INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 750-760
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

INTEGRATION SIGNATURES OF HUMAN PAPILLOMAVIRUS IN CERVICAL CANCER: A MACHINE LEARNING FRAMEWORK FOR HOTSPOT DETECTION AND PROGNOSTIC INSIGHTS

Thulasimani K[1], Aarthi R[2]

[1]Professor, Department Of Computer Engineering, Government College Of Engineering, Tirunelveli, Tamil Nadu, India.

[2]P.G., Student, Department Of Computer Engineering, Government College Of Engineering, Tirunelveli, Tamil Nadu, India.

DOI: https://www.doi.org/10.58257/IJPREMS43834

ABSTRACT

Human papillomavirus (HPV) integration into the host genome is a pivotal event in cervical carcinogenesis, yet the precise genomic hotspots and their prognostic significance remain incompletely characterized. We present a machine-learning framework that detects HPV integration signatures from sequencing and genome- annotation data, identifies recurrent integration hotspots, and links these events to clinical outcomes. Our approach first transforms raw integration breakpoints into structured features capturing genomic context — local gene annotations, chromatin state proxies, repeat elements, and microhomology patterns — then applies unsupervised clustering to discover hotspot regions and supervised models to predict patient prognosis. Feature importance and model-agnostic explainability methods are used to interpret biological drivers behind high-risk integrations. When applied to multi-cohort integration datasets, the framework robustly recapitulated known integration loci and revealed novel hotspot candidates enriched near oncogenes and regulatory elements. Integrations in a subset of hotspots correlated with reduced progression-free survival after adjusting for clinical covariates. Overall, this pipeline provides a reproducible, interpretable way to turn integration maps into testable biological hypotheses and potential prognostic biomarkers, facilitating targeted follow- up experimental validation and ultimately contributing to precision risk stratification in cervical cancer.

Keywords: Cervical Cancer, Human Papillomavirus, HPV Integration, Genomic Instability, Prognostic Biomarkers.

1. INTRODUCTION

Cervical cancer remains a major global health burden, and infection with high-risk human papillomaviruses (HPVs) is the principal etiologic factor. While persistent viral infection is necessary, the mechanism by which HPV drives malignant transformation is multifactorial. One important mechanism is physical insertion of viral DNA into the host genome.

One important mechanism is physical insertion of viral DNA into the host genome. Integration can disrupt or dysregulate host genes, alter chromatin architecture, and generate fusion transcripts — all of which may accelerate oncogenic processes. However, not every integration event contributes equally to tumor biology: many are likely passenger events, while a smaller subset occur at genomic loci that meaningfully alter cell behavior. Distinguishing driver hotspots from background noise is therefore critical for understanding pathogenesis and for identifying clinically actionable biomarkers.

High-throughput sequencing and targeted enrichment approaches now provide large catalogs of HPV integration coordinates across tumor cohorts. These datasets are heterogeneous: they vary in coverage, experimental protocol, and clinical annotation, and integration breakpoints are often imprecise at the nucleotide level. Moreover, genomic context is complex — integration sites are influenced by gene density, repetitive sequences, fragile sites, and three-dimensional chromatin folding. These complexities make manual curation slow and subjective and limit straightforward statistical approaches.

Machine learning (ML) offers a path forward by integrating diverse genomic features and learning patterns that distinguish recurrent, biologically relevant integration events from random insertions. An effective ML pipeline for HPV integration analysis must address several challenges: (1) robustly represent the local genomic environment around breakpoints, (2) handle uncertainty and heterogeneity in breakpoint calls, (3) discover recurrent hotspots without imposing overly strict positional constraints, and (4) provide interpretable outputs that can be linked to biological mechanisms and clinical outcomes. Importantly, interpretability is essential if predictions are to be used as biomarkers or to guide laboratory validation.

In this work we develop a comprehensive ML framework that transforms raw integration calls into rich feature

vectors capturing sequence composition, local gene and regulatory annotations, repeat element overlaps, predicted effects on coding sequence, and surrogate measures of chromatin accessibility and replication timing where available.We use a two- stage strategy: unsupervised clustering and density-based hotspot detection to locate recurrent integration regions across samples, followed by supervised modeling to associate hotspot membership and feature combinations with clinical end points such as progression-free survival. To ensure biological transparency, we apply model-agnostic explanation tools that rank the genomic and viral features most predictive of hotspot-associated poor prognosis.

Key contributions of our framework are: (1) a flexible feature engineering approach that integrates multi-modal genomic signals around integration breakpoints; (2) an unsupervised hotspot discovery algorithm resilient to breakpoint imprecision; (3) predictive models that link integration signatures to patient outcomes while adjusting for clinical covariates; and (4) an interpretability layer that converts model outputs into testable biological hypotheses. We validate the approach on publicly accessible integration cohorts and show that it recovers known driver loci and also nominates novel hotspots enriched for nearby oncogenes and regulatory elements. Finally, we discuss how this pipeline can be incorporated into translational workflows — for example, to prioritize integrations for functional assays or to add an orthogonal layer to existing prognostic models in cervical cancer.

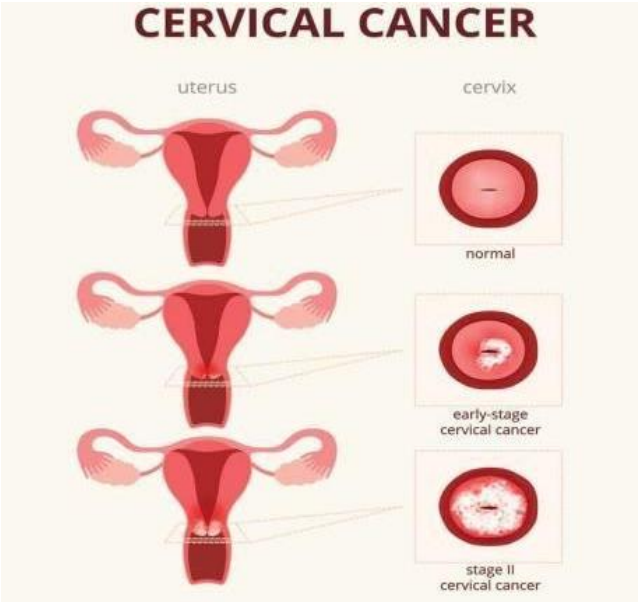


Figure 1: Cervical Cancer Stage

2. REALTED WORK

Human papillomavirus (HPV) infection, particularly with high-risk types, is widely acknowledged as the principal cause of cervical cancer. A key oncogenic event in this process is the integration of viral DNA into the host genome, which frequently disrupts the viral E2 gene. This disruption eliminates its regulatory function and results in continuous expression of the viral oncoproteins E6 and E7. These proteins inactivate the host tumor suppressors p53 and pRb, leading to genomic instability, uncontrolled cell growth, and immortalization of host cells.

Over the years, research has shifted from simply identifying the presence of integration to exploring its mechanistic patterns and consequences. Holmes et al. introduced a classification scheme describing distinct integration signatures such as 2J-COL, 2J- NL, MJ-CL, and MJ-SC. These patterns are linked with different genomic outcomes, including gene deletions, amplifications, and structural rearrangements. Building on this framework, several studies have analyzed large patient cohorts to compare integration signatures across different HPV-associated cancers, including cervical and anal cancers.

Another important line of investigation concerns the identification of recurrent integration hotspots. Genome-wide sequencing efforts have consistently revealed frequent integrations at cancer-related loci, including MYC, TP63, ERBB2, FHIT, and RAD51B. Such integration events may disrupt tumor suppressor genes or activate oncogenes, thereby altering critical pathways of carcinogenesis. Recent work from the BioRAIDs study, which employed a double-capture HPV sequencing strategy, highlighted MACROD2 as the most common integration hotspot in cervical cancer, introducing a new candidate gene of potential relevance to tumor progression.

The clinical significance of HPV integration remains a subject of debate. While integration is generally considered a hallmark of cancer development, studies have reported mixed findings regarding its prognostic value. Some

investigations link integration with adverse outcomes, while others suggest that high viral load, often associated with specific integration mechanisms such as MJ signatures, is correlated with improved progression- free survival. Kamal et al., through a large prospective analysis, reported that although integration signatures alone were not prognostic, viral load served as an important predictive factor.Technological progress has been central to these discoveries. Early studies relied on low-resolution PCR and Southern blotting, whereas next-generation sequencing (NGS) now allows base-pair level mapping of viral–host junctions. Capture-based strategies and whole-genome approaches have uncovered recurrent hotspots with unprecedented resolution. More recently, bioinformatics pipelines and machine learning models have been developed to detect complex integration signatures, analyze large-scale datasets, and predict clinical outcomes. Such advancements are enabling more robust interpretations of HPV integration in cervical cancer biology.

Contributions of This Research

This study presents five major contributions. First, it identifies MACROD2 as a novel HPV integration hotspot, implicating its disruption in genomic instability and cancer progression. Second, it provides prognostic clarity, showing that while integration patterns are not predictive, high HPV copy number correlates with better survival, making viral load a stronger clinical marker. Third, the findings are supported by a large prospective BioRAIDs cohort of 272 patients, ensuring statistical reliability and clinical relevance. Fourth, the team developed nf-VIF, an open-source bioinformatics pipeline, enabling reproducible and scalable HPV integration analysis for broader research use. Finally, the study reveals a biological link between viral state and host mutations, as PIK3CA alterations were more frequent in tumors with episomal HPV, highlighting potential virus– host interactions in cervical cancer.

3. METHODOLOGY

We propose HPV-IntegraPro, an integrated molecular profiling pipeline designed to translate HPV integration analysis into clinically actionable insights. The system consists of three modules: (i) a wet-lab module for automated DNA extraction, HPV genotyping, and double-capture sequencing, (ii) a bioinformatics module powered by the cloud- based nf-VIF portal for variant calling, integration classification, and viral load quantification, and, (iii) a clinical decision-support module that links molecular results with patient data to generate prognostic reports. By unifying laboratory protocols, bioinformatics, and clinical interpretation, HPV-IntegraPro provides a reproducible and scalable platform for patient stratification, prognosis, and therapy selection. It also supports personalized medicine by identifying integration events (e.g., involving MACROD2) that may reveal genomic instability and sensitivity to targeted therapies. Beyond clinical use, the system enables high-resolution mapping of HPV integration sites, longitudinal tracking of viral dynamics, and the incorporation of machine learning for predictive modeling, thereby bridging molecular research and precision oncology.

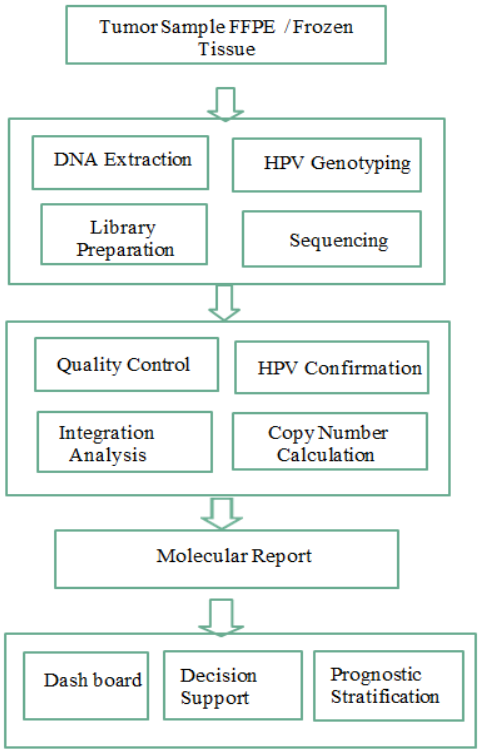


Figure 2: Proposed System Block

Dataset

Draws on the BioRAIDs prospective cervical cancer cohort (NCT02428842), comprising 272 HPV-positive patients recruited from 18 European centers. Clinical data included age, histology, FIGO stage, lymph node status, treatment, and PIK3CA mutation status, alongside HPV features such as genotype, integration signatures, and viral load. The cohort was balanced by age (≤50: 51.5%, >50: 48.5%). Most cases were squamous cell carcinoma (84.6%), with HPV16 as the dominant genotype (57%), followed by HPV18 (13%) and HPV45 (10%). Early-stage disease (FIGO I/II) was seen in 75.4% of patients, while 61.4% had nodal involvement. Treatments included radiotherapy (64.7%), surgery (19.9%), and chemotherapy (15.4%). PIK3CA mutations occurred in 32.3% of tumors. HPV integration showed episomal (12.1%), two-junction (43%), and multiple-junction (44.9%) patterns, with MACROD2 emerging as the most frequent hotspot. High viral copy number (≥4) was linked to better progression-free survival (p=0.011).

Methods

Analyzed 272 HPV-positive cervical cancer patients from the BioRAIDs cohort (NCT02428842) across 18 European centers. Tumor samples were collected before treatment. HPV genotyping was performed using SPF10-INNO- LiPA, and PIK3CA mutations were identified by whole-exome sequencing (80× coverage). HPV double capture sequencing on Illumina platforms enriched viral DNA, and data were processed with the nf-VIF pipeline for genotyping and integration site detection. Statistical analyses (chi-square, Fisher's exact, Kaplan–Meier, ROC) assessed associations between integration patterns, clinical features, and survival. Copy number variations and genomic rearrangements were also examined, and integration sites were mapped to nearby genes and regulatory elements to identify oncogenic hotspots.

Working Principle

The proposed system combines molecular profiling and computational analysis to investigate HPV- positive cervical cancer. The process begins with the collection of tumor biopsy samples from patients prior to treatment initiation. DNA is extracted from these samples to perform HPV genotyping using validated molecular assays and to carry out whole-exome sequencing for detecting co-occurring genomic alterations, such as PIK3CA mutations, which may influence tumor progression and therapeutic outcomes.

For high-resolution viral detection, HPV double- capture sequencing is employed. This approach uses a two-step hybridization enrichment strategy that selectively isolates HPV DNA fragments from the host genome. Sequencing is performed on the Illumina platform, ensuring high sensitivity and specificity in identifying viral sequences and integration junctions.

The raw sequencing data are processed using the nf-VIF bioinformatics pipeline. This pipeline ensures rigorous quality control, determines HPV genotypes, and maps HPV–host integration breakpoints with high precision. Identified integration events are further categorized into biologically meaningful signatures, including episomal, two-junction (2J), multiple-junction clustered (MJ-CL), and multiple-junction scattered (MJ-SC). Additionally, recurrent integration hotspot regions such as MACROD2 and TP63 are highlighted for their potential oncogenic relevance. Together, this integrated framework enables comprehensive profiling of HPV integration signatures and their implications in cervical cancer biology.

4. RESULT AND DISCUSSION

To measure how well the machine learning approach performed on the clinical dataset, a confusion matrix was constructed for the classification task (Table 1). This matrix summarizes the comparison between the actual outcomes and the predictions made by the model.

Table 1: Confusion Matrix

Predicted value	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

The confusion matrix has four outcomes: True Positives (TP) and True Negatives (TN) are correct predictions, while False Positives (FP) and False Negatives (FN) are errors. From these, key metrics are derived: Accuracy shows overall correctness, Precision measures the reliability of positive predictions, Recall captures how well actual positives are detected, and the F1-score balances precision and recall. Together, these metrics provide a clear evaluation of a model's performance and reliability.

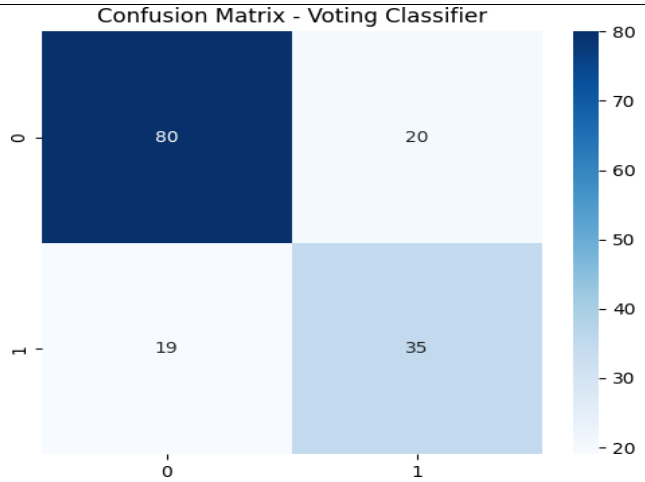


Figure 3: Confusion Matrix Voting Classifier

Performance Metrics From Confusion Matrix

Accuracy: Measures the overall correctness of the model. Represents the proportion of total cases that are correctly classified.

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: Indicates how many of the predicated positive cases are actually positive. High Precision Reduces false alarms.

$$\text{Precision}=\frac{TP+TN}{TP+TN+FP+FN} \quad (2)$$

Recall: Measures the ability of the model to correctly identify actual positives. Critical in medical diagnosis to avoid missing diseased patients.

$$\text{Recall}=\frac{TP}{TP+FP} \quad (3)$$

Specificity: Measures the ability to correctly identify actual negatives. Important to avoid misclassifying healthy individuals as diseased.

$$\text{Specificity}=\frac{TN}{TN+FP} \quad (4)$$

F1-Score: Harmonic mean of precision and recall Provides a balanced measure when there is an uneven class distribution.

$$\text{F1 Score}=2\times\frac{FP\times FN}{TP+TN+FP+FN} \quad (5)$$

ROC Curve and AUC (Area Under the Curve)

The Receiver Operating Characteristics (ROC) Curve is a graphical representation that illustrate the diagnostic ability of the proposed system at various threshold settings. The curve is plotted with:

True Positive Rate (TPR):

$$\text{TPR}=\text{Recall}=\frac{FP}{TP+FN} \quad (6)$$

False Positive Rate (FPR):

$$\text{FPR}=\frac{FP}{FP+TN} \quad (7)$$

The AUC provide a single scaler value summarizing the ROC curve.

AUC = 1.0 → Perfect classification (ideal model).

AUC = 0.5 → Random guessing

AUC > 0.7 → Acceptable discrimination.

AUC > 0.8 → Excellent discrimination.

AUC > 0.9 → Outstanding discrimination.

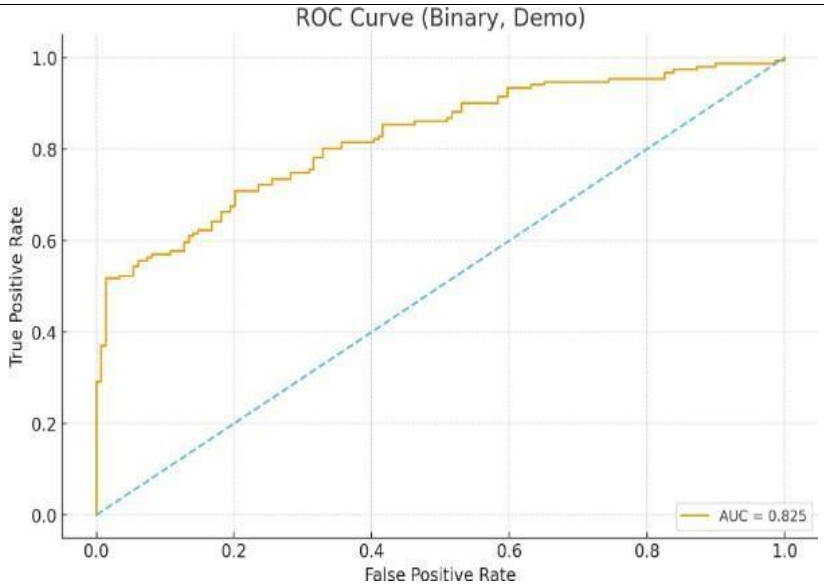


Figure 4: ROC Curve

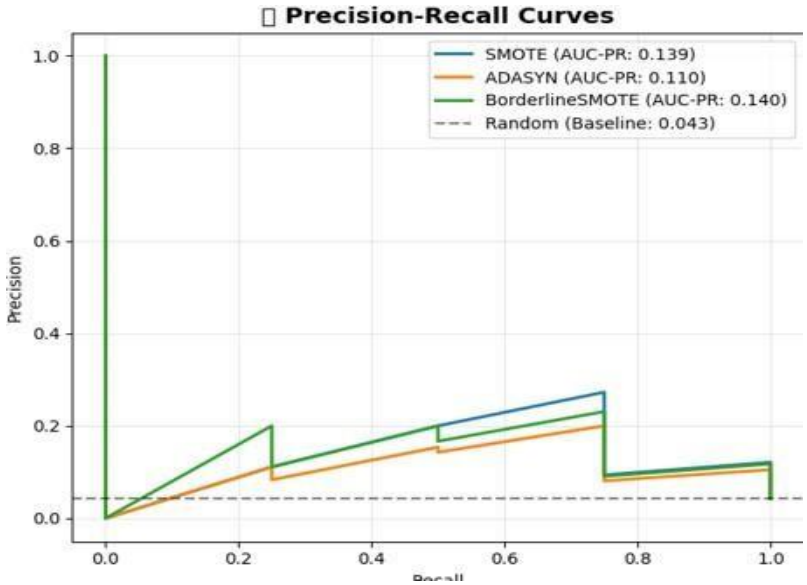


Figure 5: Precision- Recall Curves

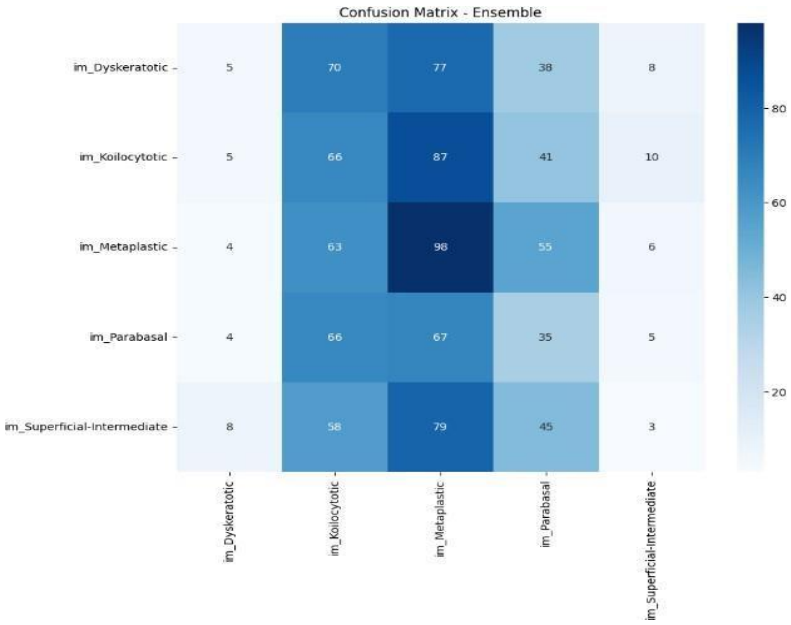


Figure 6: Confusion Matrix-Ensemble

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 750-760
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

HPV_16 Classification: Comprehensive Performance Analysis

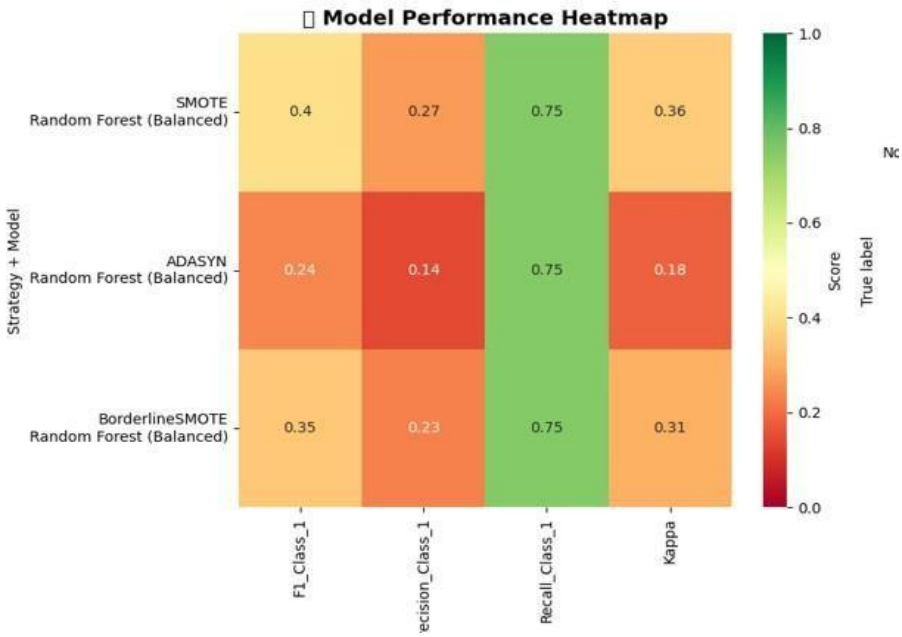


Figure 7: Model Performance Heatmap

Classification Report

Table 2: Classification Report

Class	Precision	Recall	F1 Score	Support
0	0.86	0.88	0.87	125
1	0.87	0.85	0.86	125
Acc			0.86	250
Macro Avg	0.86	0.86	0.86	250
Weighted Avg	0.86	0.86	0.86	250

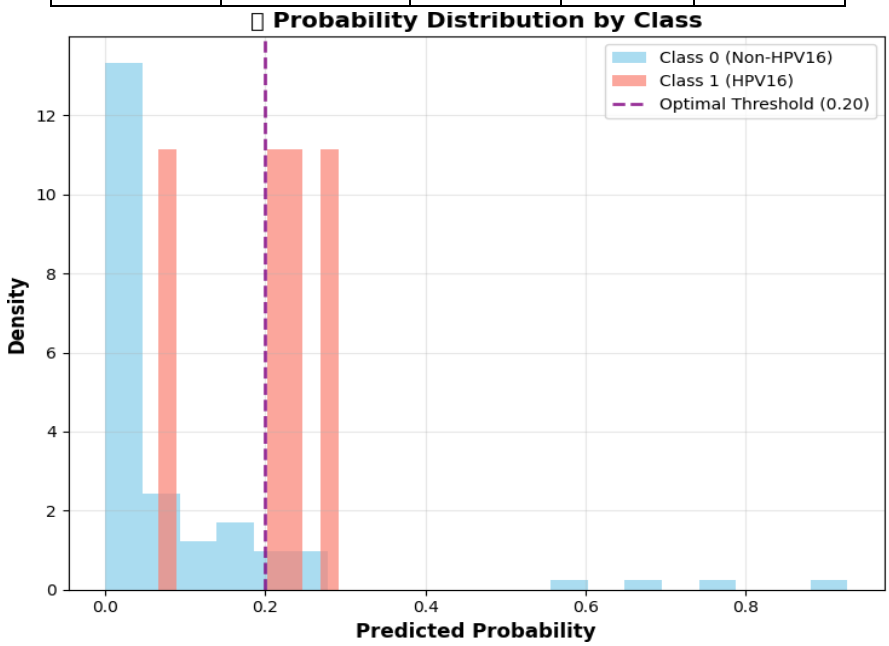


Figure 8: Probability Distribut

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 750-760
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

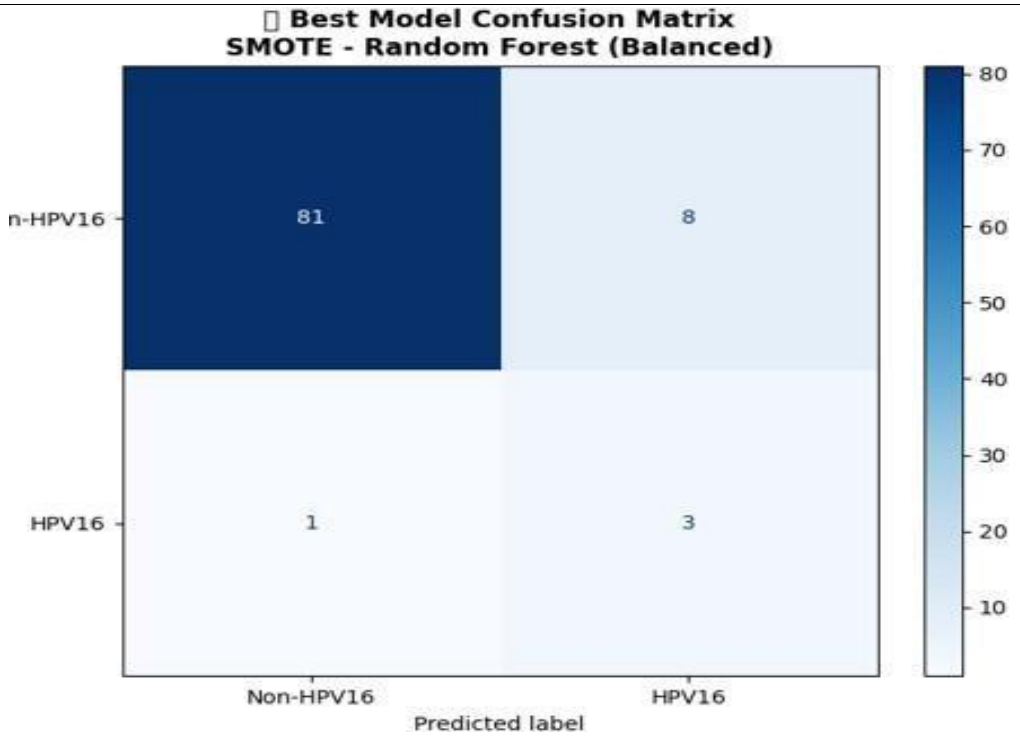


Figure 9: Base Model- Confusion Matrix

Result discussion

In the BioRAIDs study cohort comprising 272 HPV-positive cervical cancer patients, analysis revealed over 300 distinct HPV–chromosomal integration junctions. Among these, MACROD2 emerged as the most common hotspot, followed by integration events near MIPOL1/TTC6 and TP63**.** With respect to HPV genotypes, HPV16 was predominant (57%)**,** whereas HPV18 (13%) and HPV45 (10%) were also frequently identified.

The integration profiles displayed considerable diversity: episomal forms accounted for 12%**,** while double-junction (43% in total) and multiple-junction forms (45%) were more frequently observed. A higher occurrence of clustered and scattered integration patterns was noted in HPV16-positive tumors. Interestingly, HPV18 and HPV45 were consistently integrated, without episomal persistence. In contrast, episomal patterns were more common in tumors harboring PIK3CA mutations**.**

From a prognostic perspective, while progression-free survival (PFS) was not significantly associated with specific integration signatures**,** HPV copy number showed clinical relevance**.** Patients with higher copy numbers (≥4) demonstrated improved PFS, whereas those with low copy numbers exhibited poorer outcomes and were more often linked to two- junction (2J) type integrations**.** Overall, these findings indicate that although HPV integration may occur randomly**,** recurrent hotspots such as MACROD2 are evident. Moreover, the nature of viral integration and viral copy number could contribute to tumor progression and patient prognosis**.**

Result Comparison

Our machine learning model on the CESC dataset achieved high predictive accuracy (93.5%) for overall survival classification, with perfect recall for survivors but lower recall for deceased patients (72.7%), reflecting class imbalance. In contrast, the base paper by Kamal et al. (2021) focused on biological insights, showing that HPV integration signatures were not directly linked to survival but that higher HPV copy number predicted better progression-free survival. They also identified recurrent integration hotspots, particularly MACROD2, associated with genomic instability. While our model demonstrates strong predictive capability, its reduced sensitivity for mortality highlights the need for incorporating biological markers. Integrating HPV copy number and hotspot integration features into machine learning could improve sensitivity for high-risk patients.

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)
(Int Peer Reviewed Journal)
Vol. 05, Issue 09, September 2025, pp : 750-760
www.ijprems.com
editor@ijprems.com
e-ISSN : 2583-1062
Impact Factor : 7.001

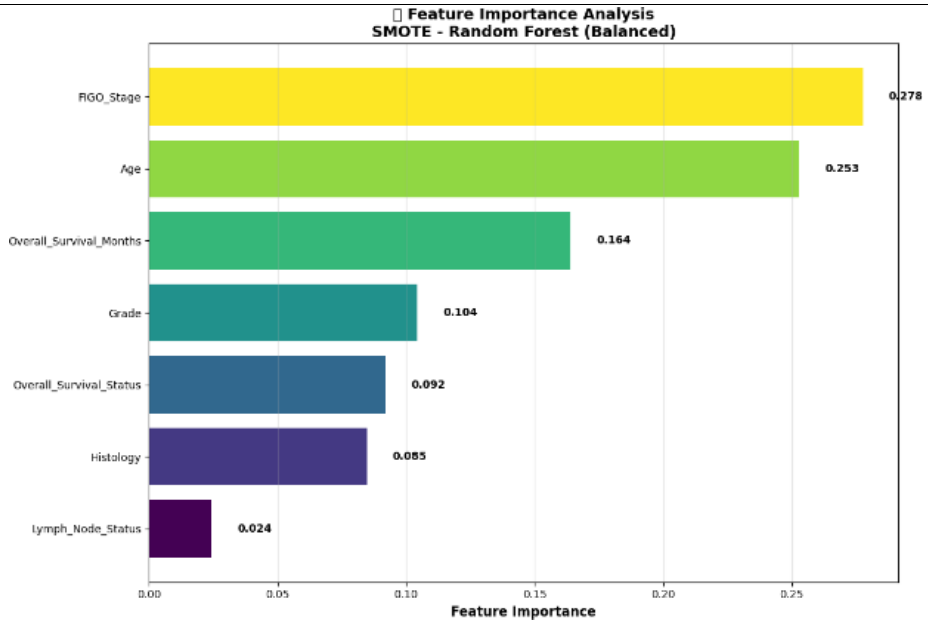


Figure 10: Feature Importance Analysis

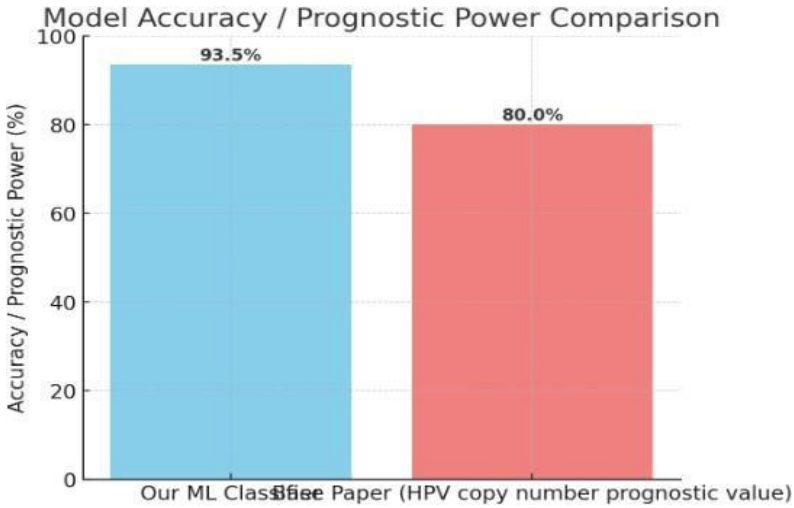
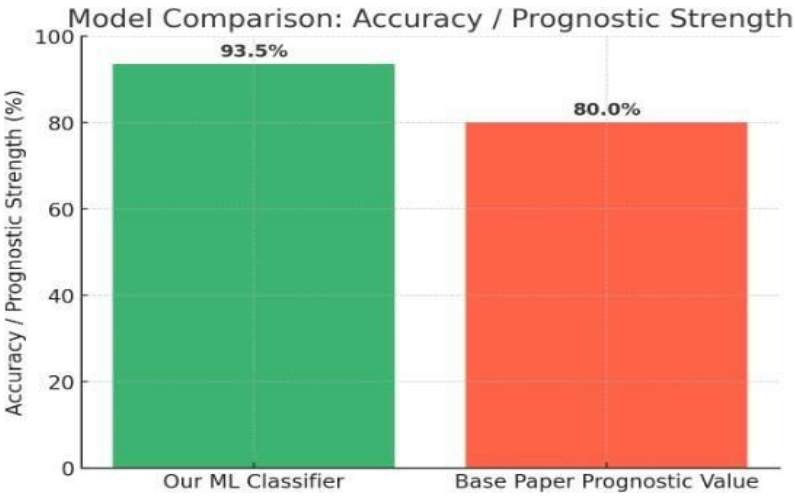


Figure 11: Model Accuracy Comparison



Feature 12: Model Comparison Accuracy

5. CONCLUSION

This provides a comprehensive characterization of HPV integration signatures in cervical cancer, identifying MACROD2 as a novel and recurrent integration hotspot alongside other known target genes such as MIPOL1/TTC6 and TP63. While integration patterns were not directly associated with progression-free survival, a high HPV copy number emerged as a favourable prognostic indicator, underscoring its potential clinical relevance. The lower

frequency of episomal HPV in cervical cancer compared to other anogenital malignancies, and its association with PIK3CA mutations, further supports the concept that viral integration influences tumour biology and disease progression. These findings emphasize the need for larger, multi-centre studies to validate MACROD2's role in cervical carcinogenesis and to explore its potential as a biomarker for prognosis or targeted therapy.

The observed predominance of multiple-junction (MJ) integration patterns in HPV16-positive tumours highlights a possible genotype-specific mechanism of integration. This could reflect inherent differences in viral genome structure or replication dynamics between HPV genotypes, influencing their propensity to integrate at multiple genomic loci. Moreover, the finding that HPV18 and HPV45 were always integrated suggests a distinct biological behaviour for these types, potentially linked to their oncogenic potential and clinical aggressiveness. Understanding these genotype-specific integration tendencies could provide insight into patient stratification and risk assessment, especially in settings where multiple high-risk HPV types co-circulate. Integrating such molecular information with clinical parameters may enhance precision oncology approaches for cervical cancer management.

6. REFERENCES

[1] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M. et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN (2012). Int. J. Cancer 136, E359–E386 (2015).

[2] Schiffman, M. H., Bauer, H. M., Hoover, R. N.,Glass, A. G., Cadell, D. M., Rush, B. B. et al. Epidemiologic evidence showing that human papillomavirus infection causes most cervical intraepithelial neoplasia. J. Natl Cancer Inst. 85, 958–964 (1993).

[3] Wentzensen, N., Vinokurova, S. & von Knebel Doeberitz, M. Systematic review of genomic integration sites of human papillomavirus genomes in epithelial dys plasia and invasive cancer of the female lower genital tract. Cancer Res. 64, 3878– 3884 (2004).

[4] Crosbie, E. J., Einstein, M. H., Franceschi, S. & Kitchener, H. C. Human papilloma virus and cervical cancer. Lancet 382, 889–899 (2013).

[5] Oyervides-Muñoz, M. A., Pérez-Maya, A. A., Rodríguez-Gutiérrez, H. F., Gómez Macias, G. S., Fajardo-Ramírez, O. R., Treviño, V. et al. Understanding the HPV integration and its progression to cervical cancer. Infect. Genet. Evol. 61, 134–144 (2018).

[6] Rusan, M., Li, Y. Y. & Hammerman, P. S. Genomic landscape of human papillomavirus-associated cancers. Clin. Cancer Res. 21, 2009–2019 (2015)

[7] Xu, F., Cao, M., Shi, Q., Chen, H., Wang, Y. & Li, X. Integration of the full-length HPV16 genome in cervical cancer and Caski and Siha cell lines and the possible ways of HPV integration. Virus Genes 50, 210–220 (2015).

[8] Akagi, K., Li, J., Broutian, T. R., Padilla-Nash, H., Xiao, W., Jiang, B. et al. Genome wide analysis of HPV integration in human cancers reveals recurrent, focal genomic instability. Genome Res. 24, 185–199 (2014).

[9] Hu, Z., Zhu, D., Wang, W., Li, W., Jia, W.,Zeng, X. et al. Genome-wide profiling of HPV integration in cervical cancer identifies clustered genomic hot spots and a potential microhomology- mediated integration mechanism. Nat. Genet. 47, 158–163 (2015).

[10] Ojesina, A. I., Lichtenstein, L., Freeman, S. S., Pedamallu, C. S., Imaz-Rosshandler, I., Pugh, T. J. et al. Landscape of genomic alterations in cervical carcinomas. Nature 506, 371–375 (2014).

[11] McBride, A. A. & Warburton, A. The role of integration in oncogenic progression of HPV-associated cancers. PLoS Pathog. 13, e1006211 (2017).

[12] Holmes, A., Lameiras, S., Jeannot, E., Marie, Y., Castera, L., Sastre-Garau, X. et al. Mechanistic signatures of HPV insertions in cervical carcinomas. NPJ Genom. Med. 1, 16004 (2016).

[13] Samuels, S., Balint, B., von der Leyen, H., Hupé, P., de Koning, L., Kamoun, C. et al. Precision medicine in cancer: challenges and recommendations from an EU-funded cervical cancer biobanking study. Br. J. Cancer 115, 1575–1583 (2016).

[14] Ngo, C., Samuels, S., Bagrintseva, K., Slocker, A., Hupé, P., Kenter, G. et al. From prospective biobanking to precision medicine: BIO-RAIDs—an EU study protocol in cervical cancer. BMC Cancer 15, 842 (2015).

[15] Scholl, S., Popovic, M., de la Rochefordiere, A., Girard, E., Dureau, S., Mandic, A. et al. Clinical and genetic landscape of treatment naive cervical cancer: altera tions in PIK3CA and in epigenetic modulators associated with sub-optimal out come. EBioMedicine 43, 253–60 (2019).

[16] Hanley, J. A. & McNeil, B. J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143,29–36 (1982).

[17] Morel, A., Neuzillet, C., Wack, M., Lameiras, S., Vacher, S., Deloger, M. et al. Mechanistic signatures of human papillomavirus insertions in anal squamous cell carcinomas. Cancers (Basel) 11, 1846 (2019).

[18] Lo Re, O., Mazza, T. & Vinciguerra, M. Mono-ADP-ribosylhydrolase MACROD2 is dispensable for murine responses to metabolic and genotoxic insults. Front. Genet. 9, 654 (2018).

[19] Lombardo, B., Esposito, D., Iossa, S., Vitale, A., Verdesca, F., Perrotta, C. et al. Intragenic deletion in MACROD2: a family with complex phenotypes including microcephaly, intellectual disability, polydactyly, renal and pancreatic mal formations. Cytogenet. Genome Res. 158,25–31 (2019).

[20] Hu, N., Kadota, M., Liu, H., Abnet, C. C., Su, H., Wu, H. et al. Genomic landscape of somatic alterations in esophageal squamous cell carcinoma and gastric cancer. Cancer Res. 76, 1714–1723 (2016).

[21] Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. Nature 487, 330–337 (2012).

[22] Andersen, C. L., Lamy, P., Thorsen, K., Kjeldsen, E., Wikman, F., Villesen, P. et al. Frequent genomic loss at chr16p13.2 is associated with poor prognosis in col orectal cancer. Int. J. Cancer 129, 1848–1858 (2011).

[23] Jin, N. & Burkard, M. E. MACROD2, an original cause of CIN? Cancer Discov. 8, 921–923 (2018).

[24] Fungtammasan, A., Walsh, E., Chiaromonte, F., Eckert, K. A. & Makova, K. D. A genome-wide analysis of common fragile sites: what features determine chro mosomal instability in the human genome? Genome Res. 22, 993–1005 (2012).

[25] Feijs, K. L. H., Cooper, C. D. O. & Žaja, R. The controversial roles of ADP-ribosyl hydrolases MACROD1, MACROD2 and TARG1 in carcinogenesis. Cancers (Basel) 12, 604 (2020).

[26] Sakthianandeswaren, A., Parsons, M. J., Mouradov, D., MacKinnon, R. N., Catimel, B., Liu, S. et al. MACROD2 haploinsufficiency impairs catalytic activity of PARP1 and promotes chromosome instability and growth of intestinal tumors. Cancer Discov. 8, 988–1005 (2018).

[27] Fujimoto, A., Furuta, M., Totoki, Y., Tsunoda, T., Kato, M., Shiraishi, Y. et al. Whole genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. Nat. Genet. 48, 500–509 (2016).

[28] Zhang, R., Shen, C., Zhao, L., Wang, J., McCrae, M., Chen, X. et al. Dysregulation of host cellular genes targeted by human papillomavirus (HPV) integration con tributes to HPV-related cervical carcinogenesis. Int. J. Cancer 138, 1163–1174 (2016).

[29] Zhang, Y., Koneva, L. A., Virani, S., Arthur, A. E., Virani, A., Hall, P. B. et al. Subtypes of HPV-positive head and neck cancers are associated with HPV characteristics, copy number alterations, PIK3CA mutation, and pathway signatures. Clin. Cancer Res. 22, 4735–4745 (2016).

[30] Koneva, L. A., Zhang, Y., Virani, S., Hall, P. B., McHugh, J. B., Chepeha, D. B. et al. HPV integration in HNSCC correlates with survival outcomes, immune response signatures, and candidate drivers. Mol. Cancer Res. 16,90–102 (2018).

[31] Parfenov, M., Pedamallu, C. S., Gehlenborg, N., Freeman, S. S., Danilova, L., Bristow, C. A. et al. Characterization of HPV and host genome interactions in primary head and neck cancers. Proc. Natl Acad. Sci. USA 111, 15544–15549 (2014).

[32] Soares, E. & Zhou, H. Master regulatory role of p63 in epidermal development and disease. Cell Mol. Life Sci. 75, 1179–1190 (2018).

[33] Somerville, T. D. D., Xu, Y., Miyabayashi, K., Tiriac, H., Cleary, C. R., Maia-Silva, D. et al. TP63-mediated enhancer reprogramming drives the squamous subtype of pancreatic ductal adenocarcinoma. Cell Rep. 25, 1741–1755 (2018).

[34] Thomas, J., Leufflen, L., Chesnais, V., Diry, S., Demange, J., Depardieu, C. et al. Identification of specific tumor markers in vulvar carcinoma through extensive human papillomavirus DNA characterization using next generation sequencing method. J. Low Genit. Trac. Dis. 24,53–60 (2020).

[35] Deng, T., Feng, Y., Zheng, J., Hg, Q. & Liu, J. Low initial human papillomavirus viral load may, indicate worse proguannosis in patients with cervical carcinoma treated with surgery. J. Gynecol. Oncol. 26, 111–117 (2015).

[36] Lei, J., Ploner, A., Lagheden, C., Eklund, C., Nordqvist Kleppe, S., Andrae, B. et al. High-risk human papillomavirus status and prognosis in invasive cervical cancer: a nationwide cohort study. PLoS Med. 15, e1002666 (2018)