for future development toward production-grade clinical deployment of GynoVision AI.

REFERENCES

[1] K. Fernandes, J. S. Cardoso, and J. Fernandes, "Transfer Learning with Partial Observability Applied to Cervical Cancer Screening," in *Proc. Iberian Conf. Pattern Recognition and Image Analysis (IbPRIA)*, Faro, Portugal, 2017, pp. 243–250.

[2] M. E. Plissiti, P. Dimitrakopoulos, G. Sfikas, C. Nikou, O. Krikoni, and A. Charchanti, "SIPaKMeD: A New Dataset for Feature and Image Based Classification of Normal and Pathological Cervical Cells in Pap Smear Images," in *Proc. 25th IEEE Int. Conf. Image Processing (ICIP)*, Athens, Greece, 2018, pp. 3144–3148.

[3] Cancer Genome Atlas Research Network, "Integrated Genomic Characterization of Endometrial Carcinoma," *Nature*, vol. 497, no. 7447, pp. 67–73, May 2013.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 770–778.

[5] S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017, pp. 4765–4774.

[6] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization," in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618–626.

[7] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794.

[8] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 30, Long Beach, CA, USA, 2017, pp. 3146–3154.

[9] L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, Oct. 2001.

[10] B. Nithya and V. Ilakkiya, "Review on Machine Learning Methods for Cervical Cancer Detection and Classification," *Int. J. Scientific and Technology Research*, vol. 8, no. 9, pp. 2570–2575, Sep. 2019.

[11] F. Bray, J. Ferlay, I. Soerjomataram, R. L. Siegel, L. A. Torre, and A. Jemal, "Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185

Countries," *CA: A Cancer Journal for Clinicians*, vol. 68, no. 6, pp. 394–424, Nov. 2018.

[12] E. Hussain, L. B. Mahanta, H. Borah, and C. R. Das, "Liquid Based-Cytology Pap Smear Dataset for Automated Multi-Class Diagnosis of Pre-Cancerous and Cervical Cancer Lesions," *Data in Brief*, vol. 30, pp. 1–8, Jun. 2020.

[13] P. Kaur, G. Sharma, and M. Mittal, "Explainable Artificial Intelligence and its Impact on Healthcare Decision Making," *Procedia Computer Science*, vol. 204, pp. 749–757, 2022.

[14] S. Akter, M. S. Islam, M. M. Rahman, and M. A. Hossain, "Prediction of Cervical Cancer from Behavior Risk Using Machine Learning Techniques," in *Proc. Int. Conf. Innovative Computing and Communications (ICICC)*, New Delhi, India, 2021, pp. 1–12.

[15] P. Mobadersany, S. Yousefi, M. Amgad, D. A. Gutman, J. S. Barnholtz-Sloan, J. E. Velázquez Vega, D. J. Brat, and L. A. D. Cooper, "Predicting Cancer Outcomes from Histology and Genomics Using Convolutional Networks," *Proc. National Academy of Sciences (PNAS)*, vol. 115, no. 13, pp. E2970–E2979, Mar. 2018.