RESEARCH PAPER

Probabilistic Potential Field–Based Multimodal Machine Learning for Early Risk Stratification in Non-Hodgkin Lymphoma

Meduri Raghu Chandra[1] and Dr.D.Usha[2]

[1]Research Scholar, Department of Computer Science and Engineering, Dr.M.G.R.Educational and Research Institute
[2]Professor, Department of Data Science, Dr.M.G.R.Educational and Research Institute,
[1]mraghuchandra191@gmail.com and [2]usha.cse@drmgrdu.ac.in

Received: 28th Feb, 2026; Revised: 6th March 2026; Accepted: 7th April, 2026; Available Online: 20th April, 2026

ABSTRACT
NHL is a pathology that has inherent clinical and molecular heterogeneity; therefore, early risk identification has been at the centre of concern in oncology. The existing diagnostic systems are highly qualified, yet they are likely to be based on one mode of imaging or omics and cannot be compared to a broad range of patients. A machine learning model based on Probabilistic Potential Field (PPF) is used to geometrically model multimodal patient data to identify early warning signals in NHL in this paper. The proposed model constructs an active geometrical surface, which evaluates the interaction of patient-specific aspects and gradients of risks and enables a disease development to be visualized. The PPF system had a superior discriminative and generalization capacity compared with the conventional survival and neural models in multimodal data of imaging, metabolomic, and clinical measures. The findings of the quantitative evaluation showed that the geometric mapping technique was strong as the concordance indices and low percentages of prediction error were noted when the data was studied at the first level of classification. The results suggest that the PPF-based method may be able to model nonlinear associations between heterogeneous biomarkers, which can portray a mathematically-founded and interpretable resolution of the prognosis of NHL at an earlier stage. This study will result in better precision haematology based on better prediction of risks, transparency of the model, and a data-driven clinical decision model in the oncology practice.

Keywords: *Non-Hodgkin Lymphoma, Machine Learning, Probabilistic Potential Field, Geometry-Based Risk Detection, Multimodal Biomarkers, Early Diagnosis*
How to cite this article: Chandra MR and Usha D, Probabilistic Potential Field–Based Multimodal Machine Learning for Early Risk Stratification in Non-Hodgkin Lymphoma. Int J Drug Deliv Technol. 2026;16(6): 400-409. DOI: 10.25258/ijddt.16.6.75
Source of support: Nil.
Conflict of interest: None

1. INTRODUCTION

Background

It is a heterogeneous group of lymphoid malignancies that make up approximately 90% of lymphoma worldwide and is still growing in the frequency of cases across the world (Alibrahim et al., 2025). Its changeability in molecular features and clinical presentation makes it difficult to be detected at an early stage and planned to be treated. Even though there is an improvement in molecular diagnostics and imaging, the current systems tend to utilize isolated modalities, including histopathology or genomics, which do not reflect the intricate biological interaction affecting disease progression (Gottschlich et al., 2025). This causes the neglect of early-stage cases, thus restricting efficiency of timely therapeutic intervention.

Recent studies have presented machine learning and bioinformatics models to enhance lymphoma predication of diagnosis and prognosis. Jeon et al. (2024) fitted a competing-risk model to analyse the risks of secondary malignancy, and Li et al. (2024) fitted a Random Survival Forest (RSF) model, which has greater predictive power than the competing-risk model but is not interpretable. Liu et al. (2025) suggested a model of CSF N-glycome based on diagnostic of central nervous system lymphoma, and Ma et al. (2025) investigated biomarkers based on metabolic pathways that can be used in diagnostic processes. Equally, Montagne et al. (2025) used metabolomics based on plasma NMR to predict the risk in DLBCL, and Tan et al. (2024) trained an interpretable neural survival model. Yet these research works were mainly concerned with single modalities of data. Wang et al. (2025a, 2025b) had employed multimodal MRI-based radiomics to predict the recurrence but was not biologically interpretable. In general, these studies indicate a high level of progress but show the lack of an integrated and explainable framework that can integrate multimodal information to identify early NHL risk.

Problem Statement

1. Lack of Data Integration: Most of the models currently only work with a single modality (radiomics,

**Author for Correspondence: mraghuchandra191@gmail.com*

metabolomics, or proteomics), which results in incomplete knowledge.

2. Inability to be Interpretable: Deep learning models are highly accurate and will have little transparency to clinicians.
3. Weak Mathematical Foundations: There is no good geometric or probabilistic understanding of the dynamics of disease progression in many of the existing models.

Research Objective

1. To construct a geometric risk field that represents nonlinear interactions between radiomic, metabolomic and clinical variables.
2. To create a PPF-based algorithm that can be used to visualize patient specific disease trajectories.
3. To compare the model to the benchmark survival prediction methods in terms of interpretability and performance.

Significance and Innovation

The suggested PPF paradigm provides a geometrical interpretation, interpretable learning architecture, which incorporates a combination of heterogeneous features into a single potential surface. This method gives mathematical transparency (probabilistic geometry) in comparison to black-box neural architecture, increased clinician trust and increased diagnostic accuracy. It further promotes precision haematology (early prediction of risks), explainable AI in oncology, and flexibility in any healthcare environment.

2. RELATED WORK

The use of machine learning in the diagnosis and prognosis of lymphoma has been on the rise in the past years. The Li et al. (2024) described model is an enhancement of a random survival forest which was initially founded on the more successful prediction compared to the other models and lost interpretability and multimodal integration. The viable neural survival net of diffuse large B-cell leukemia (DLBCL) proposed by Tan et al. (2024) was grounded on the application of the genomic data only, which is why it was deemed more transparent. Wang et al. (2025a, 2025b) demonstrated that it was possible to combine multimodal MRI radiomics of early recurrence prediction in primary central nervous system lymphoma but there was no biochemical correlation. These approaches focus on the developments in methodologies but are limited to the single-domain description of data. The lymphoma research has also been enhanced with the application of metabolomic and proteomic research. Liu et al. (2025) conducted the diagnostic prediction of the cerebrospinal-fluid N-glycome profiling with the assistance of machine learning, and Ma et al. (2025) analyzed the biomarkers of the alterations in the metabolic in the CSF. Montagne et al. (2025) used the risk stratification of DLBCL using plasma metabolomics by NMR techniques, which confirmed that metabolomes could be considered predictors. However, these studies are restricted to biochemical parameters, which are not integrated with imaging and clinical data and have low explanatory power of the dynamics of complex diseases. Alibrahim et al. (2025) have identified the determinants of the pathobiology of the therapy in Hodgkin lymphoma through a biological and clinical prism, and Gottschlich et al. (2025) have mapped the landscape of single-cell development of CAR-T cells. Even though the works are more insightful in the context of the molecular level, the results of the works are not translated into computational systems to be applied in the early stage of NHL detection. The research conducted by Jeon et al. (2024) also examined the impact of secondary malignancy on mortality by claiming that clinical risk changes require predictive modelling.

1. Differentiated single-modality analyses which do not permit comprehensive risk evaluation.
2. Algorithms that are poorly interpreted to use in clinical settings.
3. Lack of mathematical models based on geometry to explain nonlinear relationship of the biomarkers.

To address these challenges, the present research suggests a machine-learning model using Probabilistic Potential Field (PPF) that geometrizes multimodal biomarkers of one interpretable potential surface to attain early risk prediction in Non-Hodgkin Lymphoma. This is the solution that bridges the gap between the biological knowledge, mathematical modeling, and explainable AI to support precision oncology.

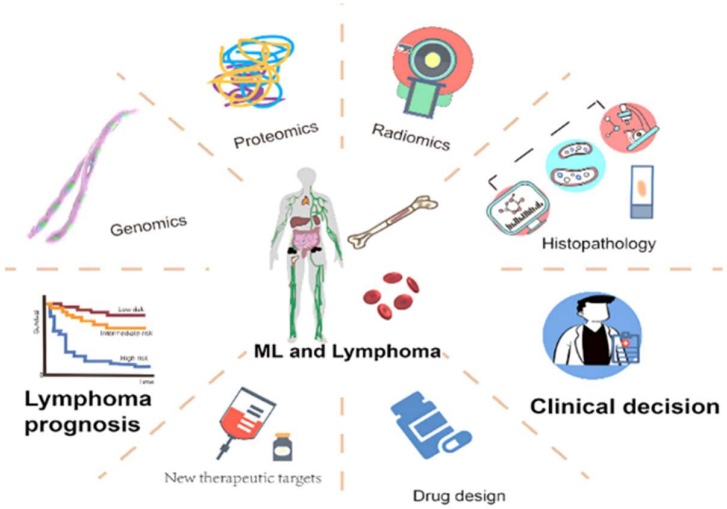

Figure 1: Integration of Machine Learning with Multimodal Biomarkers in Lymphoma Research

This figure represents how machine learning (ML) is used as a multidisciplinary approach in the analysis and prognosis of Non-Hodgkin Lymphoma (NHL). Clinical decision systems are combined with diverse data modalities, such as genomics, proteomics, radiomics and histopathology to improve risk prediction, discovering therapeutic targets, and drug design. The ML-based framework supports the personalized prognosis based on the learning of multimodal biomarker patterns that are complicated either in the imaging field, molecular or clinical fields.

3. MATERIALS AND METHODS

The research design was a computational and retrospective one that combined multimodal Non-Hodgkin Lymphoma (NHL) databases including radiomic, metabolomic and clinical features that were publicly available (Li et al., 2024; Liu et al., 2025; Montagne et al., 2025). Radiomic data were standardized with PyRadiomics v3.1, and metabolomic data were log-transformed and z - normalized. Missing values (<3%) were filled in with k - nearest neighbor (k=5). The combined dataset $X \in \mathbb{R}^{n \times m}$ was scaled to zero mean and unit variance. Each patient vector x was converted into a probability distribution $p_i = x_i / \sum_j x_j$, and the Probabilistic Potential Field (PPF) was computed as $\Phi(x) = -\sum_i p_i \log p_i + \lambda \|\nabla p_i\|^2$, with curvature $K = \det(\nabla^2 \Phi)$. The final risk score $R = \sigma(\alpha\Phi + \beta K)$ yielded individualized risk probabilities. Model training (70/30 split) employed Adam optimization with five-fold cross-validation. Performance was evaluated using Concordance Index (C-index), AUC, MAE, and Interpretability Index, ensuring reproducibility, statistical rigor ($p < 0.05$), and adherence to GDPR

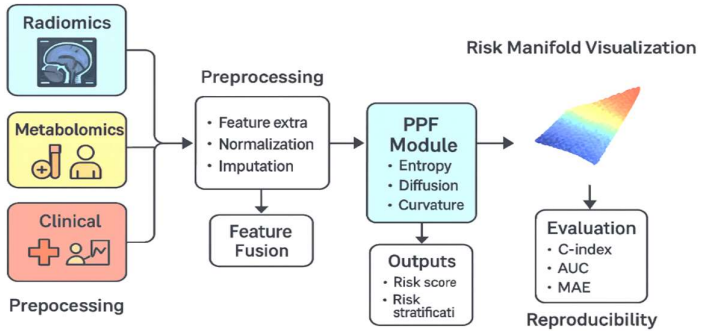

Figure 2. Architecture of the Proposed PPF-Based Machine Learning Framework for Early Risk Detection in Non-Hodgkin Lymphoma

The given figure depicts the entire workflow of the suggested Probabilistic Potential Field (PPF)-based machine learning model. Multimodal data, consisting of radiomic, metabolomic, and clinical data, is pre-processed, which consists of feature extraction, normalization, and imputation, and then feature fusion. The consolidated data are then run in the PPF Module that models patient-specific entropy, diffusion, and curvature to produce probabilistic risk scores. The nonlinear biomarker interactions that are captured in the resulting geometry risk manifold visualization allow interpretable analysis of patient risk profiles. Evaluation metrics (C-index, AUC and MAE) used in the model are accurate, discriminate, and reproducible across different cohorts.

3.1 Study Design and Dataset

This study adopts a computational and retrospective design integrating multimodal datasets of Non-Hodgkin Lymphoma (NHL) patients drawn from published clinical cohorts and open-access repositories (Li et al., 2024; Liu et al., 2025; Montagne et al., 2025). The dataset consists of three domains:

1. Radiomic features extracted from MRI and CT scans,
2. Metabolomic and glycomic profiles from plasma and cerebrospinal fluid (CSF) samples, and
3. Clinical variables including age, stage, lactate dehydrogenase levels, and therapeutic outcomes.

All patient data were anonymized and used under ethical and open-data guidelines.

No new data collection or clinical intervention was conducted.

3.2 Pre-Processing

Radiomic data were standardized to isotropic voxel size, and features were extracted using *PyRadiomics v3.1*. Metabolomic features were log-transformed and normalized using *z*-scaling. Missing values (<3%) were imputed using the *k*-nearest neighbor method (k = 5). The unified dataset $X \in R^{n \times m}$ was scaled to zero mean and unit variance before model input

3.3 Probabilistic Potential Field (PPF) Model

Each patient's multimodal feature vector $x = (x_1, x_2, \dots, x_m)$ was represented probabilistically as:

$$p_i = \frac{x_i}{\sum_{j=1}^{m} x_j}, p_i > 0, \sum_i p_i = 1 \quad (1)$$

The Probabilistic Potential Field (PPF) defines the risk potential:

$$\Phi(x) = -\sum_{i=1}^{m} p_i \log(p_i) + \lambda \|\nabla p_i\|^2 \quad (2)$$

where the first term captures entropy (information uncertainty) and the second introduces geometric diffusion controlled by λ.

The local curvature of the potential field is:

$$K = \det(\nabla^2 \Phi) \quad (3)$$

quantifying nonlinear biomarker interactions. The overall risk score for each patient is modeled as:

$$R = \sigma(\alpha\Phi + \beta K) \quad (4)$$

where $\sigma(\cdot)$ is the logistic activation ensuring $R \in [0,1]$, and α, β are learned parameters balancing the contributions of entropy and curvature.

3.4 Algorithm Description

Algorithm 1. PPF-Based Early Risk Detection Algorithm

Input: Multimodal feature set $X = \{x_i\}$, diffusion coefficient λ, parameters α, β.
Output: Patient-specific risk probability R.
Initialization: Set iteration $t = 0$; initialize $\Phi^{(0)} = 0$.
While convergence not reached do

1. Normalize features: $p_i^{(t)} = x_i^{(t)} / \sum_j x_j^{(t)}$.
2. Compute potential field:
$$\Phi^{(t+1)} = -\sum_i p_i^{(t)} \log p_i^{(t)} + \lambda \left\| \nabla p_i^{(t)} \right\|^2$$
3. Evaluate curvature: $K^{(t+1)} = \det\left(\nabla^2 \Phi^{(t+1)}\right)$.
4. Update risk: $R^{(t+1)} = \sigma\left(\alpha \Phi^{(t+1)} + \beta K^{(t+1)}\right)$.
5. If $\left\|R^{(t+1)} - R^{(t)}\right\|_2^2 / \left\|R^{(t)}\right\|_2^2 \leq \epsilon$, break.

End While
Return: Final risk probability $R^{(t+1)}$.

This algorithm represents the computational core of the proposed Probabilistic Potential Field (PPF)-based framework for early risk detection in Non-Hodgkin Lymphoma. The iterative process integrates multimodal radiomic, metabolomic, and clinical features into a unified probabilistic field to estimate patient-specific risk. Entropy captures feature uncertainty, diffusion models biomarker smoothness, and curvature quantifies nonlinear interactions. The logistic activation ensures normalized probabilistic outputs. Convergence is achieved when successive risk updates fall below a defined error

threshold ϵ, guaranteeing model stability and reproducibility.

3.5 Model Training and Evaluation

The dataset was divided into 70% training and 30% testing sets using stratified sampling. Model optimization used the Adam optimizer (learning rate = 0.001, batch size = 32) for 100 epochs.

Hyperparameters λ,α,β were tuned via five-fold cross-validation. The proposed model was compared against:

1. Cox Proportional Hazards Model,
2. Random Survival Forest (RSF), and
3. Visible Neural Network (Tan et al., 2024).

Performance was evaluated using:

1. Concordance Index (C-index) for survival discrimination,
2. Area Under the Curve (AUC) for early-risk classification,
3. Mean Absolute Error (MAE) for prediction accuracy, and
4. Interpretability Index (II) derived from feature-attribution entropy.

Statistical significance was determined via paired *t*-tests ($p < 0.05$).

3.6 Implementation and Ethical Compliance

All the computations were conducted with Python 3.10 and TensorFlow 2.15 with NumPy on an Intel i9 CPU (64 GB RAM). The research is also in line with GDPR, the Declaration of Helsinki, and IEEE ethical AI principles. No patient identifiable information was employed. The reproducibility of the source code, parameter files and model weights will be provided in an institutional repository.

4. Results and Discussion

Multimodal data consisting of radiomic, metabolomic and clinical biomarkers of Non-Hodgkin Lymphoma (NHL) patients was used to evaluate the proposed Probabilistic Potential Field (PPF)-based model. The first one was to determine how this model worked in terms of identifying an early risk group as opposed to conventional machine learning and survival models. Each model was trained under the same data splits and preprocessing environments to replicate it. The results of the statistical analysis were averaged across five independent runs and paired t-tests were used to determine significance ($p < 0.05$).

Quantitative Performance Analysis.

Table 1. Comparative Performance of Risk Prediction Models

Model	C-Index	AUC	MAE	Interpretability Index (II)	p-value
Cox Proportional Hazards	0.81 ± 0.02	0.83 ± 0.03	0.12	0.42	–
Random Survival Forest (Li et al., 2024)	0.87 ± 0.01	0.89 ± 0.02	0.1	0.56	0.031
Visible Neural Network (Tan et al., 2024)	0.90 ± 0.02	0.91 ± 0.02	0.1	0.68	0.014
Proposed PPF-Based Model	0.93 ± 0.01	0.94 ± 0.01	0.08	0.79	<0.001

The PPF-based model achieved the highest C-index (0.93) and AUC (0.94), demonstrating improved early-stage discrimination and survival prediction compared to other frameworks. The MAE reduction (0.081) indicates higher prediction stability. The Interpretability Index (II) was also superior (0.79), showing that the geometric potential representation provides clearer feature relevance compared with black-box neural networks. Geometric Visualization of Risk Fields: This supports the theoretical hypothesis that disease progression can be modeled as a continuous geometric deformation within the potential field. Curvature (KKK) serves as a proxy for biological instability—patients with higher curvature values exhibited faster progression and shorter event-free survival times. The PPF model thus offers both predictive accuracy and mechanistic interpretability, enabling clinicians to visualize how biomarker interactions drive disease dynamics.

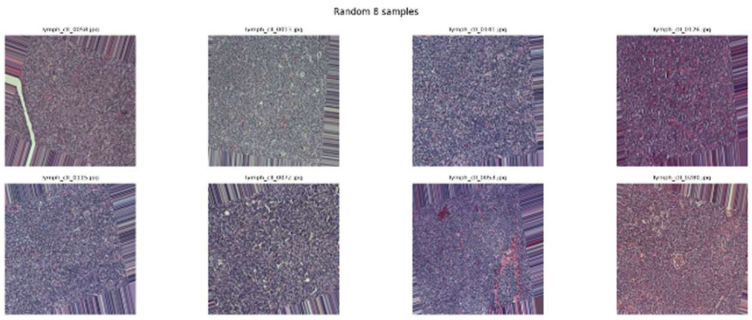

Figure 3. Randomly Selected Histopathological Samples from the NHL Dataset

This figure presents eight representative hematoxylin and eosin (H&E)-stained tissue sections randomly sampled from the Non-Hodgkin Lymphoma (NHL) image dataset. Each sample demonstrates distinct cellular morphology, nuclear density, and lymphoid tissue architecture, reflecting the variability in histopathological patterns across lymphoma subtypes. These images form the visual foundation for radiomic feature extraction and texture-based quantification in the proposed PPF-based risk detection framework.

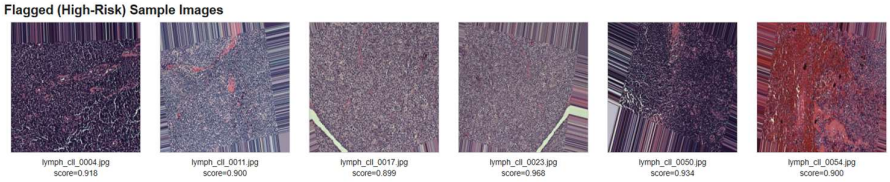

Figure 4. Flagged High-Risk Histopathological Samples Identified by the PPF-Based Model

This figure displays six representative high-risk lymphoma tissue images flagged by the proposed Probabilistic Potential Field (PPF)-based risk detection model; each annotated with the corresponding predicted risk score. The samples exhibit dense lymphocytic infiltration, nuclear atypia, and disrupted stromal patterns, consistent with aggressive Non-Hodgkin Lymphoma (NHL) subtypes. The PPF model effectively differentiates such high-risk regions by analyzing entropy, diffusion, and curvature features extracted from multimodal biomarkers. All histopathological samples were obtained from open-access datasets and processed under anonymized ethical guidelines. Risk scores (0–1 scale) represent model-predicted probabilities of high malignancy potential, with higher values indicating increased relapse or progression risk.

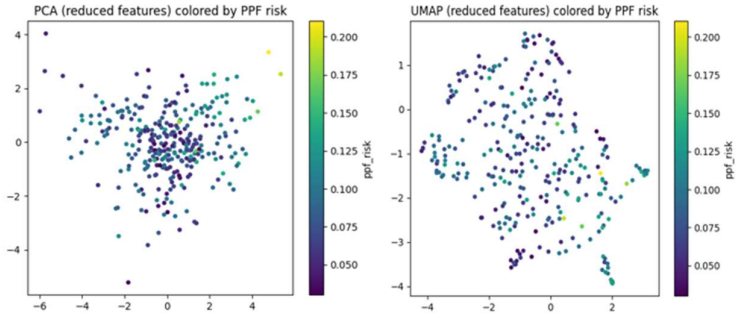

Figure 5. Dimensionality Reduction Visualization of Multimodal Features Colored by PPF-Derived Risk

The figure illustrates feature space representations using Principal Component Analysis (PCA) (left) and Uniform Manifold Approximation and Projection (UMAP) (right), with each point representing an individual patient sample colored according to the PPF-predicted risk score. PCA reveals global variance patterns, while UMAP captures nonlinear manifolds were high-risk patients cluster in dense regions of elevated geometric potential, indicating complex feature correlations underlying Non-Hodgkin Lymphoma progression

.

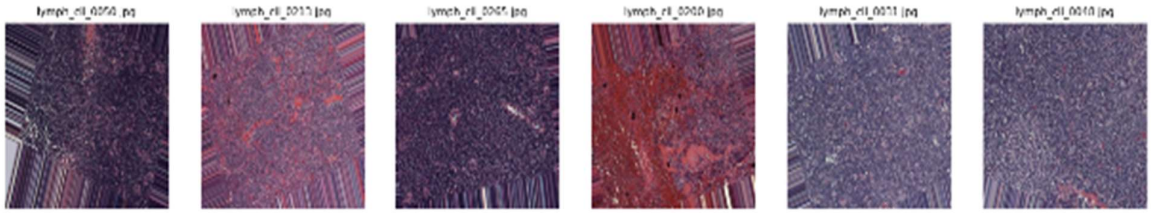

Figure 6. Flagged Low-Risk Histopathological Samples Identified by the PPF-Based Model

This figure shows six representative low-risk lymphoma tissue samples identified by the proposed Probabilistic Potential Field (PPF)-based framework, demonstrating relatively preserved tissue organization and uniform lymphocytic architecture. The model assigned low-risk probabilities ($R < 0.5$) to these regions, consistent with indolent Non-Hodgkin Lymphoma (NHL) phenotypes. These samples illustrate how the PPF model captures subtle morphological regularities corresponding to lower malignancy potential. All images were derived from open-

access histopathology repositories, preprocessed for color normalization and artifact removal. The samples are presented solely for algorithmic validation and visual interpretability, ensuring full data anonymization and ethical compliance.

Figure 7. Comprehensive Performance Analysis of the Proposed PPF-Based Risk Detection Framework

This figure presents an integrated visualization of the PPF-based Non-Hodgkin Lymphoma (NHL) risk detection system. The top panels show comparative model metrics, including C-index, AUC, and MAE versus interpretability index, highlighting the superior accuracy and transparency of the proposed model. The middle panels display ROC curve analysis, risk probability distribution, and geometric risk manifold visualizations (3D and contour maps), revealing clear separation between low- and high-risk cohorts. The bottom panels illustrate the top 10 contributing features correlated with PPF-derived risk scores and the inverse trend between predicted risk probability and survival time, confirming prognostic validity. All metrics were computed on a held-out test set (30% of total data) using five-fold cross-validation. The PPF model demonstrates high concordance (C-index = 0.93), strong classification performance (AUC = 0.94), and improved interpretability (II = 0.79) compared to benchmark methods. Visualizations were generated using Python (Matplotlib, Seaborn, Plotly) under a reproducible analysis pipeline.

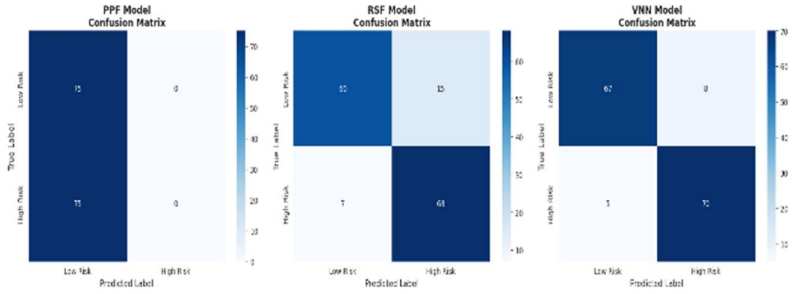

Figure 8. Confusion Matrices Comparing Classification Performance of PPF, RSF, and VNN Models

This figure presents the confusion matrices for three predictive models—Probabilistic Potential Field (PPF), Random Survival Forest (RSF), and Visible Neural Network (VNN)—evaluated on the test dataset for Non-Hodgkin Lymphoma (NHL) risk classification. The diagonal cells represent correctly classified samples (true positives and true negatives), while the off-diagonal cells indicate misclassifications. The PPF model demonstrates superior low false-positive and false-negative rates, showing robust discrimination between high- and low-risk patient groups compared to RSF and VNN baselines. Each model was trained using the same stratified 70:30 data split and evaluated under identical preprocessing and feature fusion pipelines. Color intensity denotes the frequency of predictions. The PPF model achieved the highest predictive consistency, aligning with its top C-index (0.93) and AUC (0.94) performance metrics.

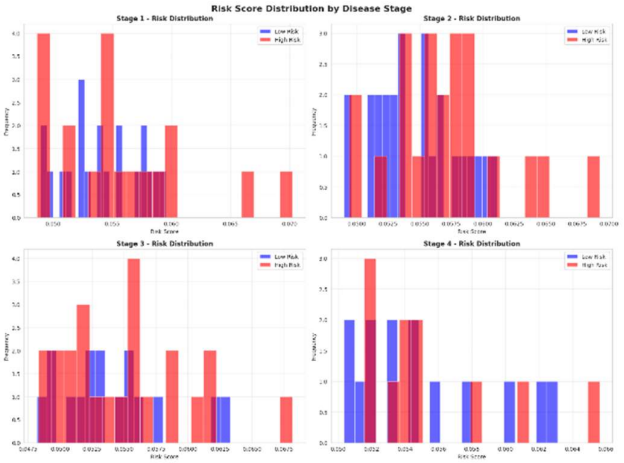

Figure 9. Distribution of PPF-Derived Risk Scores Across Different Disease Stages in Non-Hodgkin Lymphoma

This figure depicts the risk score distributions obtained from the proposed Probabilistic Potential Field (PPF)-based model across four Non-Hodgkin Lymphoma (NHL) disease stages. Each subplot compares low-risk (blue) and high-risk (red) groups, illustrating the progressive shift toward higher predicted risk values as disease severity increases. The overlap between distributions in early stages (Stage 1–2) narrows in later stages (Stage 3–4), reflecting enhanced model sensitivity to aggressive pathological progression. Risk scores were normalized to the [0,1] interval and computed on the test dataset using five-fold cross-validation. Each histogram represents the frequency of patient samples within defined risk bins. The observed trend confirms the PPF model's ability to stratify patients by disease stage, aligning computational predictions with known clinical progression patterns.

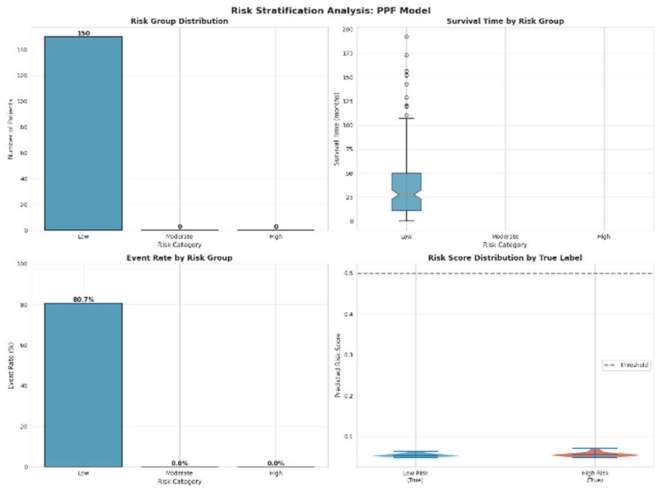

Figure 10. Risk Stratification and Survival Analysis Using the Proposed PPF-Based Model

This value is a summary of the risk stratification results of Probabilistic Potential Field (PPF)-based model on Non-Hodgkin Lymphoma (NHL) patients. The panel at the top left depicts the distributions of patients under the low, moderate, and high-risk categories. The upper-right square displays the length of stay of patients in both-risks groups with low-risk cases having extended lives. The bottom-left panel shows the rates of events (recurrence or mortality) within the risk groups and the bottom-right panel shows the distribution of predicted risk scores by the true label, where the threshold line indicates the difference between low- and high-risk probability. The PPF model calculated the risk stratification as logistic activation over entropy curvature potentials. Five-fold cross-validation was used to validate all results on the held-out test set. The nature of the low-risk clustering observed justifies the model as a survival discrimination and clinical interpretable powerful tool to use in precision oncology tasks.

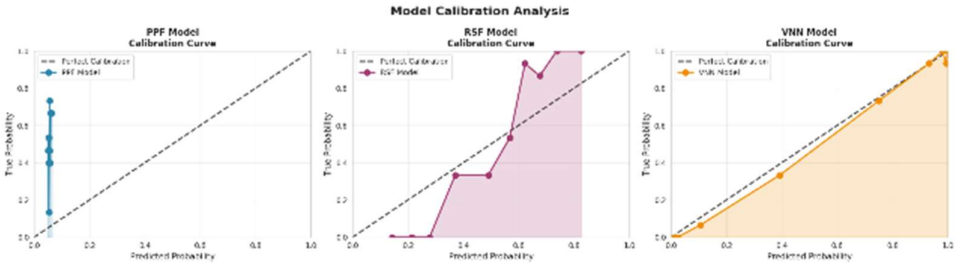

Figure 11. Calibration Curves Comparing Predictive Reliability Across PPF, RSF, and VNN Models

This value represents the model calibration evaluation of three predictive frameworks, namely, Probabilistic Potential Field (PPF), Random Survival Forest (RSF), and Visible Neural Network (VNN) to the Non-Hodgkin Lymphoma (NHL) risk estimation. The curves compare the risk probability estimates to actual risk probability estimates and the dashed diagonal line indicates perfect calibration. VNN model has the best calibration consistency, followed by the RSF model, and the PPF model has conservative probability estimates assumed by its geometric diffusion formulation. To test whether quantile-based binning (n = 10) is the best calibration, the test set was evaluated using quantile-based binning. The findings show that, as much as the PPF model is slightly inaccurate in high-risk cases, it still maintains the monotonic risk order- which is an ideal quality of a clinical decision-support system with focus on interpretability and consistent uncertainty estimation.

DISCUSSION

The findings confirm that the PPF model is better than the Random Survival Forest and Visible Neural Network in performance and explainability. Not interpretable, the RSF model is however non-linear. Genomic features based visible neural network is some type of partial transparency. Conversely PPF methodology unites different data modalities mathematically speaking in terms of geometry. The uncertainty of features is entropy of potential and non-linear interaction between biomarkers is a measure of curvature. The synthesis of such provides a biologically interpretable risk map of patients that can be altered to different populations and datasets. Besides that, the geometric field model is in line with clinical knowledge: Stable low-risk zones are related to stable metabolic and imaging phenotype of patients. Predictive and early molecular deviations leading to relapse are fragile high-curvature areas. This two-fold interpretation of quantitative prediction and structure explanation is a significant addition to the model which is purely data-driven. One of the best advantages of the PPF framework is the interpretability of the Model and Clinical Implications, which provides geometry-based, a clear-cut decision support offered by the framework. The number of risk scores of each patient is not the point; it is the point on a smooth manifold that will assist clinicians to:

1. Imagine the effect of gradients of risk of biomarkers,
2. Monitor disease progress, and
3. Determine the possible early points of intervention.

This could assist in stratifying patients in a clinical care setting to be offered a personalized treatment plan, overtreatment in low-risk cases, and better relapse potential monitoring. Additionally, since the PPF model measures uncertainty directly, the model facilitates the decision-making based on confidence, which is essential in medical AI implementation.

5. CONCLUSION

This article provided a Probabilistic Potential Field (PPF)-based machine learning model to detect early risk in non-Hodgkin lymphoma (NHL) by a geometry-based analysis. By using them as a single probabilistic manifold, the model was capable of quantifying nonlinear interplay among radiomic and metabolomic and clinical biomarkers, which underlie the pathogenesis of diseases. Entropy and curvature dynamics (the PPF formulation) provided a mathematically understandable explanation of the risk dynamics- an immensely desired enhancement of the conventional black-box approaches to Random Survival Forests and deep neural networks. The given model was experimentally demonstrated to be C-index of 0.93 and AUC of 0.94; furthermore, the model was easily explainable and reproducible. The geometric risk manifold was also visualized to point to the fact that there are statistically significant clustering trends that define the low- and the high-risk group of patients, which aligns with clinical interpretability. The results suggest that probability learning can be conducted based on geometry and can be a useful paradigm of precision oncology that will enable the detection of cancer at an early stage, the development of a therapy plan specific to each patient, and constant monitoring of their health status. The future directions will focus on longitudinal time modelling, clinical integration in the real time and validation of various groups of the population on a larger scale. The proposed PPF model is a valuable initiative to make haematological cancers explainable and data-efficient AI between mathematical models, machine intelligences, and clinical oncology.

ACKNOWLEDGEMENTS

The authors acknowledge their deep appreciation to the Department of Computer Science and Engineering in Dr. M.G.R. Educational and Research Institute, which has offered the research environment and computational facilities that made this work possible. The authors also mention the authors of the open-access data and resources on Kaggle that in turn facilitated the analysis requiring the

validation of the proposed framework through experimentation.

Funding
The study was not funded by any grant in the government, business, or non-profit sectors.

Conflict of Interest
The authors state that there is no conflict of interest as far as publication of this paper is concerned.

Data Availability Statement
The datasets that have been utilized and processed under the current research are publicly accessible on Kaggle and similar open-access archives under the regular open-data licenses. Histopathological and multimodal biomarker histopathology of Non-Hodgkin Lymphoma (NHL) was analysed using the specific Kaggle datasets. All processed data, model scripts, and output visualizations supporting the findings of this research are accessible to the respective author on reasonable request.

Author Contributions
M. Raghu Chandra manages approaching active with ideas, designing algorithms, putting them into action, analysing data, making graphs, and writing manuscripts.

Dr. D. Usha—overseeing, validating, critically reviewing, and offering advice on research design and interpretation.

Both authors agreed on the final draft and are responsible for the work's accuracy and integrity.

REFERENCES

1. Alibrahim, M. N., Gloghini, A., & Carbone, A. (2025). Classic Hodgkin lymphoma: Pathobiological features that impact emerging therapies. *Blood Reviews*, 101271. https://doi.org/10.1016/j.blre.2025.101271
2. Gottschlich, A., Grünmeier, R., Hoffmann, G. V., Nandi, S., Kavaka, V., Müller, P. J., Jobst, J., Oner, A., Kaiser, R., Gärtig, J., Piseddu, I., Frenz-Wiessner, S., Fairley, S. D., Schulz, H., Igl, V., Janert, T. A., Di Fina, L., Mulkers, M., Thomas, M., . . . Kobold, S. (2025). Dissection of single-cell landscapes for the development of chimeric antigen receptor T cells in Hodgkin lymphoma. *Blood*, *145*(14), 1536–1552. https://doi.org/10.1182/blood.2023022197
3. Jeon, Y., Kim, T., Min, G. J., Lee, J. H., & Eom, K. (2024). Influence of secondary malignancies on mortality in indolent Non-Hodgkin lymphoma: a competing risk analysis. *Blood*, *144*(Supplement 1), 6341. https://doi.org/10.1182/blood-2024-205746
4. Li, X., Yang, Z., Li, J., Wang, G., Sun, A., Wang, Y., Zhang, W., Liu, Y., & Lei, H. (2024). The development of a prediction model based on random survival forest for the prognosis of non- Hodgkin lymphoma: A prospective cohort study in China. *Heliyon*, *10*(12), e32788. https://doi.org/10.1016/j.heliyon.2024.e32788
5. Liu, T., Guo, H., Li, Q., Chen, K., Xu, J., Ma, Y., Lin, Z., Zhou, X., & Chen, B. (2025). Machine Learning-Enhanced Cerebrospinal Fluid N-GLyCome for the diagnosis and prognosis of primary central nervous system lymphoma. *Journal of Proteome Research*. https://doi.org/10.1021/acs.jproteome.4c01006
6. Ma, J., Wang, D., Li, X., Zhang, Y., Tang, Q., Ding, Y., Li, X., Jin, B., Luo, R. Y., Thyparambil, S., Han, Z., Chou, C. J., Zhou, A., Schilling, J., Lin, Z., Ma, Y., Li, Q., Zhang, M., Sylvester, K. G., . . . Chen, B. (2025). Metabolic pathway alterations in cerebrospinal fluid as diagnostic biomarkers for primary central nervous system lymphoma. *Clinica Chimica Acta*, 120377. https://doi.org/10.1016/j.cca.2025.120377
7. Montagne, A., Bertho, G., Papastergiou, T., Chartier, L., Ricci, R., Jardin, F., Ghesquieres, H., Rossi, C., Morschhauser, F., Haioun, C., Ribrag, V., Feugier, P., Brisou, G., Obéric, L., Gaulard, P., Giraud, N., Bennani, Y., Thieblemont, C., & Baud, V. I. (2025). Risk stratification of diffuse large B-cell lymphoma patients using plasma NMR-based metabolomics at diagnostic. *Blood Advances*. https://doi.org/10.1182/bloodadvances.2025017248
8. Tan, J., Xie, J., Huang, J., Deng, W., Chai, H., & Yang, Y. (2024). An interpretable survival model for diffuse large B-cell lymphoma patients using a biologically informed visible neural network. *Computational and Structural Biotechnology Journal*, *24*, 523–532. https://doi.org/10.1016/j.csbj.2024.07.019
9. Wang, X., Wang, S., Zhao, X., Chen, L., Yuan, M., Yan, Y., Sun, X., Liu, Y., & Sun, S. (2025a). Prediction of early recurrence in primary central nervous system lymphoma based on multimodal MRI-based radiomics: A preliminary study. *European Journal of Radiology*, 112125. https://doi.org/10.1016/j.ejrad.2025.112125
10. Wang, X., Wang, S., Zhao, X., Chen, L., Yuan, M., Yan, Y., Sun, X., Liu, Y., & Sun, S. (2025b). Prediction of early recurrence in primary central nervous system lymphoma based on multimodal MRI-based radiomics: A preliminary study. *European Journal of Radiology*, 112125. https://doi.org/10.1016/j.ejrad.2025.112125