A Hyperbolic Evidential Learning Approach For Hierarchical Consistency And False-Negative Minimization In Lung Carcinoma Subtyping

by

Shoaib Ahmed
22301715
Zahin Anan Hyder
21301052
Tahmid Islam
20301028
Bless Peter Biswas
24241195
Ahmed Fahim Kabir
22301067

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science and Engineering

Department of Computer Science and Engineering
BRAC University
January 2026.

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Student Name: Shoaib Ahmed

22301715

Student Name: Zahin Anan Hyder

21301052

Student Name: Tahmid Islam

20301028

Student Name: Bless Peter Biswas

24241195

Student Name: Ahmed Fahim Kabir

22301067

Approval

The thesis/project titled "A Hyperbolic Evidential Learning Approach For Hierarchical Consistency And False-Negative Minimization In Lung Carcinoma Subtyping" submitted by

1. Shoaib Ahmed (22301715)
2. Zahin Anan Hyder (21301052)
3. Tahmid Islam (20301028)
4. Bless Peter Biswas (24241195)
5. Ahmed Fahim Kabir (22301067)

of Fall, 2025 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science in Spring 2026.

Examining Committee:

Supervisor:
(Member)

Md. Aquib Azmain

Senior Lecturer
Department of Computer Science and Engineering
Brac University

Thesis Coordinator:
(Member)

Dr. Md. Golam Rabiul Alam

Professor
Department of Computer Science and Engineering
Brac University

Head of Department:
(Chair)

Dr. Sadia Hamid Kazi

Associate Professor
Department of Computer Science and Engineering
Brac University

Ethics Statement

The authors of this thesis would like to state that this work is an original piece of work resulting from honest and extensive research. All sources of information, data and concepts used in this research have been properly acknowledged and cited. This work has not been copied, plagiarized or misrepresented in any way. This research has not been submitted to any university or institution in whole or in part for the purpose of awarding any degree or any other award. The authors take full responsibility for the integrity, accuracy and authenticity of the research presented.

Abstract

Lung carcinoma is among the most prevalent and fatal cancers across the globe, and an accurate clinical classification of the disease subtypes is necessary to plan the treatment.Computed tomography (CT) imaging has become the primary tool in the diagnosis of lung cancer, and automated subtyping with the help of deep learning continues to present significant obstacles. Most of the models available consider subtyping as a flat classification problem and do not consider the hierarchical nature of relationships between the categories of lung carcinoma. Furthermore, these models tend to give confident predictions, but without much indication of reliability of these predictions, thus restricting their application in practical clinical practice. This thesis proposes a hierarchical subtyping of lung carcinoma utilizing hyperbolic evidential learning. The suggested methodology simulates the inherent hierarchy of the lung cancer subtypes in a hyperbolic space, which promotes predictions that are consistent with clinical experiences. Parallel to this, predictive uncertainty is estimated using evidential learning, which enables the model to produce lower confidence when the information in its possession is unclear or insufficient.This combination assists to enhance the consistency as well as reliability of the outputs of the model. The performance of the proposed method is assessed with the use of CT scan dataset to assess the classification performance and the quality of uncertainty estimation. Additionally, four backbone architectures (EfficientNet-B0, DenseNet121, ViT-B/16, and PVT-V2-B0) are evaluated, and Vision Transformer (ViT) performs best among the hyperbolic models, achieving a False Negative Rate of 0.0029 (approximately 99.7 percent sensitivity). The findings suggest that the use of a hierarchical structure and recognizing uncertainty causes more reliable subtype predictions and the improvement of more consistent confidence measures. In general, the present study is expected to encourage more informed clinical decisions and make a contribution to the safe and practical application of artificial intelligence in the diagnosis of lung carcinoma.

Keywords: Lung carcinoma, CT imaging, hierarchical classification, hyperbolic learning, evidential deep learning, uncertainty estimation, medical image analysis

Dedication

This work is dedicated to everyone who inspired and guided us along our journey of research and development. We offer our deepest gratitude to our families, whose constant support, belief, and encouragement have been our greatest source of strength. We are especially thankful to our professors and teachers for their advice, insight and faith in our abilities. Their contribution has influenced our academic and personal progress in addition to this research. This dedication also pays tribute to everyone who pushed us to aim for success, creativity and knowledge.

Acknowledgement

To begin with, we would want to express our deepest gratitude to the Almighty, whose blessings allowed us to finish this thesis without facing any major obstacles. We sincerely thank our supervisor for his crucial guidance, continuous support and constructive feedback throughout this research journey. Finally, we are grateful to our instructors, peers and families for their support and inspiration throughout this work.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

Lung cancer is the most lethal cancer in the world.Low survival rates and late-stage diagnosis rates are particularly problematic in underdeveloped regions like Bangladesh, where environmental pollutants and a lack of suitable screening facilities aggravate the problem. Early and accurate diagnosis is crucial, but conventional imaging procedures and even current AI-based diagnostic devices are not reliable or clinically trustworthy. Traditional deep learning systems use Euclidean vector spaces, which cannot adequately support the inherently hierarchical structure of medical taxonomy, including relationships between malignant lung carcinoma and its subtypes. Also,these geometric constraints can lead to medically inconsistent predictions, including a malignant tumor being false predicted as benign, or pathological subtypes being confused both of which are associated with diagnostic errors and decreased confidence in the clinician. The issue of predictive overconfidence also cannot be overestimated. Softmax outputs are commonly the default output of the standard neural networks, and give point-estimate probabilities, but do not estimate uncertainty. This kind of overconfidence is perilous in medical decision-making subjected to high stakes, where false-negative may be very expensive in terms of intervening late and avoidable death. To overcome these two shortfalls hierarchical inconsistency and uncalibrated uncertainty this thesis offers a single *Hyperbolic Evidential Network* of lung carcinoma subtyping that is well-behaved and clinically interpretable. The designed architecture is designed on a Poincaré ball model of the hyperbolic geometry which is naturally exponential with respect to the radius and thus is very suitable to encode hierarchical structures. It allows classification that preserves ontological relationships (i.e., between malignancy and its subtypes e.g. Adenocarcinoma, Squamous Cell Carcinoma). The model applies structural, medical consistency by incurring classification with hyperbolic distances to learned prototype embeddings. Simultaneously, it has an evidential deep learning mechanism that produces Dirichlet concentration parameters $\boldsymbol{\alpha}$, which express both class belief and predictive uncertainty. This will enable clinicians to distinguish low and high confidence predictions, which facilitate safe clinical triage and decision-making. Training is based on a purely evidential loss that combines Bayesian risk and Kullback Leibler (KL) divergence regularization to encourage against overconfidence, and does not use cross-entropy. `RiemannianAdam` is used to optimize the curved

space and keep the hyperbolic geometry intact when learning on the curved space through gradient based optimization.
In the assessment, clinically significant safety indicators, especially the false-negative rate (FNR) and the hierarchy violation rate are taken into consideration as long as competitive accuracy is maintained to ensure compatibility with the real-world diagnostic needs. In general, the research can fill a critical gap in the field of oncology diagnostics by integrating hierarchy-aware learning based on hyperbolic representation and calibrated learning of evidential uncertainty. The final system should be interpretable, reliable, and deployable to provide early and accurate subtyping of lung cancer, and particularly in underserved and resource-constrained healthcare environments.

1.2 Motivation

An accurate and on-time subtyping of lung carcinoma is a fundamental concern of cancer therapy and has far-reaching implication on patient prognosis, therapy selection and survival. This diagnostic task is greatly hampered by systemic constraints in most areas, particularly in the developing world such as ours where access to specialised thoracic pathologists, access to advanced immunohistochemical and genomic profiling, and access to high-quality images is restricted. Due to these restrictions, it is more challenging to make a conclusive diagnosis, which frequently necessitates the subjective interpretation of radiography.These limitations make it more difficult to provide a definitive diagnosis, which often forces the subjective interpretation of radiology. This can lead to inconsistent diagnoses, improper treatment initiation, and ultimately worse than ideal clinical outcomes.However, there are still several significant obstacles that need to be addressed before these methods can be considered trustworthy clinical tools and move beyond laboratory standards. Structured medical knowledge that is taxonomic is conceptually incompatible with the standard deep learning design. The intrinsic hierarchical structure of pathological ontologies, such as the clear categorical distinction between benign and malignant tissues and the additional classification of malignancies into histological subtypes, cannot be encoded by them because they operate in Euclidean vector spaces. This geometric weakness may provide logically incoherent predictions that violate medical logic, such as confusing subtypes along hierarchical lines, resulting in a loss of clinician trust and the usefulness of AI as a trusted diagnostic companion.Another issue is equally vital and that is the problem of predictive overconfidence. Standard models make use of softmax activations which provide point-estimate probabilities that do not measure the underlying uncertainty of a particular prediction. This lack of quantification of uncertainty is not just a technical weakness in the high-stakes environment of oncology where a false-negative diagnosis, or, more precisely, the inability to detect a malignant lesion, can result in death. Clinicians are deprived of an indication of model confidence, they cannot tell whether a prediction is strong or conjectural, which is highly constraining in terms of its usefulness in triage and critical decision-making.Therefore, there is a great and unmet need to develop a new paradigm for AI-assisted diagnosis that strikes a compromise between clinical pragmatism and computational intelligence. This research is motivated by the need to create a system that is not just accurate but also ontologically consistent and cog-

nisant of uncertainty. It uses a novel paradigm based on hyperbolic geometry and evidentiary deep learning to directly address the two issues of hierarchical incompatibility and unreliable conviction.The goal of the proposed Hyperbolic Evidential Network is to make AI clinically reliable. Because the model embeds data in hyperbolic space, it naturally respects the hierarchical taxonomy of lung cancer. It uses an evidential output to provide precise, calibrated predictions of risk. This integrated approach will ensure that the predictions are logically consistent with medical knowledge, lower the probability of false-negative errors, and give doctors interpretable confidence metrics. The ultimate goal is to create a robust, comprehensible, and scalable decision-support system that can improve the quality of diagnostics, expedite patient examinations, and expand access to high-quality cancer subtyping in the most under-resourced settings, such as our nation, where such an innovation is most needed.

1.3 Problem Statement

Common deep learning hypernet architectures of lung cancer subtyping are structurally mismatched with the hierarchical structure of biomedical knowledge, especially in medical areas where disease categories follow tree-like taxonomies with both parent-child relationships. The currently favored neural network design, Euclidean embedding schemes, views classes as flat, discrete objects in a vector space, thus creating logical errors like cases of hierarchy, such as a model predicting the particular subtype adenocarcinoma without identifying its supertype of malignancy. The result of such a mismatch undermines the ability of the model to learn genuine biological associations and make those superficial shortcuts that are based on the artifact of data and not substantially meaningful features. Moreover, the softmax based classifiers, commonly used in these models, have a tendency to be overconfident when presented with ambiguous or out-of-distribution data, thus, increasing the likelihood of a catastrophic false-negative event in a safety critical task such as oncologic screening, where any omission of a malignant case may produce a severe clinical consequence. These concerns are further compounded by the fact that there is no strong uncertainty quantification and therefore the models are incapable of marking out untrustworthy forecasts, limiting its viability and preventing smooth incorporation into clinical diagnostic procedures. Resolving this geometric discrepancy is essential to enable the creation of AI systems, which are logically consistent, uncertainty-aware, and aligned with clinical taxonomies, which eventually leads to the minimization of diagnostic errors and improvement of patient safety.

Research Questions

The following research questions guide this study:

- **RQ1:** How can hierarchical relationships in lung cancer subtypes be effectively modeled within a deep learning framework?
- **RQ2:** How can predictive uncertainty be quantified to distinguish reliable diagnoses from ambiguous cases?
- **RQ3:** To what extent does hyperbolic representation learning improve hierarchical consistency compared to Euclidean approaches?
- **RQ4:** What is the impact of uncertainty-aware learning on reducing the False-Negative rate while maintaining competitive diagnostic accuracy?

Therefore, these restrictions really bring down the confidence and reliability in the clinical diagnosis. In our thesis, we attempt to address these issues by assembling a framework that explicitly addresses hierarchical structure and factors in predictive uncertainty, which makes the predictions geometrically sound and safe for clinical use by minimizing False Negative cases.

1.4 Research Objective

Fundamentally, the main goal of this research is to build a *Hyperbolic Evidential Deep Learning (Hyp -EDL)* framework which is able to classify lung carcinoma in a hierarchical manner considering uncertainty based on CT imaging data.

Specifically, this study aims to achieve the following objectives:

- The aim would be to create a hyperbolic embedding space with the Poincare ball model that reflects the hierarchical relationships between categories and subtypes of benign and malignant lung cancer.
- We are trying to merge Evidential Deep Learning in order to model both Aleatoric (Data driven) and Epistemic (Model driven) uncertainty in order to identify ambiguous and out-of-distribution clinical cases.
- We are going to implement and evaluate some deep learning backbone architectures which include EfficientNet-B0, DenseNet, Vision Transformers (ViT), and PVT-V2-B0 in the hyperbolic evidential framework we developed for our research.
- To reduce the number of safety-critical diagnostic errors, we are concentrating on reducing the false-negative rate (FNR), because in that way, we can be more confident that malignant cases are being reliably spotted or not.

- To assess the proposed framework using clinically relevant performance metrics, including accuracy, sensitivity, uncertainty calibration, and hierarchy violation rate, ensuring alignment with real-world diagnostic requirements.

Furthermore, the present paper suggests a Hyp-EDL structure that would encompass the hierarchy of lung cancer and also estimate uncertainty by way of CT pictures. Moreover, it evaluates several robust backbones such as EfficientNet-B0, DenseNet, ViT and PVT-V2-B0 in a single configuration. Above all, it is aimed at minimizing false negatives in order to minimize the likelihood of missing malignant cases. Lastly, the framework is tested with clinical measures that are practical such as accuracy, sensitivity, calibration and the rate of hierarchy violation to remain near actual diagnosis requirements.

1.5 Methodology in Brief

This thesis proposes and analyzes a Hyperbolic Evidential Learning system to subtype lung carcinoma that explicitly aims at hierarchical consistency and minimization of false-negatives. We pre-process an external image dataset that is publicly available, which is split into training, validation, and test sets and we normalize and resize the input pipeline. Furthermore, since the issue of class imbalance may favor majorities and exaggerate the error rate of missing cancers, we utilize a balanced sampling approach with weighted sampling because minority subtypes are adequately represented in the optimization. Simultaneously, we code a clinically significant hierarchy in which each subtype label is tagged by a parent category (Malignant vs. Benign/Normal) which allows the downstream evaluation of violations of the desired medical taxonomy in predictions.
Then we introduce a Hyperbolic Evidential Network, with a representation space the Poincaré ball manifold, as this has the advantage of exhibiting such a hierarchical embedding. Alternatively, concretely, deep image features are obtained using a pretrained backbone (tested on numerous architectures, such as EfficientNet-B0, DenseNet, ViT, and PVT-V2-B0), and projected into a tangent space before being mapped onto the hyperbolic manifold through exponential mapping. It is classified by doing hyperbolic computations between each sample embedding and a set of learnable class prototype embeddings, which are required to lie on the manifold; thus, the decision about subtypes is guided by geometry which is more hierarchical structure aware than a Euclidean embedding. At the same time, the network, rather than softmax probabilities, produces Dirichlet concentration parameters (alpha), using distance-based scores as input, transformed to non-negative evidence using a softplus, and hence, the model produces both class belief and a direct uncertainty estimate using total evidence.We use RiemannianAdam to optimise hyperbolic components because these parameters reside on a curved manifold. To control generalisation, we employ a learning-rate schedule, early halting, and a tiered training approach (warmup on frozen backbone, then full fine-tuning). Last but not least, we concentrate more on clinical safety in the evaluation process. We report the standard accuracy and recall values, but we pay particular attention to the false-negative rate (FNR) on missed malignancies and the hierarchy violation rate on medically inconsistent predictions. To validate calibrated decision support,

we also test FNR at confidence levels and uncertainty separation between correct and incorrect predictions. The same procedure is also used to train a comparable Euclidean evidentiary baseline in order to determine the advantages of hyperbolic geometry in terms of performance and safety.

1.5.1 Workplan

This thesis will be implemented as an experimentally controlled staged workplan which is clearly consistent with the two main clinical aims of the study: the requirement to subtype lung carcinoma hierarchically, and the requirement to reduce safety-critical false negatives. In order to provide each instance a fine-grained subtype label and a parent malignancy label (such as Malignant vs. Benign), we will first formalise the clinical label ontology by defining the parent-child relationships between classes and entering them directly into the dataset interface. Practically, a hierarchy mapping will be employed to identify subtype categories(e.g., Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma and Normal) and their parent category, so as to be able to downstream verify their hierarchical consistency both in training and in evaluation.Then we will build a data preparation and loading pipeline that can be reproduced into reality and has the resources to manage the variability and constraints of reality. Standardization of images will be done by a unified preprocessing procedure and enhanced with clinically plausible augmentation to enhance inter-acquisition generalization, i.e. resizing, random crop, horizontal flip, rotation and controlled color perturbation, and then convert them to tensors and normalize them.Furthermore, since in medical data distributions of subtypes are usually skewed, to ensure balanced sampling will be applied through weighted sampling to avoid the systematic underrepresentation of minor classes during the optimization process.Simultaneously, train/validation/test splits will be operationalized by the use of special loaders, and all reported results will be calculated on held-out test data, and under the same experimental protocol.After data preparation, the main model development process will provide the suggested Hyperbolic Evidential Learning model and its Euclidean counterpart on controlled comparison. The feature embeddings in the Euclidean evidential baseline will be matched with learnable class prototypes, and distances transformed into evidence and subsequently converted to Dirichlet concentration parameters (i.e. evidence $+1$) which will replace the softmax point estimates with uncertainty-sensitive evidential outputs.Learning objectives instead of cross-entropy will be a pure evidential loss, a combination of Bayesian risk and KL-divergence regularization to a uniform Dirichlet prior, with an annealing process that progressively intensifies regularization, and which is supposed to dishearten predictive overconfidence in high-stakes tasks.Additionally, Riemannian optimization will be used (via RiemannianAdam) to preserve geometric validity when updating the gradient because the hyperbolic version will require optimization on a curved manifold.Lastly, the focus of evaluation will include clinically meaningful safety and consistency measures, primarily false-negative rate (FNR) and hierarchy violation rate, consistency with accuracy as well, and the results will be compared across backbones systematically to confirm robustness and minimize architecture-specific inference.

1.5.2 Workflow

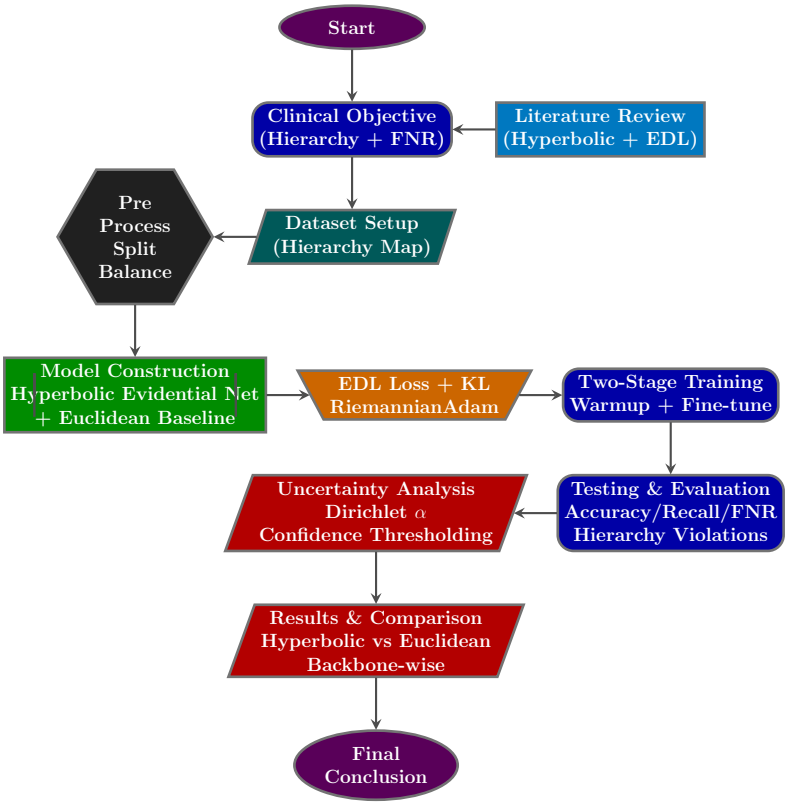

Figure 1.1: Workflow of the proposed Hyperbolic Evidential Learning framework for hierarchy-consistent lung carcinoma subtyping and false-negative minimization.

1.6 Scopes and Challenges

1.6.1 Scope

The thesis examines a *Hyperbolic Evidential* Learning model of subtyping lung carcinoma by two clinical goals, namely an enforcement of hierarchical consistency in predictions and reduction of false-negative during malignancy-related decision-making. In particular, the research constructs an integrated model to integrate imaging-based representations in hyperbolic space in such a way that the classifier is more inclined towards ontological relationships between diagnostic categories (such as the distinction between malignant and non-malignant findings and further subdivision of malignancies into histological subtypes). The proposed method further substitutes the usual point-estimate softmax output with a manifested formulation that generates Dirichlet concentration parameters such that they can be used to explicitly quantify uncertainty so that they can be used to support safer triage and clinician-friendly interpretability.

In addition, the work is limited in terms of scope to learning on publicly available medical imaging datasets (e.g., CT and/or X-ray modalities, based on the availability of datasets) with realistic variability of scanning conditions and patient characteristics. As a result, the experimental protocol focuses on splits cross-generalizability and domain shifts cross-generalizability in the presence of heterogeneous acquisition environments. Additionally, the paper also assesses the methodology on a typical feature backbones, paired with the core contribution centered on the geometric and evidential learning units and not on custom data collection or commercial clinical systems. Lastly, the system is made deployment friendly such that it is possible to stay within limited computational resources with efficient training and inference settings.

1.6.2 Challenges and Limitations

First, geometry-aware and uncertainty-aware methods hold potential, there are a number of significant barriers to the adoption of clinically reliable lung carcinoma subtyping. To begin with, medical imaging datasets gathered within resource-constrained environments are frequently subjected to noisy acquisition as well as annotations, and the ensuing labeled samples are typically relatively few and unbalanced in terms of classes. As a result, the models need to learn to have strong decision boundaries without being overfitted by the idiosyncrasies of the dataset. At the same time, subtype distinctions can be visually faint and in some cases ambiguous even to a skilled radiologist, hence, introducing label noise into the training procedure and can give brittle performance unless carefully handled. Reduction in false negatives and an increase in false positives is therefore a long term trade-off that needs to be maneuvered with the loss function design, calibration and the setting of a threshold.

Second, even though hyperbolic geometry is more compatible with hierarchies of data, optimisation on curved manifolds is computationally problematic, such as being sensitive to curvature parameters, can be unstable during training, and can collapse to representational minima when prototypes and embeddings are not regularised. In addition, evidential learning also presents novel sources of failure: a lack of uncertainty calibration can cause the model to be either too confident on out-of-distribution examples or too unconfident on standard clinical examples, and thus restrict its applicability. In line with this, uncertainty evaluation should not only incorporate numerical scores of confidence but also how they are related to predictive accuracy and safety of errors especially when it comes to omission of malignant lesions.

Third, the barriers of clinical integration are still nontrivial. Clinicians might be reluctant to use systems that are incapable of supporting the outputs; when predictive performance is strong, explainability mechanisms, including saliency-based visualisations and clinician friendly confidence measures, are needed to promote trust and adoption. Also, privacy and governance requirements limit the full real-world testing without the approvals of the institutions, which may slow down the external validation. Lastly, the performance might differ significantly based on the hospital environment which might have vintage hardware and limited infrastructure; thus, real-time performance might require optimisation methods, lightweight inference pipelines and workflow compatible reporting format .

Chapter 2

Literature Review

2.1 Preliminaries

From the reviewed studies, we have summarized relevant research papers by following a structured approach to ensure clarity and consistency in analysis. Each study has been reviewed based on specific criteria, including its research objectives, the core problem it aims to address, the methodologies and techniques applied, as well as the challenges or issues encountered during the study. Additionally, we have noted the limitations acknowledged by the authors and identified existing research gaps. This systematic breakdown allows for a comprehensive understanding of prior work and helps in establishing the foundation for the current study.

2.2 Review of Existing Research

Dunn et al. (2023) focus on the development of an automated system of lung cancer subtypes reconstruction, which is grounded on CT scan and radiomic analysis. The main purpose of the study is to reduce the application of the manual interpretation method by radiologists and provide a non-invasive instrument of establishing lung cancer histological subtypes. To address this challenge, in the automated segmentation of lung tumors in CT images, the authors implement a deep learning-based segmentation network known as incremental multiple resolution residual network (iMRRN) and subsequently identify radiomic features. It subsequently categorizes these features using the assistance of a number of machine learning frameworks, including support vectors machines, k-nearest neighbors, ensemble tools, decision trees, naive Bayes, and a small neural network.The study addresses the problem of tumor segmentation and subtype classification by combining deep learning with the traditional supervised classifiers. The findings indicate that the support vector machine performed optimally, with the classification accuracy of 92.7% and the AUC of 0.97 in differentiating adenocarcinoma, small cell carcinoma, and squamous cell carcinoma. Nevertheless, the authors state that there were problems associated with inaccurate segmentation of non-tumor areas, and some of them needed the help of radiologists partially. Class imbalance and the use of one set of CT are also constraints of the study. The authors identify the gap in research by noting that more generalized and fully automated systems are required to be applied in varied datasets and clinical environments in a reliable manner.[13]

Fayjie et al. (2024) research the ability of uncertainty estimation to enhance deep learning models applied to lung carcinoma classification in digital pathology using digital pathology images, particularly when the test statistics are not similar to those of the training statistics. The primary aim of the study is to learn whether uncertainty-aware models are capable of creating more reliable predictions when there is a shift in the real-world data. To solve this issue, the authors test various methods of uncertainty estimation, such as Monte Carlo dropout, deep ensembles, and few-shot learning, on popular deep learning models, such as ResNet50, VGG19, DenseNet121, Xception, and EfficientNetB0. The research explores several situations of dataset shift, such as cancer subtype change, tissue type change, and imaging source change. The models under normal (in-distribution) conditions perform well in classification and reported accuracies tend to be between about 85% and more than 95% depending on the model and the task performed. Nevertheless, the findings reveal that model accuracy significantly decreases with the addition of dataset shifts, and numerous deep learning models are overconfident even when they are wrong. In such difficult conditions, the ensemble-based and few-shot learning methods enhance the robustness and awareness of uncertainty, but more resources are required in terms of computer computations and training time. Besides, the research primarily deals with digital pathology images and assesses only a small number of uncertainty estimation methods, which could limit generalizability to other medical imaging fields. The authors thus determine a research gap in the creation of uncertainty-aware models which are computationally efficient and trustworthy in a variety of and unseen clinical data distributions.[16]

Kowsari et al. (2020) present the Hierarchical Medical Image Classification (HMIC) framework to enhance the classification of medical images based on the hierarchical nature of diseases and subtypes. The primary purpose of the research is to overcome the shortcomings of the conventional flat classification systems that tend to disregard the links between the high-level categories of diseases and their smaller subdivisions. To address this issue, the authors suggest a two-stage deep learning model that is founded on convolutional neural networks in which the first model categorizes the images by general disease types and the second model categorizes the severity levels of the disease. The method is assessed on histopathological biopsy images to classify normal tissue, environmental enteropathy and celiac disease and then the severity is graded using the Marsh scores. The experimental findings indicate that the HMIC framework performs better than the typical flat CNN models with the highest parent-level classification accuracy of more than 90% and child-level accuracy with F1-scores in the interval of 86–90%. In spite of these advancements, the research states that there are difficulties connected to computational complexity owing to huge image patches and widespread preprocessing criteria. Besides, the method is restricted to two-level hierarchy, and it is evaluated on one medical field. The model also gives less attention at the overall classification accuracy and fails to explicitly test in false-negative error, which is essential in medical diagnosis. Subsequently, the authors emphasize the necessity to develop more flexible hierarchical models, which will be applicable to a wider range of diseases as well as enhance reliability and minimize the clinically risky misclassifications.[3]

Khrulkov et al. (2020) introduce a hyperbolic method of image embedding, which is aimed at better reflecting the hierarchical relationships in visual data. The study

has its primary purpose of filling the gaps that exist in the traditional Euclidean embedding spaces that fail to capture tree-like structures that are often present in hierarchical image classification. To address this issue, the authors present image embeddings on hyperbolic geometry, relying on the Poincaré ball model, and combine them with the deep convolutional neural networks. The proposed approach is tested on hierarchical image classification and retrieval, on subsets of ImageNet in which the labels of classes are organized in a natural hierarchy. The experiment results show that hyperbolic embeddings are always better than the Euclidean embeddings at hierarchy-aware classification and retrieval tasks where they achieve better top-k recall and better retrieval performance on several datasets. The study however indicates difficulties in training stability and the escalated complexity of the optimization in the hyperbolic space. To produce good results, the method also needs precise fine tuning of the curvature parameters. Moreover, the approach has primarily been assessed using general computer vision data sets and has not been assessed on medical imaging applications. The authors propose the expansion of hyperbolic learning methods into the fields of real life that require structured hierarchies of labels and safety-related decision-making as a research gap. This work justifies the application of hyperbolic learning models to structured medical image classification problems like subtyping of lung carcinoma by showing enhanced performance on representation of hierarchical label structures.[2]

The study by Mettes et al. (2024) provides an overall survey of hyperbolic deep learning techniques applicable to computer vision with much emphasis on learning representations, which naturally reflect the hierarchical relationship. The primary goal of the paper is to structure and describe the recent studies that indicate that hyperbolic geometry is more appropriate than the Euclidean space when working on problems that involve tree-like labeling. The research does not offer a solution to a single classification problem, but it deals with a greater challenge of making hierarchical data more effectively represented in visual learning tasks. In order to accomplish this, the authors study a broad set of current procedures and classify them as supervised and unsupervised: prototype-based and sample-based classification as supervised methods, and clustering, generative models, and learning in hyperbolic space as unsupervised methods. Some of the typical difficulties that these methods encounter, which are discussed in the survey, include training instability, numerical problems along the boundary of hyperbolic space, and higher computational cost. One of the shortcomings that are pointed out is the fact that most existing models employ hybrid architectures, with the embedding layer hyperbolic and the remaining network Euclidean. Moreover, most methods are not very efficient at scaling to large data sets. The authors highlight the research gap with fully hyperbolic neural networks and improved design options of core elements like convolutions and normalization. In general, the paper is a well-founded motivation to use the hyperbolic learning in structured and hierarchy-based tasks, such as in medical image classification issues like lung carcinoma subtyping.[20]

The scientific aim of the survey was the extensive review and syntheses of the literature in hyperbolic deep neural networks to demonstrate how the extension of deep learning models to hyperbolic geometry, as opposed to the traditional Euclidean space, can be done to provide more realistic representations of the hierarchical and relational data structure. It seeks to address the drawbacks of regular deep net-

works in learning structures such as trees, graphs and semantic hierarchies using the geometrical characteristics of hyperbolic space. The authors adopt a survey approach where they examine methodically the available architectures of hyperbolic archetypes, mechanical embedding, methods of optimization and deployment in tasks of machine learning and pattern recognition. The mathematical and computational complexity of non-Euclidean space operation, numerical instability during training and the absence of standardized frameworks to implement hyperbolic layers and models are identified as among the major challenges. There are also issues related to the fact that Hyperbolic learning has not been integrated into the mainstream deep learning libraries and its ability to be effectively utilized in a wide variety of datasets. Weaknesses mentioned are; the infancy of the entire hyperbolic network design, the fact that the hyperbolic design has more to be empirically confirmed in real world tasks and the fact that specialized knowledge is required to interpret and apply the hyperbolic techniques successfully.[11]

The research by Janßen et al. (2022) is aimed at enhancing lung cancer subtyping by integrating the images provided by digital pathology with the data of molecular analysis revealed with the help of MALDI mass spectrometry imaging. The primary goal of the research is to correctly differentiate the two most prevalent non-small cell lung cancer subtypes adenocarcinoma and squamous cell carcinoma, and decrease the amount of work done by pathologists during regular diagnoses. Their issue is that manual subtyping by immunohistochemical staining is time-consuming and can be constrained by availability of tissue. The authors recommend a multimodal deep learning pipeline to address this issue, which initially involves segmenting the whole slide image using a U-Net-based segmentation model to automatically identify the locations of the tumor and subsequently classifies the tumor subtypes with MALDI MSI data applying a neural network known as IsotopeNet. The model was trained using tissue microarrays of 232 patients and tested using independent whole slide sections. The proposed system had a test accuracy of 94.7% at the spectral level and identified 15 of 16 whole tissue sections with an accuracy of 100%, depending on whether a quality control threshold was met. Though the study has been very successful, it has cited problems like misclassification in areas of inflammation, necrosis, or abnormal tissue structures. One of the main constraints is that the method is based on the specific equipment of MALDI MSI, which is not always accessible in all clinical facilities. Also, the model is not structured to support hierarchical subtype relationships and is only defined to classify subtypes flatly. This demonstrates the necessity of future studies on learning frameworks that can extend to support hierarchical structure and more robust decision-making, specifically to enhance reliability and minimize clinically critical error in lung carcinoma subtyping.[10]

The aim of the research was to create and prove a non-invasive machine learning model to identify the pathological-subtypes of non-small cell lung cancer (NSCLC) of lung adenocarcinoma versus squamous cell carcinoma based on clinical, laboratory, and PET/CT radiomic features in aiding clinical decision-making. Their goal is to address the issue of a more appropriate assessment of the subtypes of NSCLC without using the invasive biopsies, which, however, may be unsafe or not very convenient in certain cases. The techniques included the following: 99 combined clinical, laboratory and radiomic features were picked, the Boruta algorithm is used to filter the features, and after that, several classifiers, such as Support Vector

Machine (SVM) and Random Forest (RF) are trained to determine the most suitable model. Among the critical issues were the handling of high-dimensional radiomic data, the selection of the best feature subsets, among others and the suitability of the few selected models in case of their retrospective single-center data and small sample size. The limitations mentioned are, a relatively small cohort of the patients, retrospective design, the possible absence of an external validation in a wide range of population, and the reliance on quality and consistency of the PET/CT image and clinical data which could limit the extrapolation to wider populations unless new multicenter studies are conducted.[12]

Fan et al. (2025) present a cross-modal hierarchical fusion framework on clinical prior to enhance histological subtyping of lung cancer by analyzing CT scans. The primary aim of the research is to increase the classification accuracy through the minimization of redundant information in CT images and the successful incorporation of clinical prior knowledge and imaging characteristics. The addressed problem is that the current single-modal and multimodal methods are not able to differentiate between subtle histological differences between subtypes of lung cancer and they cannot make use of clinical data in an appropriate way. To address this, the authors propose the CM-DAC framework that incorporates a YOLO-based slice-detecting module to identify the location of lesions, a 3D ResNet backbone to extract features, and a cross-modal hierarchical fusion component, which combines the CT features and clinical records by using the attention mechanism. Moreover, multimodal contrastive learning is employed to match image representations to ontology-driven textual label representations. Evaluation of the method is conducted on a publicly available dataset (LPCD) and a confidential clinical dataset. The suggested solution gets an accuracy of 80.88% and an AUC of 84.91% on the LPCD data, and an accuracy of 80.06% and 90.26% AUC on the private data, beating a variety of state-of-the-art solutions. The study however mentions challenges of increased complexity in the model and of dependence on accurate detection of the lesion at its early stages. The reliance on multiple types of data and manually maintained clinical priors can be considered a major constraint, as it can be limited in terms of scalability between institutions. The identified research gap is a requirement of more generalized and efficient hierarchical learning frameworks that can ensure high-performance and less dependence on multimodal inputs, as well as less clinically significant misclassification.[22]

The study hypothesis was to construct a radiomics-based machine learning pipeline to classify the histological subtype of non-invasive CT imaging features (adenocarcinoma versus other types) of pulmonary cancer through CT images. They are addressing the issue of effectively differentiating between lung cancer subtypes and classic CT scans without the need to make use of invasive pathology reports only that are prone to time-consuming tasks, not to mention subjectivity. They included measurements of radiomic features extracted via annotated CT regions of interest using PyRadiomics, a reduced signature of significant features, and the training of classification models measured over a reduced signature on both internal and external datasets. Primary challenges and concerns are the methods of handling the high-dimensional radiomic data, feature overlap, uneven classes, and the ability of the models to be usable in dissimilar cohorts and imaging vendors. The study has limitations which are moderate external validation performance (AUC equals ap-

proximately 0.70), reliance on the use of manual lesion annotation by radiologists, possible bias of the study due to composition of the data sets and the fact that it needs to be validated in larger multi-centers to ascertain the robustness of study and its application in various populations.[8]

Imran et al. (2024) suggest a hierarchical deep learning model of transformers to identify and classify non-small cell lung cancer (NSCLC) in histopathological images. The main objective of the study is to improve diagnostic accuracy by capturing both fine local features and long-range relationships within tissue images, which are often missed by traditional convolutional models. The addressed issue is that the manual analysis of whole slide images is time consuming and subject to human error, and the current deep learning techniques cannot effectively model the global context. To address it, the authors implement a hierarchical architecture, which fuses convolutional neural networks to extract local features with vision transformers to capture the long-range image patch dependencies. The model is categorized into three types of images namely normal tissue, adenocarcinoma, and squamous cell carcinoma. The method is tested with LC25000 benchmark dataset and it has a classification accuracy of 98.8%, F1-score of 98.0%, specificity of 99.1%, recall of 98.2% and precision of 98.0%. The challenges mentioned in the study include the increased complexity of architecture, and the longer training time than simpler CNN models, despite its high performance. One of the limitations is that the model is trained and evaluated with a single and well-curated dataset, which might prevent its application to real-world clinical data. Moreover, the method also emphasizes flat subtype classification and it does not explicitly mention the hierarchical error propagation or false-negative risk between subtypes. This creates a research gap of creating stronger hierarchical learning systems that can generalize across datasets and minimize clinically critical misclassifications.[17]

The research question aimed at developing a high-tech deep learning system to enable lung cancer pathological subtypes with high accuracy and using multi-modal medical images, which would meet the clinical requirement of non-invasive precision in diagnosing lung cancer. This is to overcome the drawbacks of the other existing models which use only a single modality input and lack both pathological data and representation space. The suggested approach, CC-DCNet, takes into consideration a dynamic convolutional neural network that is able to adapt to either paired CT and pathological images or CT itself, and a contrastive constraint module that is used to learn cross-modality associations and enhance diagnostic performance. Among them, there was managing the varied and inconsistent clinical image input, trade-offs between flexibility and accuracy across modalities, and learning significant minimal representations based on high-dimensional radiological and pathological data. There were also issues due to the absence of clinical data like missing pathological images and the need to have excellent generalization between large-scale multi-centers. The study has limitations that have been identified as: reliance on the accessibility of multi-modality imaging and quality of the imaging, the possibility of computational complexity of dynamic networks and the necessity to further validate in more generalized fields of patients and prove applicability and generalizability of their populations.[19]

The article by Javed et al. (2024) is a systematic review of the deep learning solutions

used to detect and classify lung cancer in medical images. The principal goal of this research is to provide an overview of the latest developments in the field of deep learning methods and investigate their application in enhancing automated lung cancer diagnosis. The authors conduct a systematic literature review of 66 journal articles published in 2015-2024 and use extensive academic databases and apply a rigid inclusion and quality evaluation criterion. Their review emphasizes that convolutional neural networks are the most widely used models, such as AlexNet, VGGNet, ResNet, DenseNet, Inception, and U-Net, where computed tomography has become the most common imaging modality since it provides a detailed view of the lung structure. Nevertheless, the authors reveal that there are a number of technical limitations, such as the dependence on predetermined input image sizes, which can lead to isometric resizing or cropping and can cause the removal of valuable diagnostic data or the creation of noise. Moreover, most studies rely on small or homogeneous data, which casts doubt on the extrapolation of the model to other populations and clinical settings. Another weakness mentioned in the review is the absence of standard evaluation practices and international standards, which makes it hard to make fair comparisons of reported performances. Also, the problem of data security and privacy as well as long-term model adaptability are not fully resolved. In general, the research demonstrates an obvious research gap in the creation of more robust, generalizable, and clinically reliable lung cancer detection systems, and in particular, those that will be more responsive to uncertainty and more consistent with real-world clinical practice.[18]

Yang et al. (2020) propose a study of hierarchical classification of pulmonary lesions using a large-scale radio-pathomics method which involves both radiological and pathological data. The primary goal of the research is to enhance the diagnosis of lung lesion through learning the disease types in a coarse-to-fine way instead of flat classification. The addressed problem is that the traditional approaches are not capable of capturing the hierarchical similarities between benign lesions, malignant tumors, and detailed pathological subtypes. To address this, the authors develop a hierarchical deep learning architecture in multiple stages, which combines CT-based radiomic features with pathology-based features, and the predictions are performed at various levels of the disease hierarchy. The framework employs convolutional neural networks to extract features and divides predictions of general lesion types into a specific subtype by using hierarchical classifiers. The model is tested on a large radio-pathomics dataset and this model has an overall classification accuracy of approximately 86% with an AUC of about 0.87 in key classification stages. Nonetheless, the research also indicates problems concerning the alignment of features between radiology and pathology data and discrepancies across data sources. A key shortcoming is that the architecture is complex and requires large, better-annotated datasets, which presents a research gap of less complex and more generalizational hierarchical models.[4]

Alsallal et al. (2025) provide a model of subtype classification of lung cancer based on CT scans using a combination of deep learning and radiomic features. Their primary aim is to optimize subtype prediction without changing the features, which are likely to be effective only in certain scan conditions. In order to address this, they apply an attention-based DeepCNN to ensure that the network concentrates more on the significant tumor regions and afterwards, the deep features are combined

with radiomics to classify them better. They evaluate the method on 2,725 CT slices of 160 patients, and feature-selection methods (including NMF and RFE) with machine-learning models, in which XGBoost performs well. Its highest results are 92.47% test accuracy, 93.99% AUC and 92.11% sensitivity. One of their major issues is that most profound features are inimitable which may lead to diminished confidence when the scanning conditions vary. They also report minor overfitting and state that bigger and more heterogeneous datasets and broader validation are yet to be done before actual clinical application.[21]

Yazdi et al. (2015) investigated why patients with non-small cell lung cancer might get false-negative results when staging mediastinal lymph nodes using endosonography methods like EBUS and EUS. False-negative outcomes can cause patients to be incorrectly diagnosed and receive wrong treatment. That's why, the study retrospectively reviewed 775 patients who were classified as lymph node-negative based on EBUS, EUS or combined procedures. The factors associated with missed lymph node metastases were found via statistical analysis using logistic regression. The researchers discovered that central tumor location, enlarged lymph nodes on CT scans and increased FDG uptake on PET scans were important predictors of false-negative findings. Based on these factors, they developed a clinical risk model to calculate the probability of false-negative diagnosis, which helps physicians decide whether further surgical staging is needed. However, this methodology is based only on clinical and imaging features instead of ML and AI and for this reason it lacks AI-driven false-negative minimization mechanisms. Also, uncertainty is not explicitly addressed and it can't make very detailed (subtype-level) predictions.[1]

Bruschi (2021) aimed to develop an understandable deep learning system that uses CT scans to categorize lung cancer types. The main challenge mentioned is that traditional CNN usually works as "black boxes", providing little explanation for their predictions. To overcome this, a 3D CNN was designed that was enhanced with SENet and Grad-CAM++ to offer visual interpretability. CT images were carefully preprocessed through normalization, resampling and lung segmentation to improve model performance. The method was trained to differentiate between adenocarcinoma and squamous cell cancer. It generates heatmaps that highlight significant areas in the scans along with classification results. Due to these visual explanations, clinicians are more able to understand and trust the AI's decisions. Yet, the model fails to include hierarchical disease structure and performs only flat classification between classes. Also, it does not specifically reduce false-negative errors or evaluate prediction uncertainty. Furthermore, advanced techniques like hyperbolic representations and evidential learning are not included. As a result, the system's applicability in high-risk clinical decision-making is limited.[9]

The study objective was to build a deep learning and radiomics-based model to predict the histological subtypes and the survival of lung adenocarcinoma with the use of preoperative CT, with the aim of aiding clinical decision making. Their goal is to address the problem of true classification of the subtypes of lung adenocarcinoma non-invasiveness, and determination of the patient prognosis, which in the past was only possible to illustrate invasive pathology, and subjectively. The approaches are a deep learning network ResNet-34 with a slight modification, radiomics feature extraction, as well as a hybrid deep radiomics classifier, tested on a huge multi-center

sample of 1,222 patients. The main issues here have to do with handling a high dimensional image set, strong subtype stratifications of more than two categories (2-, 3-, and 8-class prediction) and creating dependable prognostic models utilizing excessively survival information. Other problems were performance matching between internal and external validation cohort and the fusion of deep learning outputs with clinical relevance. The study has limitations listed; it is retrospective, there might be bias due to the heterogeneity of the dataset in the different centers and the study would need a further prospective validation to prove the generalizability and clinical use in wider populations.[7]

Bartlett et al. (2021) looked at the problem of false-negative outcomes in low-dose CT scans for lung cancer screening, where the main concern presented is that false negatives are defined and reported inconsistently across different screening programs. After thoroughly analyzing major international screening trials, such as NLST, NELSON and I-ELCAP, the researchers identified a number of causes of false negatives, such as screening process failures, errors in interdisciplinary clinical decision-making and radiological interpretation. The review reported that up to 15% of lung cancers detected in screening settings may be false negatives. To address this issue, the authors suggested using standardized definitions and tracking frameworks for false-negative cases as this would improve transparency, comparison across studies and overall screening program quality assurance. Still, neither AI based solutions nor any computational or predictive models are introduced in the paper. Because of this, false-negative reduction remains reactive and subjective instead of being proactive and data-driven. Although uncertainty is discussed conceptually, it is not formally assessed. Also, the work does not use hierarchical medical structures and algorithmic strategies to minimize false negatives. As a result, the study's main objective continues to be descriptive and theoretical in nature.[6]

The research objective was to create a non-invasive machine learning model whose features are based on radiomics via CT images to predict pathological subtypes of non-small cell lung cancer (NSCLC) and this aims at increasing the accuracy of preoperative diagnosis. They are addressing this clinical problem of decreasing reliance on invasive pathological sampling with the help of easily accessible CT imaging to subtype classification. The techniques involved identifying 1, 834 radiomic features on the CT scans of 317 histologically confirmed cases of NSCLC, feature selection by LASSO regression, and assessment of 11 machine learning algorithms (Random Forest, XGBoost, Extra Trees, Gradient Boosting, and SVM) to obtain predictive models. Among the problems and issues that they encountered included the handling of high-dimensional radiomic data, the ability to provide reliable cross-validation across cohorts and independent testing, and the projection of performance with multiple classes of NSCLCs. The study has several limitations, such as moderate predictive accuracy on independent datasets (e.g. AUC of around 0.618-0.644 of some subtypes), possible vulnerability to variations in CT image quality, and larger multi-centre validation would be necessary to support clinical usefulness and generalizability.[23]

Zhang et al. (2020) explores if hyperbolic attention models can better handle multi-scale histopathology images for medical image classification. The primary concern raised is that traditional deep learning models rely on Euclidean space, which is

not well suited for representing hierarchical or multi-level patterns found in medical images. That's why, the researchers propose a hyperbolic-attention model by mapping CNN features into hyperbolic space using the Poincaré ball model, and redesign standard attention and classification layers to work in this geometry. After using ResNet18 to extract features from different image scales, hyperbolic linear layers and hyperbolic attention are applied to combine the information. The suggested approach performs better than baseline and Deep MIL models, especially in multi-scale scenarios, according to experiments done on the Camelyon16 (cancer metastasis detection) and TCGA lung cancer classification datasets. This method not only implies that hyperbolic space can better portray hierarchical relationships in medical images but also helps preserve both global context and fine details. However, the study is limited to comparatively small datasets and simple tasks, and the attention mechanism is poorly interpretable.[5]

The main research hypothesis of the paper was to create a deep learning-based solution of histomorphological subtyping and risk stratification of small cell lung cancer (SCLC) using routinely acquired H&E-stained whole slide images, which would be used to enhance prognostication and clinical decision-making. The authors aimed to overcome the issue of traditional molecular subtyping being resource-demanding and challenging to apply to clinical practice through a hybrid unsupervised learning framework of deep representation and clustering one that identifies image-based histopathology phenotypes and clusters patients with clinically meaningful subtypes. Their approaches involved deep learning extraction of feature on WSIs, consensus clustering to identify subtypes, survival analysis, and combined multi-omic analysis to examine biological surrogacy. The main issues encountered were in managing the complex and high-dimensional image data, guaranteeing the reproducibility of the main subtypes under independent cohort, and multimodal data usage in biological explanation. The study has the following limitations: its retrospective nature with the possible selection bias, the requirement of large and heterogeneous datasets to confirm the application of such computational model to population and center-wide situations, and the difficulty of implementing such computational models into ordinary pathology processes without massive technical infrastructure and expertise.[24]

Khadirnaikar et al. (2023) aimed to provide a learning framework that improves lung carcinoma subtyping by enforcing hierarchical consistency and reducing false-negative predictions. The main concern addressed is that conventional deep learning models often overlook the natural hierarchy in medical diagnosis (e.g., nodule $\rightarrow$ malignancy $\rightarrow$ subtype), resulting in predictions that are medically unsafe and logically contradictory. In order to represent hierarchical medical structures more accurately, the authors suggest a hyperbolic evidential learning method that integrates features in hyperbolic space. The approach combines evidential deep learning with hyperbolic embeddings, modeling predictions as probability distributions instead of fixed labels. To penalize diagnostic order violations and enable consistent predictions across levels, a hierarchical loss function is used. The model also contains uncertainty estimation to lower overconfident false-negative errors. Moreover, experiments are conducted on lung cancer histopathology datasets for subtype classification tasks. In comparison to Euclidean baselines, the results reveal higher classification accuracy and more control over false negatives. The solution shows that hyperbolic geometry is effective for modeling diagnostic hierarchies in medical

imaging. It also benefits clinical decision-making by providing uncertainty-aware outputs. Despite this, the method requires complex hyperbolic operations that increase computational cost. Also, the study's low dataset size and lack of actual clinical validation make it difficult to generalize to real-world situations.[15]

2.3 Summary and Research Gap

The literature review shows a deeply ingrained separation between the hierarchical characteristics of medical diagnosis and the flat architectural assumptions of the majority of the available deep learning models applied to lung cancer analysis[13], [18]. Even though some studies find high classification rates with convolutional and transformer-based networks on CT images the methods generally consider subtypes of lung cancer as independent and equally spaced points in Euclidean embedding spaces[17], [21]. These notions cannot express the natural parent child connections existing in oncological taxonomies, such as the difference between malignant tumors and histological subtypes, resulting in logically inconsistent predictions[3], [4].

Furthermore, it has been demonstrated in earlier research that traditional softmax-based classifiers are more likely to produce overconfident predictions, especially when dealing with ambiguous or imbalanced samples[16]. This is particularly severe in safety-critical clinical contexts since a clinical system may lose out on timely clinical interventions and cancers that are not detected due to a firmly inaccurate forecast. While some research has attempted to address the issue of uncertainty estimation or hierarchical learning independently[2], [3], these approaches typically rely on ensemble techniques, manually created radiomic features, or multimodal inputs, which are more computationally demanding and unfeasible in healthcare applications with limited resources[10], [22].

As a result, there is an apparent research gap concerning a unified learning framework that will both consider geometric inconsistency and diagnostic safety. As far as we understand, there exists no methodology that combines hyperbolic geometry to parameterize hierarchical disease structure with Evidential Deep Learning to do the task of estimating uncertainty calibration with standard CT images only. This thesis aims to fill this gap by suggesting a Hyperbolic Evidential Learning framework which achieves this goal by imposing hierarchical consistency and reducing false-negative diagnoses using explicit uncertainty modeling and hence allowing a clinically reliable and logically consistent decision support system.

Chapter 3

Requirements, Impacts and Constraints

3.1 Final Specifications and Requirements

The complete design of Hyperbolic Evidential Network (HEN) will focus on the consistent combination of non-Euclidean geometry and uncertainty estimation to achieve clinical confidence in the subtyping of lung carcinoma. The architecture demands an instance of a high-performance feature extractor (e.g. a Densenet or Vision Transformer (ViT) backbone) to produce latent representations which are subsequently projected to a Poincare Ball manifold with the help of exponential maps, achieved by the Geoopt library. One of the fundamental design conditions is the binarial coding of the Kaggle Lung CT Dataset into an implicit hierarchical code that reflects the biological taxonomy of lung diseases with fine-grained diagnostic codes linked to parent ones. The network needs to substitute the usual Softmax results with an Evidential Deep Learning (EDL) head that values a Dirichlet distribution so that when presented with ambiguous or unsatisfactory spans, the model, at least, recognizes that it does not comprehend. The practical model uses an Evidential Deep Learning (EDL) objective with Bayes risk loss and KL-divergence regularization to forecast predictive uncertainty and is instead pursued by use of weighted sampling strategies, hyperbolic geometric embeddings, and safety-oriented evaluation metrics instead of using loss penalties. Hardware requirements Hardware resources (GPU acceleration with CUDA driver) are needed to address the challenges of Riemannian optimization and the software stack uses PyTorch to be modular. Finally, the system is defined so as to attain high recall and minimal hierarchical violation rates, this will present a strong, human-in-the-loop diagnostic tool that would match the logic of AI with the natural taxonomy of medical pathology.

3.2 Societal Impact

The societal consequences of detecting lung carcinoma using Hyperbolic Evidential Networks (HEN) lies in the fact that medical AI will change its current appearance as opaque/black box design to one that is clinically safe and logically consistent,

making it a reliable diagnostic instrument. This study positively impacts patient safety specifically by reducing false negatives to catastrophic concentrating on the hierarchical complexity of lung cancer, which is a primary cause of mortality in many parts of the globe, such that high-risk cases can be actively dealt with to offer them the early intervention required to increase their chances of survival. In addition to safety, HEN tackles the severe issue of trust gap in healthcare by compelling to explicitly quantify diagnostic uncertainty with Evidential Deep Learning; as a result, the system can indicate when it operates on questionable data, which will enable clinicians to use their knowledge where it is most relevant and minimise the possibility of incorrect treatment due to algorithmic overconfidence. It is also supported by the fact that hyperbolic geometry (including Poincare embeddings) is used to bring it into harmony with the internal logic of the AI and the natural hierarchical taxonomy of medical knowledge. In selecting tumor subtypes in the right context in the greater malignant cones, the architecture helps the model to avoid being exposed to any biological shortcuts, leading to improved and fair healthcare results. Finally, through direct medical logic integration into the neural network geometry, this model promotes a trustful human-in-the-loop model enhancing the diagnostic accuracy of medical systems, anxiety in patients and offering the possibility to lower the cost of health care through simplifying the clinical processes.

3.3 Environmental Impact

The main attributes of the environmental impact of the Hyperbolic Evidential Network (HEN) are that the computational efficiency is greatly enhanced in conjunction with the non-Euclidean geometry. In a traditional Euclidean deep learning model, embedding into a high dimensional space, typically into the hundreds of dimensions, is often needed to properly model highly structured biological hierarchies, and the cost, in terms of number of parameters and power needed to train and entropy consumed during each inference is astronomical. As a comparison, the application of the Poincare Ball manifold demonstrates that this project can fit the tree-like taxonomy of lung carcinoma into the two-dimensional space with significantly less distortion and, as a result, significantly reduced the computational footprint of the model and the corresponding carbon emissions of the high-performance computing. Moreover, a switch to more efficient backbones, such as Vision Transformers (ViT) and Riemannian optimization methods (e.g. Riemannian Adam) guarantees that the network will converge better, using less overall kilowatt-hour of CUDA-enabled hardware. In addition to the direct digital infrastructure, the fact that the system can be used to quantify uncertainty through Evidential Deep Learning minimizes diagnostic ambiguity. This indirectly reduces the environmental footprint of medical waste and the energy-intensive nature of to-nothing follow-up medical procedures, which are the consequence of the second Terabype and which often feature in black-box overconfidence. This study shows a dedication to the idea of Green AI by matching the AI reasoning with the natural efficiency of biological systems; hence, showing that clinical safety and strong accuracy do not require an ever-growing environmental cost.

3.4 Ethical Issues

The Hyperbolic Evidential Network (HEN) is an ethical environment in the depletion of lung cancer through the application of an opaque black-box accuracy to the domain of systemic responsibility, transparency, and clinical safety. One of the main ethical issues that this study will deal with is the elimination of overconfidence in algorithmic predictions which in the past induced violation of logical hierarchy and misclassification of life-threatening malignancies. The system explicitly measures uncertainty by incorporating Evidential Deep Learning, to put the ethical burden of blindly using the automation system on a human-in-the-loop system whereby the AI can also indicate that it does not know when it encounters ambiguous or out-of-distribution samples. This clarity is critical in the creation of a sense of Doctor Trust as well as the prevention of being misled by the overconfident but structurally misaligned output which is a major danger in typical Euclidean architectures where disease categories are considered flat.
In addition, the project is able to address problems in health equity and algorithmic bias due to the application of class-balanced evidential procedures that enhance the performance on rare cancer subtypes, meaning that groups of patients with less common pathologies are not underserved by the imbalances inherent in medical data. Biological consistency by hyperbolic geometry makes the inner logic of the AI follow scenarios of established medical taxonomies- like that a Squamous Cell subtype always appears within the larger Malignant cone- and thus does not learn biased shortcuts which might lead to disastrous false negatives. Finally, the HEN architecture inherently has an ethically positive aspect as its 12-week safety integration and assessment program is stringent, and reduction of diagnostic errors and the explicit declaration of risk are the key leadership roles of the creation of trustworthy medical AI.

3.5 Project Management Plan

The Project Management plan of "The hyperbolic evidential learning approach to hierarchical consistency and false-negative minimization in lung carcinoma subtyping" assumes the position of a categorical technical roadmap that will be utilized in order to move a standard Euclidean black-box model towards a clinically dependable diagnostics system. The first stage aims at scope definition and information governance, through which the grouping of the Kaggle Lung CT Dataset into a hierarchical labeling hierarchy to help the biological relationships among benign and malignant classes of lung cancer. The subsequent task is the technical implementation phase, which focuses on the implementation of the Hyperbolic Evidential Network (HEN) architecture, where a state-of-the-art feature extractor such as a Densenet121 or a Vision Transformer is designed and add the disease taxonomy with minimal distortion into the Poincare Ball manifold. One of the most important milestones of the project is the "Safety Integration," in which Evidential Deep Learning (EDL) is used to substitute standard Softmax outputs with Dirichlet distribution to enable the model to directly measure aleatoric as well as epistemic uncertainty. Risk management is an ongoing process across the project, aimed at minimizing catastrophic

false negatives and hierarchy violations in place of explicit hierarchy-penalizing loss terms, using evidential uncertainty modeling, hyperbolic geometric inductive bias and class balanced sampling. Quality assurance will be performed with the help of intensive training on either Riemannian Adam or SGD optimization so that to ascertain numerical stability in the curved space and then an evaluation phase based on clinical performance indicators like recall and constraint violation rates. The last stage of the project is a thesis preparation step that summarizes the work to prove that slowing down an automated oncology system with the use of embedded medical logic leads to the establishment of a trust-producing "Doctor Trust" system, and results in a safer execution of the entire system process.

3.6 Risk Management

Hyperbolic Evidential Network (HEN) project risk management is a proactive model that is aimed at reducing clinical, technical, and structural risks related to automated lung cancer diagnosis. The harmful clinical risk is the most critical, whereby failure to detect a malignancy causes loss of life due to delayed treatment. To alleviate this, the project focuses on recall using class-balanced sampling, evidential uncertainty modeling and safety-oriented evaluation measures such that high-risk malignant cases are prioritized without using a recall-specific loss. Structurally, the threat of logic failures, that a model generates a subtype that is inconsistent with its parent category, is handled by employing Poincaré embeddings and hyperbolic entitlement cones. The labels with which the system is woven in a non-Euclidean space serves to maintain tumor hierarchies, which means that the model does not learn biological shortcuts and makes the hierarchies.
The technical risks are addressed by using Riemannian Adam or SGD-based optimization to avoid model divergence over the 12-week period in development, specifically its numerical instability that is common with training in curved space. Moreover, Evidential Deep Learning is the mitigation factor in the risk of black-box overconfidence in unclear samples, based on the Kaggle Lung CT Dataset used in this study. Through Dirichlet distributions to measure diagnostic uncertainty, the model can exhibit indications when a sample is out-of-distribution or logically inconsistent as an important instrument to guard clinicians against failure modes. This represents a complete plan of management, to compare HEN to Euclidean baselines (such as HMCN) to check the accuracy of the final architecture as well as the structural soundness of it, and the general ethical considerations of the human in the loop requirements of the healthcare system of today.

3.7 Economic Analysis

The financial case of the Hyperbolic Evidential Network (HEN) highlights its ability to raise the extent of cost savings in the healthcare system through the realization of a high level of diagnostic accuracy and efficiency. The high cost of treating late-stage lung cancer, which costs approximately two to four times less than intervention at an early stage, can be viewed as one of the major economical forces; the cost

to treat late-stage lung cancer will be $16,207 or higher against $4,326 to treat it in the early stages. HEN allows the reduction of catastrophic false negatives by evidential uncertainty modeling, weighted training, and safety-based evaluation. enabling earlier intervention and reduced treatment costs.

In addition, the Evidential Deep Learning (EDL) is also integrated to measure the diagnostic uncertainty, which directly responds to the high costs of unnecessary invasive procedures. Lung biopsies are one of the most costly diagnostic methods as the average cost of $14,634 for this test can exceed $37,745 in case of complications, e.g., pneumothorax. The system allows clinicians to avoid negative biopsies (previously 43.1 percent of all diagnostics workup costs) by explicitly indicating ambiguity in the cases of low confidence or out of distribution.

Technically, computation with hyperbolic geometry has computational benefits; the complex hierarchical taxonomy of lung diseases are computationally represented in a Poincaré manifold with exponentially fewer dimensions than alternative systems, which reduces the overall training time by about 17.8 per cent and the overall energy and cloud compute costs. Finally, the HEN architecture offers an inexpensive yet efficient aspect of human-in-the-loop solution to improve the productivity of radiologists and provide high value in terms of screenings in the health of a population based on negative incremental cost-effectiveness ratio.

Chapter 4

Proposed Methodology

4.1 Design Process or Methodology Overview

This study addresses a major problem in automated lung cancer diagnosis: achieving high diagnostic accuracy while minimizing safety-critical errors, especially false negatives. We propose an integrated framework combining hyperbolic geometry with **Evidential Deep Learning (EDL)** to facilitate hierarchical classification aligned with clinical taxonomies. The model not only offers diagnostic predictions but also explicitly quantifies predictive uncertainty, enabling more confident decision-making and appropriate referrals for unclear cases.

4.1.1 Core Innovation

Traditional neural networks represent medical images in flat, Euclidean feature spaces, which inadequately captures the natural hierarchical structure of disease taxonomies. For example, adenocarcinoma is a subtype of malignant cancer, yet most classifiers treat them as independent, unrelated categories without encoding their parent-child relationship. This research hypothesizes that embedding features in **hyperbolic space** (specifically, the Poincaré ball) naturally captures these hierarchical relationships, thereby reducing hierarchy violations and improving clinical safety.

4.1.2 Methodology Pipeline

The complete design process follows five sequential stages:

1. **Data Acquisition & Preprocessing:** CT scan images are standardized and annotated with dual-level labels: fine-grained diagnostic categories (e.g., Adenocarcinoma) and coarse-grained parent categories (Malignant vs. Benign).

2. **Feature Extraction:** Four state-of-the-art deep learning architectures (EfficientNet-B0, DenseNet121, ViT-B/16, PVT-V2-B0) serve as backbone networks to extract high-level visual representations from lung images.

3. **Hyperbolic Geometric Mapping:** Rather than conventional Euclidean projection, features are mapped to hyperbolic space using the Poincaré ball

model, which naturally accommodates tree-like hierarchical structures with minimal distortion.

4. **Evidential Classification:** The network outputs a Dirichlet distribution parameterized by evidence parameters α, rather than point-estimate probabilities, enabling explicit quantification of predictive uncertainty.

5. **Safety-Focused Evaluation:** Performance is assessed using **False Negative Rate (FNR)** and **Hierarchy Violation Rate**, prioritizing patient safety over raw accuracy.

4.2 Preliminary Design or Design (Model) Specification

To test the hypothesis that hyperbolic geometry provides superior organization of hierarchical medical data, this research implements two configurations: a standard **Euclidean baseline** and the proposed **Hyperbolic Evidential Network**.

4.2.1 Backbone Architectures

To ensure results generalize across different neural architectures, four diverse backbones are employed:

Table 4.1: Backbone Architecture Specifications

Architecture	**Key Characteristics**
EfficientNet-B0	A lightweight convolutional neural network that uses compound scaling to balance depth, width, and resolution efficiently. It provides strong performance with significantly fewer parameters and was pre-trained on ImageNet.
DenseNet121	Features dense connectivity patterns where each layer receives inputs from all preceding layers, enabling efficient feature reuse and smooth gradient flow.
ViT-B/16	Vision Transformer that divides images into 16×16 patches and processes them as sequences, capturing global spatial relationships through self-attention mechanisms.
PVT-V2-B0	Pyramid Vision Transformer with hierarchical feature extraction, combining transformer-based global attention with multi-scale spatial representations suitable for medical imaging tasks.

All backbones are initialized with ImageNet pre-trained weights and fine-tuned on the lung cancer dataset.

4.2.2 Geometric Framework Specification

Euclidean Baseline

The baseline model maps extracted features to a standard flat vector space (R^d). Classification is performed by:

1. Computing Euclidean distances between image embeddings and learned class prototypes
2. Converting distances to evidence scores via a neural network
3. Generating Dirichlet parameters α through a softplus activation

Hyperbolic Proposed Model

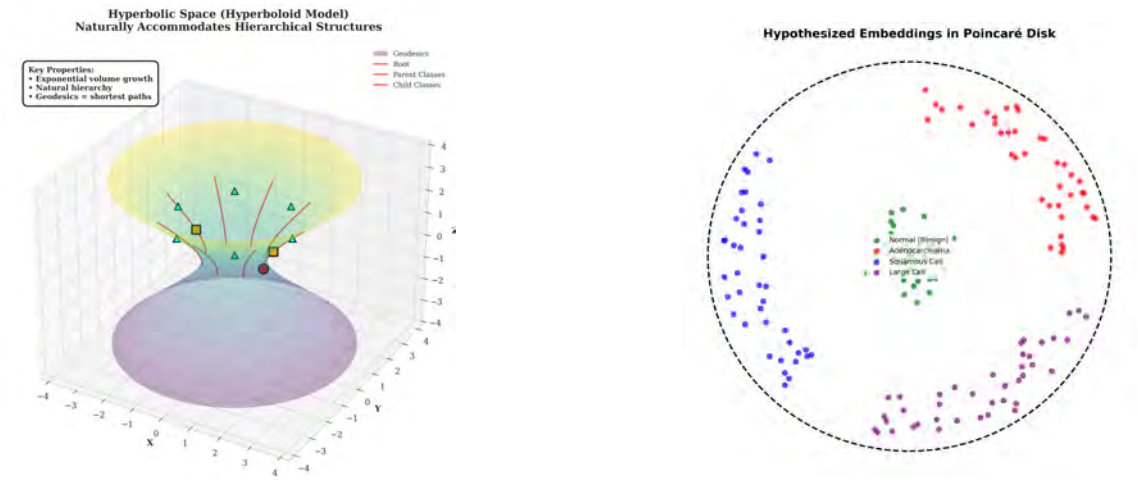

Figure 4.1: Poincaré disk model illustrating hyperbolic space used for hierarchical embedding

The proposed model replaces the flat space with the Poincaré ball

$$B^d = \{x \in R^d \mid \|x\| < 1\},$$

a model of hyperbolic geometry where:

- Features are mapped to the ball via the exponential map:

$$\exp_0(\mathbf{v}) = \tanh\left(\frac{\|\mathbf{v}\|}{2}\right)\frac{\mathbf{v}}{\|\mathbf{v}\|}$$

- Class prototypes are **Riemannian manifold parameters** optimized directly on the curved surface
- Classification uses hyperbolic distance (geodesic distance) instead of Euclidean distance

Why Hyperbolic Space? In hyperbolic geometry, the circumference of a circle grows exponentially with radius, allowing hierarchical tree structures to be embedded with minimal distortion. This property aligns naturally with medical taxonomies: the "Malignant" category encompasses all its subtypes within a proximal hyperbolic neighborhood, preserving parent-child relationships geometrically.

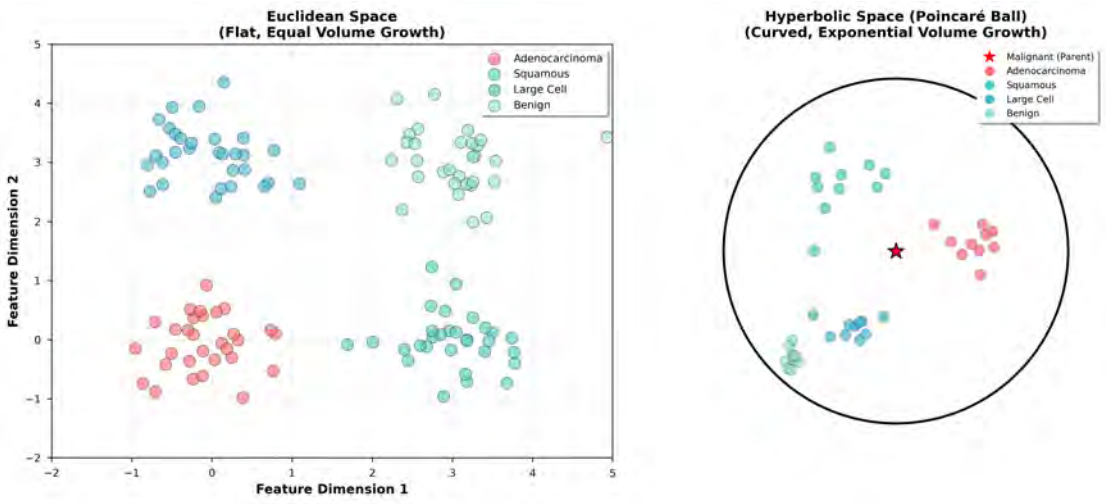

Figure 4.2: Visualization of hyperbolic geometry and comparison with Euclidean embedding behavior

4.3 Data Collection and Preprocessing

4.3.1 Data Collection

This research utilizes the **Kaggle Lung CT Dataset**[14], a publicly available collection of 2,077 chest CT scans. This dataset encompasses a variety of lung conditions across six categories:

- **Benign Categories:**
  - Normal: Healthy lung tissue (455 images)
  - Benign: Non-cancerous abnormalities (80 images)
- **Malignant Categories:**
  - Adenocarcinoma: Glandular tissue origin (195 images)
  - Large Cell Carcinoma: Undifferentiated large cells (115 images)
  - Squamous Cell Carcinoma: Epithelial tissue origin (155 images)
  - Malignant: General malignant classification (460 images)

This dataset was selected due to its clinical relevance, public availability, hierarchical label structure, and sufficient sample size for deep learning experiments.

4.3.2 Data Cleaning

We have used the Kaggle Lung CT [14] Dataset in the format it was provided, i.e., images are sorted in class-specific folders in the standard directory structure. When we loaded the data using the `ImageFolder` utility of PyTorch, in essence, it checked the integrity of the data. This helper class attempts to locate a consistent folder structure, and actually attempts to read each image. If there was any corrupted file or unsupported file then it would throw an error while iterating - but on our runs, nothing broke.

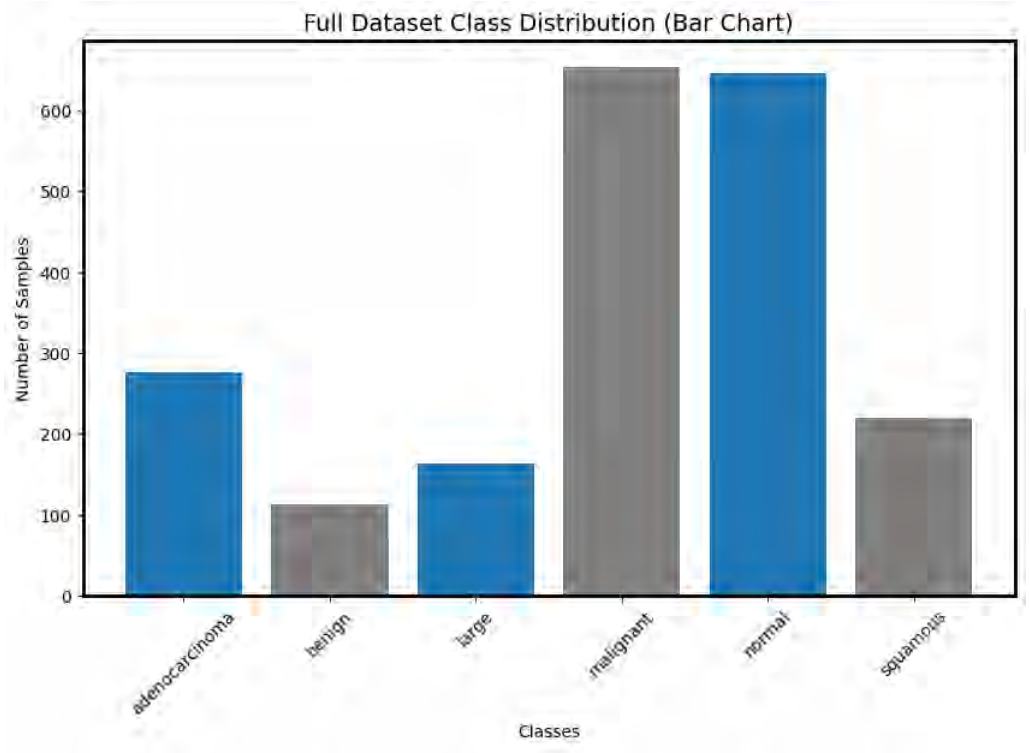

Figure 4.3: Overall class distribution of the Kaggle Lung CT Dataset

Label consistency was ensured by flexible pattern matching in the hierarchical classification scheme. Class names were normalized to lowercase when assigning parent categories so as to ensure robust handling of naming variations (e.g., "Squamous Cell Carcinoma", "squamous", and "SQUAMOUS" all are correctly assigned to the Malignant parent category through keyword matching).
In total, 2,077 images were successfully loaded across the training, validation, and testing subsets without encountering any loading errors.
All 2,077 samples successfully loaded without errors and were included in the final dataset used for model training and evaluation.

4.3.3 Data Transformation

Medical images require standardization to match the input requirements of pretrained neural networks:

1. **Resizing:** All scans were resized to 224×224 pixels, the standard input dimension for ImageNet-pretrained models. For training data, an additional random crop from 256×256 was applied as augmentation.

2. **Input Format Standardization:** The dataset images are stored in a 3-channel format to maintain compatibility with ImageNet-pretrained backbone architectures, which expect RGB input tensors of shape (3, H, W). The CT scan images were originally grayscaled image. So the intensity values were duplicated in the three RGB channels so that pretrained models can work easily without any structural change.

3. **Normalization:** Pixel intensities were normalized using ImageNet statistics: mean = $[0.485, 0.456, 0.406]$ and standard deviation = $[0.229, 0.224, 0.225]$ per channel.

4. **Data Augmentation (Training Only):** To improve generalization and prevent overfitting, training images underwent stochastic transformations:

   - Random horizontal flips (probability = 0.5)
   - Random rotations ($\pm$15 degrees)
   - Color jitter (brightness $\pm$20%, contrast $\pm$20%, saturation $\pm$10%)

Validation and test sets were only resized and normalized, without augmentation, to ensure consistent evaluation across all samples.

4.3.4 Data Integration and Reduction

The dataset exhibits **class imbalance**, a typical characteristic of medical datasets where certain diagnostic categories are underrepresented. To address this, the following data partitioning and sampling strategies were implemented:

Table 4.2: Dataset Split Strategy

Subset	**Percentage**	**Samples**	**Purpose**
Training	70.3%	1,460	Model parameter optimization
Validation	6.8%	142	Hyperparameter tuning, early stopping
Testing	22.9%	475	Final unbiased performance evaluation
Total	**100%**	**2,077**	**Complete dataset**

The per-class distribution in the complete dataset is: Adenocarcinoma (195 samples), Benign (80 samples), Large Cell (115 samples), Malignant (460 samples), Normal (455 samples), and Squamous (155 samples). This imbalance motivates the weighted sampling strategy described below.

Dataset Partitioning: The dataset was pre-partitioned into training, validation, and testing subsets following a stratified split strategy provided by the dataset organization. This ensures proportional representation of all classes across splits and preserves the original class distribution in each subset. Such stratification prevents minority classes (e.g., Benign with only 80 samples) from being underrepresented or absent in validation or test sets. The resulting split follows the distribution shown in Table 4.2.

Weighted Random Sampling: During training, a weighted sampler assigns higher selection probability to underrepresented classes. Class weights are computed as:

$$w_i = \frac{\max_j(n_j)}{n_i}$$

where n_i is the number of samples in class i, this ensures the model encounters minority classes frequently enough to learn their distinctive patterns effectively, mitigating the bias toward majority classes that would otherwise dominate gradient updates.

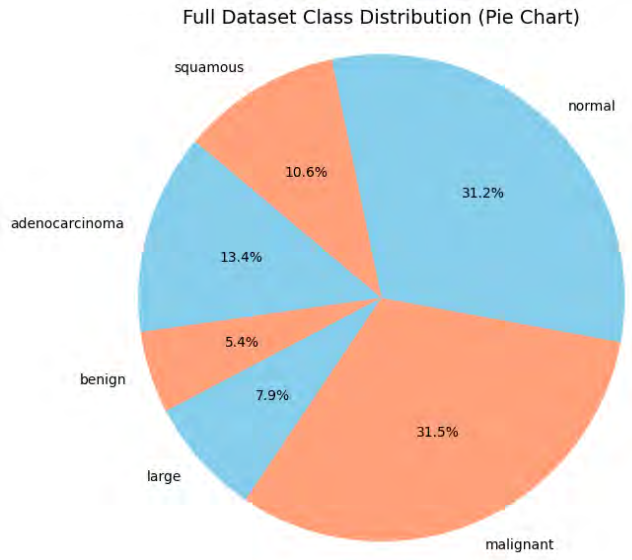

Figure 4.4: Class-wise distribution of lung cancer categories shown as pie and bar charts

4.3.5 Hierarchical Label Encoding

A key feature of this methodology is dual-level labeling that captures the medical taxonomy inherent in lung cancer classification:

- **Fine-Grained Label (y_{fine}):** The specific diagnosis at the subtype level (6 classes total: Normal, Benign, Adenocarcinoma, Large Cell Carcinoma, Squamous Cell Carcinoma, Malignant)
- **Coarse-Grained Label (y_{coarse}):** The parent category in the disease hierarchy (2 classes: Benign, Malignant)

This dual-encoding scheme enables the monitoring of hierarchical consistency during model evaluation. The hierarchical structure is illustrated in Figure 4.5, where each fine-grained subtype is explicitly associated with its corresponding parent category.
Implicit Hierarchical Consistency Through Hyperbolic Geometry:

While the primary loss function (Evidential Deep Learning with Bayes risk and KL divergence) optimizes for fine-grained classification accuracy, the use of hyperbolic geometric embedding provides an **implicit inductive bias** toward hierarchically consistent representations. In the Poincaré ball model, the exponential expansion of space naturally organizes data in a tree-like structure, where:

1. Semantically similar classes (e.g., Adenocarcinoma, Squamous Cell Carcinoma, Large Cell Carcinoma) are embedded in proximal regions corresponding to their shared Malignant parent category.
2. The learned class prototypes on the hyperbolic manifold inherently respect parent-child relationships, as geodesic distances preserve hierarchical proximities.

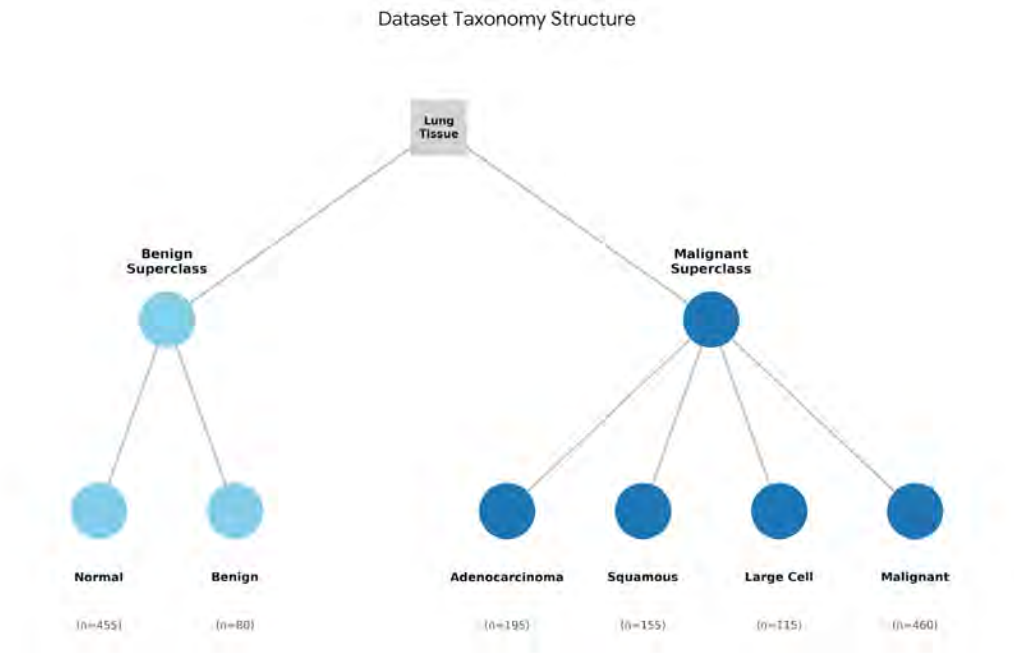

Figure 4.5: Hierarchical organization of lung cancer labels into coarse and fine classes

3. The model is discouraged from making biologically inconsistent predictions (e.g., predicting a specific malignant subtype while assigning high probability to the Benign parent) due to the geometric constraints imposed by the curved embedding space.

Evaluation of Hierarchical Consistency:
During the evaluation phase, we use the coarse-grained labels in order to calculate the **hierarchy violation rate**, which is basically the fraction of the predictions for which the fine-grained class disagrees with its parent category. For example, if the model says that it is predicting Adenocarcinoma (some malignant sub-type of Adenocarcinoma) and the coarse-level prediction, i.e. the probability mass, indicates Benign, that's a hierarchy violation. This type of metric captures how true to the clinical taxonomy the model is and is an indicator of how effective hyperbolic geometric embedding is in preserving structural consistency even without explicit supervision in form of loss.
This method is different from methods that explicitly punish the violations of hierarchy in the loss function. Instead, our framework takes advantage of the **representational geometry itself** as a structural prior that leads the model to medically coherent predictions, but at the same time preserves the flexibility to optimize for primarily discriminative performance on fine-grained classification.

4.3.6 Summary of Preprocessed Data

The final preprocessed dataset consists of:

- **Dimensions:** 2,077 images of shape (3, 224, 224) as PyTorch tensors, where the 3-channel format ensures compatibility with ImageNet-pretrained backbone networks

- **Format:** Float32 with ImageNet normalization
- **Labels:** Dual hierarchical encoding (fine-grained + coarse-grained)
- **Storage:** Loaded dynamically via PyTorch DataLoader with balanced sampling for training

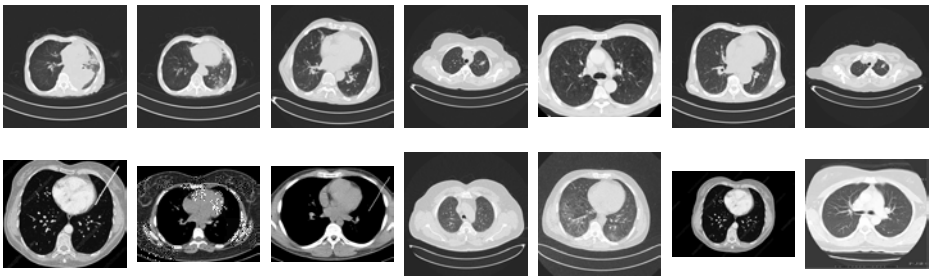

Figure 4.6: Sample lung CT images: Top row shows malignant cases, bottom row shows normal cases.

4.4 Implementation of Selected Design

The proposed Hyperbolic Evidential Network is implemented using the PyTorch deep learning framework with custom Riemannian optimization via the `geoopt` library.

4.4.1 Hyperbolic Evidential Head

The main architectural innovation is the classification head which operates on the Poincaré ball manifold:

Projection to Hyperbolic Space

Given a feature vector $\mathbf{f} \in R^d$ from the backbone network, it is projected into the Poincaré ball via the exponential map at the origin:

$$\mathbf{h} = \exp_0(\mathbf{f}) = \tanh\left(\frac{\|\mathbf{f}\|}{2}\right)\frac{\mathbf{f}}{\|\mathbf{f}\|}$$

This mapping ensures $\|\mathbf{h}\| < 1$, placing the embedding strictly within the unit ball.

Hyperbolic Distance Calculation

Classification is based on the hyperbolic distance between the image embedding $\mathbf{h}$ and the learned class prototypes $\{\mathbf{p}_k\}_{k=1}^{K}$ (where K is the number of classes). The hyperbolic distance in the Poincaré ball is:

$$d_H(\mathbf{h}, \mathbf{p}_k) = \operatorname{arcosh}\left(1 + 2\frac{\|\mathbf{h} - \mathbf{p}_k\|^2}{(1 - \|\mathbf{h}\|^2)(1 - \|\mathbf{p}_k\|^2)}\right)$$

Smaller distances indicate stronger class membership.

Evidence Generation

Hyperbolic distances are converted to evidence scores through a learnable neural network:

1. Scale distances: $\mathbf{d}_{\text{scaled}} = \mathbf{d}/(\tau + 0.1)$ where τ is a learned temperature parameter
2. Pass through evidence network: $\mathbf{e} = \text{EvidenceNet}(-\mathbf{d}_{\text{scaled}})$
3. Apply softplus: $\alpha_k = \text{softplus}(\mathbf{e}_k) + 1$

The resulting $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_K]$ parameterizes a Dirichlet distribution over class probabilities, where higher α_k indicates stronger evidence for class k.

4.4.2 Loss Function Implementation

The training objective balances accuracy and uncertainty calibration through an evidential loss formulation:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Bayes}} + \lambda_{\text{KL}} \mathcal{L}_{\text{KL}}$$

Bayes Risk Loss ($\mathcal{L}_{\text{Bayes}}$)

This term minimizes prediction error for the correct class. Let $\mathbf{y}$ be the one-hot true label and $\boldsymbol{\alpha}$ the predicted Dirichlet parameters:

$$\mathcal{L}_{\text{Bayes}} = \sum_{k=1}^{K} (y_k - \hat{p}_k)^2 + \sum_{k=1}^{K} \frac{\hat{p}_k(1 - \hat{p}_k)}{S + 1}$$

where $\hat{p}_k = \alpha_k / S$ is the expected probability and $S = \sum_{k=1}^{K} \alpha_k$ is the Dirichlet strength. The first term penalizes misclassification, while the second term accounts for prediction variance.

KL Divergence Regularization ($\mathcal{L}_{\text{KL}}$)

To prevent overconfidence on ambiguous samples, the KL divergence between the predicted Dirichlet and a uniform Dirichlet prior is added:

$$\mathcal{L}_{\text{KL}} = \text{KL}[\text{Dir}(\boldsymbol{\alpha}) \| \text{Dir}(\mathbf{1})]$$

This regularization encourages the model to express high uncertainty when evidence is weak. The weight λ_{KL} is annealed from 0.01 to 1.0 over the first 20 epochs to allow initial learning before enforcing strict uncertainty calibration.

Note on Hierarchical Consistency

It's worth noting that the loss doesn't really have a dedicated term to penalise violations of the coarse- fine hierarchy (for example predicting Adenocarcinoma while giving a high confidence Benign label). Instead, hierarchical consistency manifests itself as an **implicit property** of the hyperbolic embedding space. The Poincare

ball geometry tends to group together related subtypes near their parent categories, which reduces the likelihood of hierarchy violations without requiring any direct supervision. During evaluation we monitor the rate of hierarchy violations as a measure to ensure that indeed the geometric inductive bias maintains the medical taxonomy. This is in contrast to methods which explicitly incorporate hierarchical constraints in the loss; we use the representational geometry as a structural prior.

4.4.3 Optimization Strategy

Challenge: Standard optimizers such as Adam compute gradients under the assumption that parameters reside in flat Euclidean space. However, the class prototypes $\{\mathbf{p}_k\}$ are constrained to the Poincaré ball manifold, requiring Riemannian optimization.
Solution: The `RiemannianAdam` optimizer from `geoopt` is used, which:

- Computes gradients in the tangent space of the manifold
- Projects gradient updates back onto the manifold using retraction operations
- Ensures prototypes remain valid points inside the Poincaré ball throughout training

Training Schedule:

1. **Stage 1: Warmup (10 epochs)** – Backbone frozen, only classification head trained with learning rates: projection head (1e-3), prototypes (2e-2), evidence network (1e-3).
2. **Stage 2: Fine-tuning (up to 100 epochs)** – All parameters unfrozen with reduced learning rates: backbone (5e-5), projection head (1e-4), prototypes (3e-3).
3. **Early Stopping:** Training halts if validation accuracy plateaus for 10 consecutive epochs (standard deviation < 0.001).

Gradient clipping (max norm = 1.0) is applied to prevent exploding gradients during manifold optimization.

4.5 Computational Resources and Reproducibility

4.5.1 Hardware Configuration

All experiments were conducted on cloud-based GPU infrastructure:
Automatic Mixed Precision (AMP) training was enabled to reduce memory usage and accelerate computation using Tensor Cores.

4.5.2 Software Environment

The implementation relies on the following open-source software stack:

Table 4.3: Computational Environment

Component	**Specification**
GPU	NVIDIA Tesla T4 (16 GB VRAM)
Platform	Google Colab Pro+
CPU	Intel Xeon (2 vCPU)
RAM	12 GB
Training Time	∼2-4 hours per backbone

Table 4.4: Software Dependencies

Library	**Version/Purpose**
Python	3.10
PyTorch	2.0+ (deep learning framework)
torchvision	Pre-trained backbone models
geoopt	0.5.0 (Riemannian optimization)
NumPy	Numerical computations
Pandas	Data analysis and logging
scikit-learn	Evaluation metrics (ROC, PR curves)
Matplotlib/Seaborn	Visualization

4.5.3 Reproducibility Measures

To ensure independent verification of results:

1. **Random Seed Control:** Global seed (42) was set for all random number generators:
   - `torch.manual_seed(42)`
   - `np.random.seed(42)`
   - `random.seed(42)`
2. **Deterministic Algorithms:** CUDA benchmarking disabled via `torch.backends.cudnn.dete = True` to ensure identical results across runs.
3. **Version Pinning:** All library versions logged and fixed in `requirements.txt`.
4. **Code Availability:** Complete source code, including data preprocessing scripts and model implementations, is documented and available for verification.

4.6 Ethical Considerations

The application of artificial intelligence in medical diagnostics carries significant ethical responsibilities, which are addressed directly in this research.

4.6.1 Patient Safety and False Negative Minimization

The most important thing to look for is **false negative**. When a malignant cancer is shown as benign, then the patient can be mistreated, which would lead to a fatal situation. So identifying the fnr is crucial for clinical treatment. So in this study, the fnr is minimized through these steps

- **Explicit FNR Tracking:** The FNR is used as the primary evaluation metric for both training and testing.
- **Uncertainty Quantification:**The evidential framework gives the scores of uncertainty results for each of the predictions. So the high certainty results are flagged for further doctor review rather then a certain accuracy.
- **Conservative Thresholding:** The system follows the medical principle of "first, do no harm" by being cautious. It means a false positive (unneeded follow-up) is far better than a false negative (missing cancer).

Ethical Principle: The system adheres to the medical principle of "first, do no harm" by erring on the side of caution. A false positive (triggering unnecessary follow-up) is ethically preferable to a false negative (missing cancer).

4.6.2 Data Privacy and Compliance

Dataset Characteristics: The Kaggle Lung CT Dataset is:

- Publicly available for research and educational use
- Fully de-identified (no patient names, ages, medical record numbers, or other personally identifiable information)
- Compliant with medical data sharing regulations and HIPAA guidelines

Data Handling: No additional patient data was collected. All experiments used only the published Kaggle dataset. No re-identification attempts were made, and all data processing followed ethical research guidelines.

4.6.3 Algorithmic Transparency and Explainability

While deep neural networks are often criticized as "black boxes," this research enhances interpretability through:

- Visualization of hyperbolic embeddings showing class clustering patterns
- Uncertainty scores that indicate model confidence levels
- Hierarchical violation tracking that reveals structural prediction errors

Clinical Integration Recommendation: This system should be deployed as a decision support tool to alert radiologists to potential cases of concern, rather than as an autonomous diagnostic system.

4.6.4 Bias and Fairness Considerations

Potential Limitations: Although Kaggle Lung CT Dataset has useful clinical samples, there is risk of limitations:

- Geographic and demographic heterogeneity could be restricted according to initial data collection points.
- Differences in manufacture of CT scanners, imaging protocols and resolution among various hospitals.
- Possible variations in cancer rates as well as patterns of presentation among different population groups in the world
- Class imbalance of medical data (some of the subtypes represented poorly)

Mitigation Strategies:

- Weighted sampling addresses class imbalance during training
- Data augmentation improves robustness to imaging variations
- Multi-backbone validation ensures results generalize across architectures
- Future work should validate the model on multi-institutional datasets from diverse geographic regions before clinical deployment to assess generalization and detect potential biases

4.7 Summary

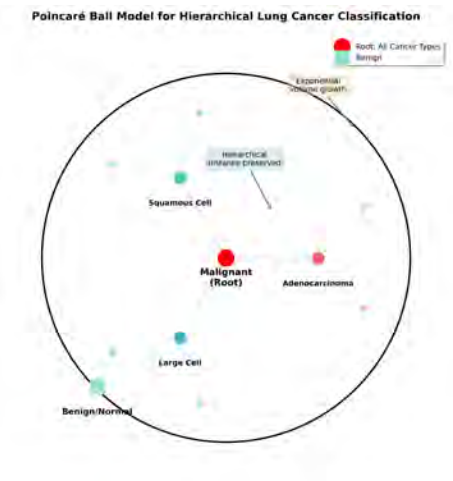

Figure 4.7: Visualization of Poincaré Ball for Hierarchical Lung Cancer Classification

This chapter presented a comprehensive methodology for constructing a clinically safe lung cancer classification framework that addresses limitations of conventional Euclidean neural networks. The key contributions include:

1. **Hyperbolic Geometric Embedding:** By leveraging the Poincaré ball model, the system naturally represents hierarchical relationships among cancer subtypes, reducing prediction errors that violate medical taxonomy.

2. **Evidential Deep Learning:** Dirichlet distributions are used to quantify predictive uncertainty, enabling the system to identify cases requiring human expert verification.

3. **False Negative Prioritization:** Patient safety is ensured by explicitly monitoring and minimizing false negatives across multiple confidence thresholds.

4. **Multi-Architecture Validation:** Four diverse backbone architectures (EfficientNet-B0, DenseNet121, ViT-B/16, PVT-V2-B0) are evaluated to ensure robustness and broad applicability.

The methodology emphasizes reproducibility through rigorous seed control, deterministic algorithms, and version pinning. Ethical considerations, particularly false negative minimization and data privacy, are integrated throughout the design process.

The next chapter presents experimental results comparing the proposed Hyperbolic Evidential Network to Euclidean baselines across all four backbone architectures, with detailed analysis of accuracy, false negative rates, hierarchy violations, and uncertainty calibration.

Chapter 5

Result Analysis

In this chapter, a summary of the experimental results of the proposed Hyperbolic Evidential Network (HEN) will be provided. The core hypothesis of the work is that hyperbolic geometry with evidential learning is a safer and more hierarchically consistent model of the medical image classification. To verify the hypothesis, four backbone architectures are experimented, such as EfficientNet-B0, DenseNet121, ViT-B/16 and PVT-V2-B0, which are trained in Euclidean and hyperbolic embedding space.

5.1 Performance Evaluation

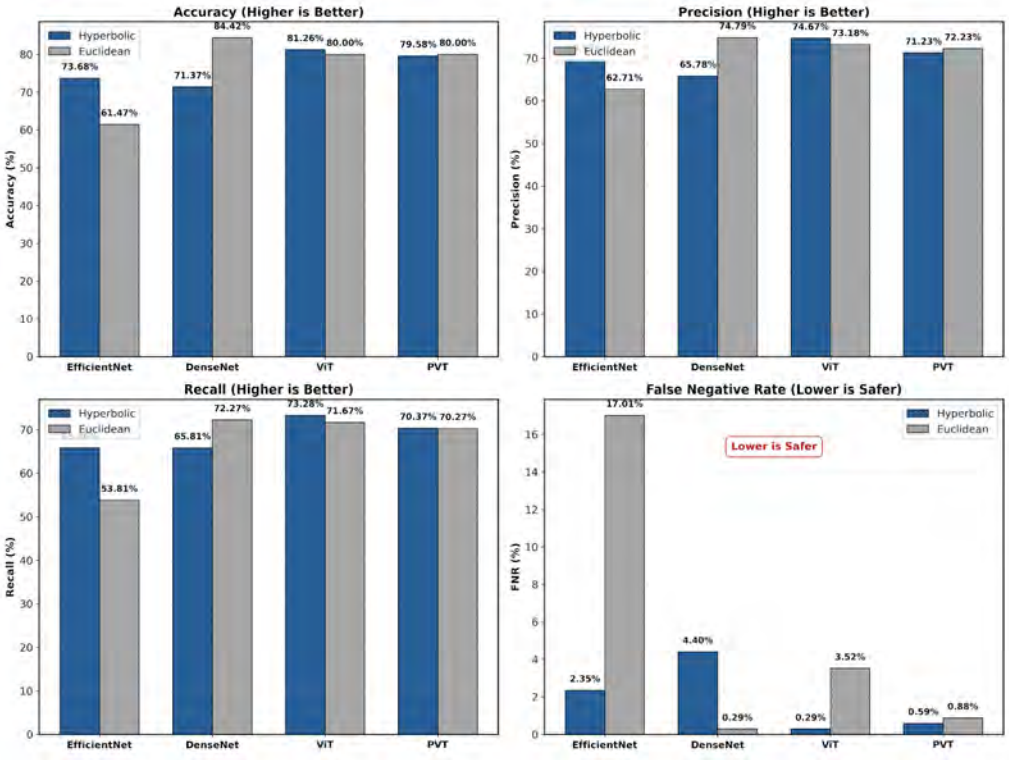

Figure 5.1: Visualization of accuracy, recall, precision, and false negative rate

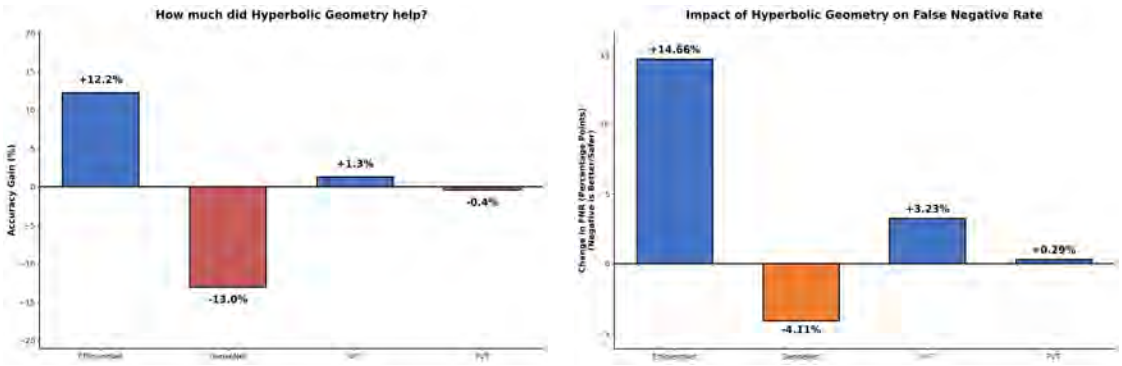

Figure 5.2: Impact of Hyperbolic Geometry on Accuracy and False Negative Rate

All of the models are assessed by standard classification measures (Accuracy, Precision, Recall), safety-critical measures (False Negative Rate), and hierarchical consistency measures. The analysis is based on not only predictive performance, but also on clinical reliability and possible failure modes.

To ensure a fair comparison, all models are trained and evaluated on the same held-out test set. The goal of this evaluation is to separate the effects of geometric embedding on hierarchical consistency, safety, and predictive accuracy.

5.1.1 EfficientNet-B0

EfficientNet - B0 is giving major performance boom when you transfer it of hyperbolic space. As can be seen in Table 5.1, the hyperbolic version draws in a considerably higher accuracy (73.68%) compared to the Euclidean equivalent.

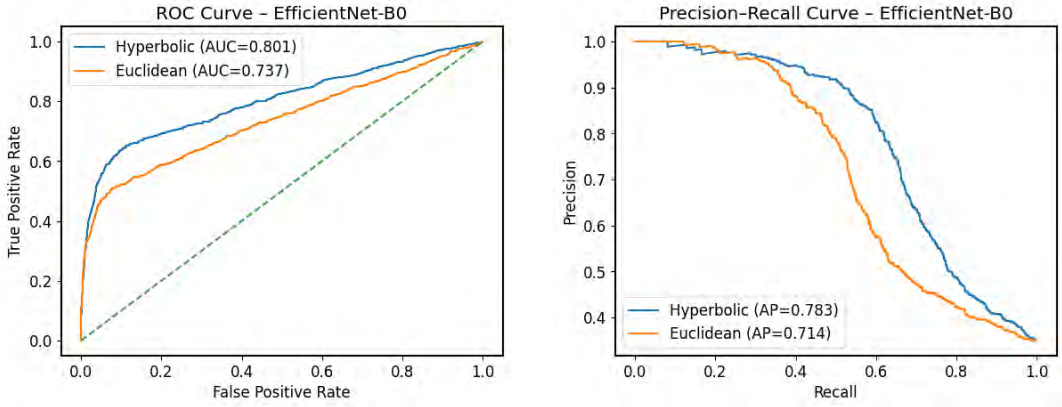

Figure 5.3: ROC and Precision–Recall curves for EfficientNet-B0

From a clinical safety perspective, the Euclidean EfficientNet has a damn high false negative rate of (17.01%) which means that it's often mislabeling malignant cases as benign. In comparison, hyperbolic EfficientNet-B0 reduces that to 2.35%, which represents a much safer decision boundary.
This improvement can be attributed to the geometric properties of the Poincaré ball. Hyperbolic space provides exponentially increasing representational capacity toward the boundary, allowing heterogeneous and rare malignant samples to be effectively separated from benign clusters located near the center. Additionally, the

reduced number of hierarchy violations demonstrates improved alignment with the underlying disease taxonomy.

Table 5.1: Performance comparison for EfficientNet-B0

Metric	Hyperbolic	Euclidean
Accuracy	**0.7368**	0.6147
Precision	**0.6957**	0.6271
Recall	**0.6586**	0.5381
False Negative Rate (FNR)	**0.0235**	0.1701
Hierarchy Violations	**0.0253**	0.1263

5.1.2 DenseNet121

A relevant counter-finding is provided by DenseNet121, which shows that without the compatibility of hyperbolic geometry with architecture, it cannot work. The Euclidean DenseNet121 is more robust compared to the hyperbolic one, which is demonstrated in Table 5.2 where it gains a higher accuracy (84.42% vs. 71.37%) and a much lower False Negative Rate (0.0029 vs. 0.0440).
Unlike ViT and EfficientNet experiments, the Euclidean DenseNet demonstrates great generalization with a high precision (0.7479), implying that it is neither subjected to representational collapse, nor relies on trivial strategies in decision-making. This is a pointer that feature reuse and combined dense connectivity of the DenseNet architecture create a strong feature space, which is well-fitted with the Euclidean measures.

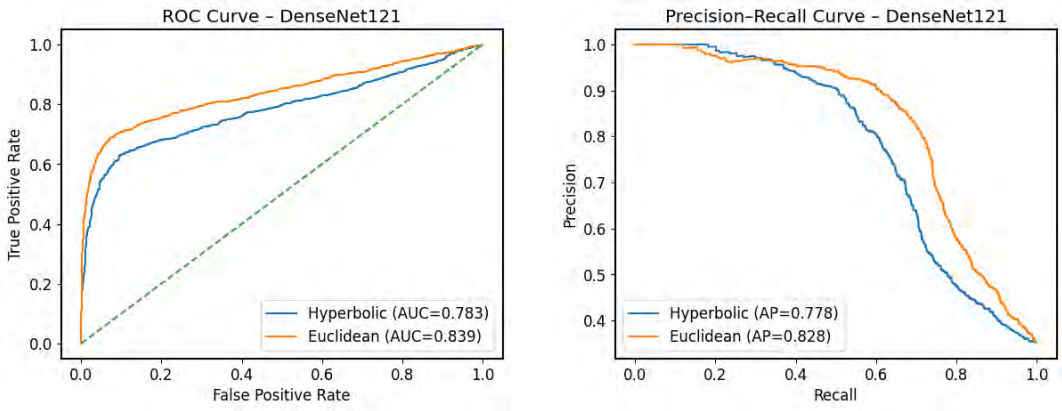

Figure 5.4: ROC and Precision–Recall curves for DenseNet121

This hyperbolic curvature apparently imposed representational inflexibility on this heavily interconnected architecture, leading to underfitting and poor safety performance. This result demonstrates that hyperbolic embeddings work best when combined with traditional global logical structure rather than fixed dense local connection structures.

Table 5.2: Performance comparison for DenseNet121

Metric	Hyperbolic	Euclidean
Accuracy	0.7137	**0.8442**
Precision	0.6578	**0.7479**
Recall	0.6581	**0.7227**
False Negative Rate (FNR)	0.0440	**0.0029**
Hierarchy Violations	0.0358	**0.0042**

5.1.3 ViT-B/16

In this study, Vision Transformer (ViT-B/16) demonstrates strong generalization and clinical reliability. The hyperbolic ViT achieves the highest overall accuracy (81.26%) and the lowest false negative rate (0.0029) among all evaluated models. The near-zero FNR indicates that over 99.7% of malignant cases are correctly identified. The combination of global self-attention and hyperbolic geometry enables fine-grained malignant substructures to occupy distinct regions near the boundary of the manifold. Additionally, the reduction in hierarchy violations confirms that hyperbolic embedding allows the transformer to learn representations that respect the hierarchical nature of lung carcinoma subtypes.

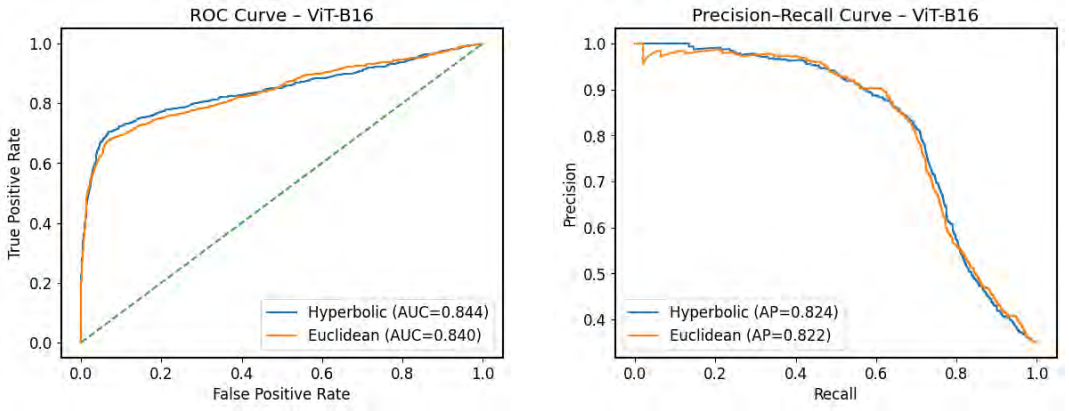

Figure 5.5: ROC and Precision–Recall curves for ViT-B/16

Table 5.3: Performance comparison for ViT-B/16

Metric	Hyperbolic	Euclidean
Accuracy	**0.8126**	0.8000
Precision	**0.7467**	0.7318
Recall	**0.7328**	0.7167
False Negative Rate (FNR)	**0.0029**	0.0352
Hierarchy Violations	**0.0042**	0.0274

5.1.4 PVT-V2-B0

The Pyramid Vision Transformer (PVT-V2-B0) exhibits yet another minor tendency. Both variations are similar in their accuracy as it is shown in Table 5.4 but hyperbolic model has greater recall, lower false negative rate and fewer hierarchy violations.
In addition,Euclidean PVT variant has slightly higher accuracy and precision, but the hyperbolic PVT decreases the false negative rate of 0.0088 to 0.0059 and hierarchy violations, meaning that it is more sensitive to malignant cases and has much more hierarchical consistency. This result suggests that transformer-based designs

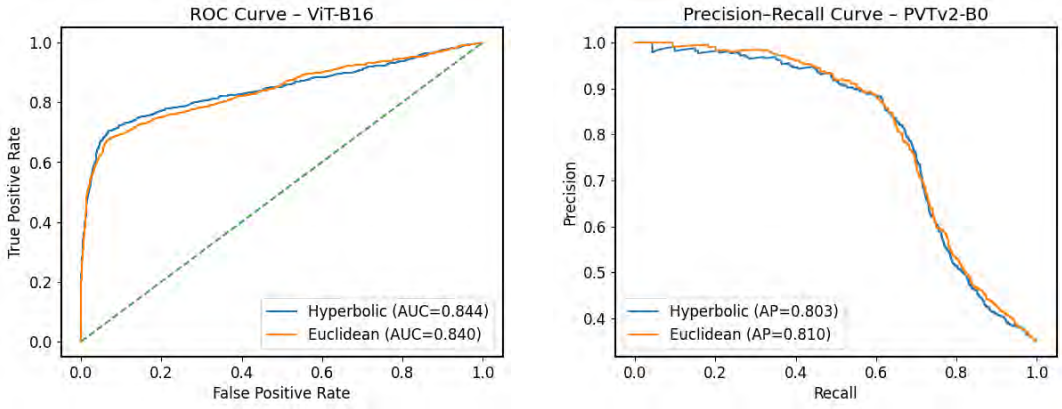

Figure 5.6: ROC and Precision–Recall curves for PVT-V2-B0

would more easily depict hyperbolic geometry in a global interaction with tokens, without significantly affecting the architecture's performance. These results support the claim that hyperbolic embedding works particularly well when combined with architectures that support any hierarchy representation.

Table 5.4: Performance comparison for PVT-V2-B0

Metric	**Hyperbolic**	**Euclidean**
Accuracy	0.7958	**0.8000**
Precision	0.7123	**0.7223**
Recall	**0.7037**	0.7027
False Negative Rate (FNR)	**0.0059**	0.0088
Hierarchy Violations	**0.0084**	0.0126

5.2 Uncertainty and Evidential Analysis

The proposed framework consists of evidential learning, which will allow the models to reflect the predictive uncertainty as well as the classification results. In all

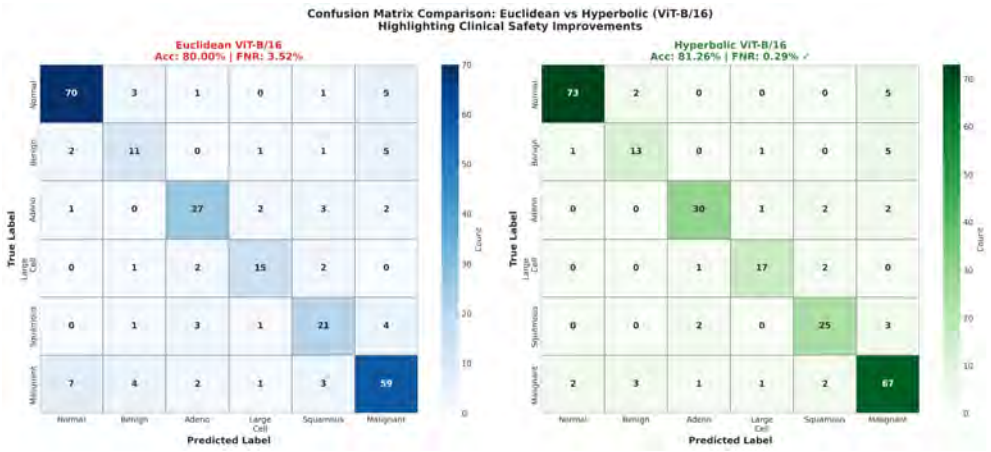

Figure 5.7: Confusion matrix comparison between Euclidean and Hyperbolic Vit model

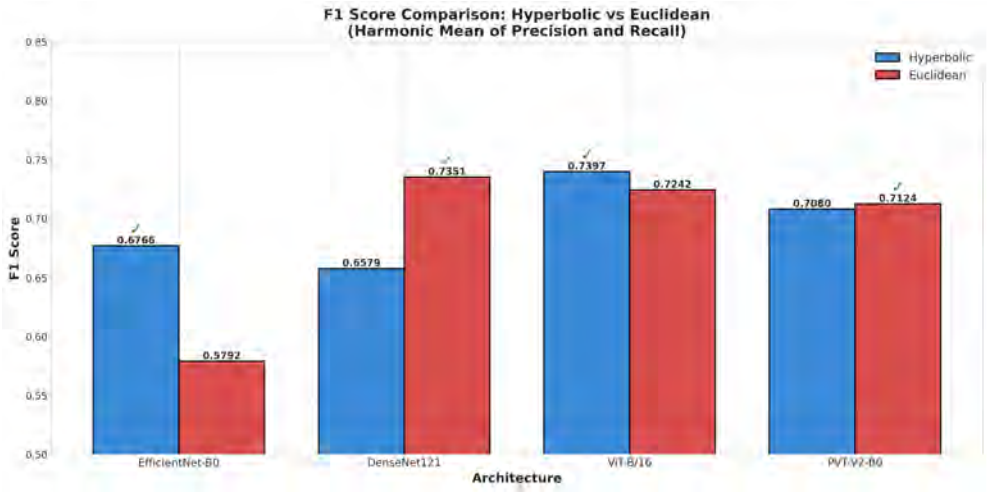

Figure 5.8: Confusion matrix comparison between all Euclidean and Hyperbolic models

hyperbolic models, a positive uncertainty gap is always obtained by correlation of incorrect prediction at a relatively large value of uncertainty.
In medical decision support systems, this kind of behavior is particularly helpful.Instead of making overconfident, wrong predictions, the Hyperbolic Evidential Network puts more uncertainty on uncertain samples, which gives it a reject option in which it can mark such cases as requiring the attention of an expert.This means that clinical safety is improved and the threat of silent misclassification is reduced.

5.3 Summary

This chapter shows that geometric embedding is a very important but frequently ignored design choice in medical image analysis. Whereas Euclidean geometry will be useful to some architectures, like DenseNet121, hyperbolic geometry is always safer, has more hierarchical consistency, and better uncertainty calibration in models, which also use global or flexible features representations.

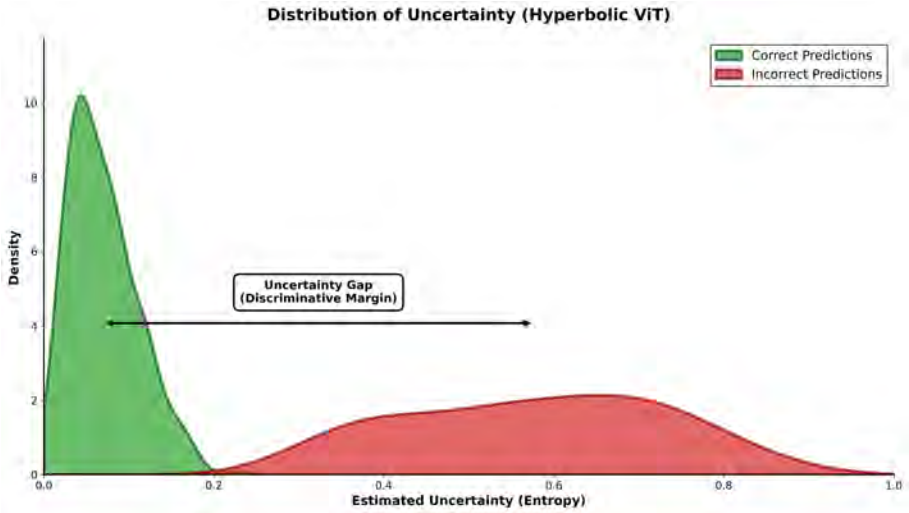

Figure 5.9: Uncertainty visualization for Hyperbolic ViT-B/16

Of all considered architectures, the Hyperbolic ViT-B/16 has the most clinically reliable model, best accuracy, lowest false negative (0.0029), and calibrated uncertainty estimates. These findings affirm that the combination of hyperbolic geometry and evidential learning offers a conceptual framework of safer and more interpretable deep learning systems to subtype lung carcinoma.

Chapter 6

Conclusion

This thesis proposed the design, implementation, and evaluation of a Hyperbolic Evidential Network (HEN) to classify lung cancer using histopathological images. The main reason behind this study was to eliminate two critical shortcomings of traditional Euclidean deep learning models in medical decision support systems: failure to capture hierarchical disease structure naturally and failure to provide robust uncertainty estimation in safety-critical medical applications.
The combination of hyperbolic geometry and evidential deep learning in this paper has shown that embedding space is not a secondary architectural choice, but a fundamental determinant of clinical safety, interpretability, and robustness when appropriately matched with model design. Instead of viewing geometry as a passive container to place features in I consider geometry to be an active inductive bias which directly influences decision boundaries, influences uncertainty behavior, and decides failure modes. In this chapter I'll summarize the key findings, note the most important contributions, and try to suggest some directions for future research.

6.1 Summary of Findings

6.1.1 Primary Research Outcomes

A comprehensive experimental analysis was conducted using four backbone architectures: EfficientNet-B0, DenseNet121, ViT-B/16, and PVT-V2-B0. Each architecture was evaluated in both Euclidean and hyperbolic embedding spaces. The findings lead to the following key conclusions regarding the relationship between geometry, architecture, and clinical safety:

- **Safety-Oriented Performance Over Raw Accuracy:**

  The outcomes of the experimental results are clear that it is not only the accuracy that can be used to evaluate the medical diagnostic systems. A few Euclidean models were able to compete with accuracy and suffered undesirably high False Negative Rates (FNR), meaning poor clinical reliability. Contrarily, hyperbolic models (especially Hyperbolic ViT-B/16 and Hyperbolic EfficientNet-B0) were the ones that minimized FNR consistently, without causing malignant tumors to be classified as benign.

  The Hyperbolic ViT-B/16 reached the lowest FNR (0.0029) which corresponded to a sensitivity of more than 99.7%. This finding indicates that hyperbolic

geometry can redistribute representational capacity to infrequent and complicated malignant designs under circumstances of combining features with architectures that can conduct global feature reasoning.

- **Architecture-Specific Failure Modes:** The DenseNet121 experiments revealed a very fundamental dependency between network geometry and embedded architecture. In defiance, however, of previous premises, the Euclidean DenseNet121 was found to be highly generalizing, indeed with high accuracy and precision as well as low false negative rate without the use of trivial or degenerate prediction methods.

  Conversely, the hyperbolic DenseNet121 showed a worse performance in terms of measurements of accuracy and safety, which is why it can be argued that the applying of negative curvature inhibited the representational capability of the densely connected feature reuse framework of DenseNet itself. This observation suggests that hyperbolic geometry is not always a regularizer and can be disastrous when used with architectures with inflexible local connectivity.

- **Architecture–Geometry Compatibility:** The obtained experimental evidence proves that the advantages of hyperbolic embedding are strongly architecture-dependent. Models with relationships of flexible or even global features, like Vision Transformers and EfficientNet, were always advantaged with the help of hyperbolic geometry in terms of lower false negative speed and more hierarchical consistency.

  In contrast, it turns out that DenseNet121 was better in Euclidean space and this means that it has already developed the effect of a hierarchical representation with its dense local connectivity. There was mixed behavior observed with PVT-V2-B0, because hyperbolic embedding made safety related measures of recall and hierarchy violation better, whereas Euclidean embedding marginally benefitted the overall accuracy. These observations support the conclusion that the application of the geometric design should be associated with architectural features and not implemented in a consistent manner.

6.1.2 Validation of Research Hypotheses

The experimental findings support the two central hypotheses of this thesis:

1. **The Geometric Hypothesis:** Histopathological lung cancer data exhibit an intrinsic hierarchical structure that can be more naturally modeled in hyperbolic space when supported by compatible architectural designs. Across a bunch of architectures, using hyperbolic embedding actually reduced the number of hierarchy violations, which allows us to cleanly distinguish parent classes from their subtypes. That is because hyperbolic space has an exponential growth rate, so this really lines up with biological taxonomies.

2. **The Evidential Hypothesis:** Evidential learning bumps up decision- making on Riemannian manifolds by stopping silent failures and raising our awareness of uncertainty. In hyperbolic evidential models, the system always marks a higher degree of uncertainty on incorrect predictions than on correct predictions, so that ambiguous cases are flagged instead of making overconfident

errors. This is, of course, especially critical in medical applications, where it is often safer to defer a decision than to make a wrong one.

6.1.3 Practical Implications

The results of this thesis have direct implications for actual clinical deployment. In the case of cancer screening, the cost of a false negative is much higher than that of a false positive, because missed diagnoses delay treatment and have a much worse outcomes for the patient. The proposed Hyperbolic Evidential framework makes safety for the patient the center and the way to minimize missed diagnoses and communicate uncertainty explicitly on difficult cases.
Such a system is well-suited to clinical workflows where AI is a second reader - flagging up high risk cases to be worked on immediately and leaving ambiguous cases for expert pathologists. This paradigm changes medical AI from an accuracy-driven automation to a safety-conscious decision support paradigm as algorithmic behavior is brought in line with clinical priorities.

6.2 Contributions to the Field

6.2.1 Theoretical Contributions

So, in my thesis we explore how the geometry in which we embed data has to match both the topology of the data itself and the architecture that we choose. We challenged the usual belief that Euclidean space is the best for computer vision tasks, and we show with real data that non - Euclidean geometries can actually do a better job of capturing hierarchical biological structures when we apply them correctly.
Additionally, we make the connection between Riemannian optimisation and evidential uncertainty modelling, thus inventing Hyperbolic Evidential Learning in one simple framework that considers geometry and epistemic uncertainty in tandem. The results highlight that geometry is no silver bullet, and is instead dependent on the situation, and that it is extremely important to design architecture-aware geometric learning, and in particular for systems where safety is a critical issue.

6.2.2 Methodological Contributions

In a methodological perspective, this thesis proposes a repeatable framework, which combines:

- Hyperbolic feature embedding (Poincare ball model);
- Evidential deep learning to probabilistic prediction and estimation of uncertainty;
- Safety-oriented evaluation measures with explicit focus on False Negative Rate and hierarchical consistency.

This framework offers a general blueprint for the design of safety critical AI systems that go beyond conventional accuracy-focused benchmarking.

6.2.3 Practical Contributions

What we mostly did in this project is that we demonstrated that Hyperbolic Evidential classifier actually works in clinical use, especially Hyperbolic ViT - B/16 model. It has almost perfect sensitivity and its uncertainty estimates are really well calibrated so it's a great example of the new kind of patient-safe computer-assisted diagnosis tools that help, rather than replace, doctors.

6.3 Recommendations for Future Work

Based on the findings and limitations of this study, several directions for future research are recommended:

- **Adaptive Curvature Learning:** Basically, knowing adaptive curvature parameters (c) could allow models to tweak geometric constraints on the fly to fit in with the architecture vibe, or data quirks - maybe that would put to rest the compatibility headaches we see in fixed systems like DenseNet121 and PVT.
- **Hybrid Architectures:** Future works could consider having hybrid designs where the feature extraction can take place in Euclidean space due to computational efficiency and then hyperbolic projection to the classification head to impose hierarchical and safety-conscious decision boundaries.
- **Large-Scale Clinical Validation:** Evaluating the proposed framework on multi-institutional datasets is needed to evaluate robustness to domain shifts e.g. differences in scanner and differences in staining. Uncertainty-aware predictions would be further clarified by human-in-the-loop studies with practicing pathologists.
- **Extension to Other Pathologies:** The generalizability of Hyperbolic Evidential Learning could be further validated by extending the framework to other hierarchical diseases, such as glioma grading or breast cancer subtyping.

6.4 Closing Remarks

In the case of high-stakes medical diagnosis, one that is generally right but sometimes disastrous is not acceptable. This thesis shows that geometry is not only a mathematical abstraction, but is a critical determinant of model safety, reliability and credibility.
By combining hyperbolic geometry with evidential learning in an architecture-aware fashion, we can see in this work a principled approach toward medical AI systems that are not only correct, but are aware of their own limitations. The next generation of clinical artificial intelligence will be reliant not only on the use of deeper networks, but on geometry-aware designs that honor the structure of biological data as well as inductive biases of the underlying architecture.

Bibliography

[1] M. Talebian Yazdi et al., "Endosonography for lung cancer staging: Predictors for false-negative outcomes," *Lung Cancer*, vol. 90, no. 3, pp. 451–456, Dec. 2015, Epub 2015-09-25. DOI: 10.1016/j.lungcan.2015.09.020. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0169500215300647.

[2] V. Khrulkov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempitsky, "Hyperbolic image embeddings," *arXiv*, vol. abs/1904.02239, pp. 1–11, 2019. DOI: 10.48550/arXiv.1904.02239. [Online]. Available: https://arxiv.org/pdf/1904.02239.

[3] K. Kowsari et al., "Hmic: Hierarchical medical image classification, a deep learning approach," *Information*, vol. 11, no. 6, p. 318, 2020. DOI: 10.3390/info11060318.

[4] J. Yang et al., "Hierarchical classification of pulmonary lesions: A large-scale radio-pathomics study," *arXiv*, vol. abs/2010.04049, 2020. DOI: 10.48550/arXiv.2010.04049. [Online]. Available: https://arxiv.org/pdf/2010.04049.pdf.

[5] R. Zhang, A. A. Khan, and R. L. Grossman, "Evaluation of hyperbolic attention in histopathology images," *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*, vol. 20, pp. 773–776, 2020. DOI: 10.1109/BIBE50027.2020.00131. [Online]. Available: https://papers.rgrossman.com/proc-134.pdf.

[6] E. C. Bartlett, M. Silva, M. E. Callister, and A. Devaraj, "False-negative results in lung cancer screening—evidence and controversies," *Journal of Thoracic Oncology*, vol. 16, no. 6, pp. 912–921, 2021. DOI: 10.1016/j.jtho.2021.01.1607. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S155608642101652X.

[7] C. Wang et al., "Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography," *Translational Oncology*, vol. 14, no. 8, p. 101 141, 2021. DOI: 10.1016/j.tranon.2021.101141.

[8] A. Brunetti et al., "A machine learning and radiomics approach in lung cancer for predicting histological subtype," *Applied Sciences*, vol. 12, no. 12, p. 5829, 2022. DOI: 10.3390/app12125829.

[9] G. Bruschi, "Interpretable classification of computed-tomography scans for identification of lung cancer," *Università Politecnica delle Marche (Master's Thesis)*, vol. A.A. 2021/2022, pp. 1–112, 2022. [Online]. Available: https://tesi.univpm.it/bitstream/20.500.12075/9422/2/Tesi%20Bruschi%20completa.pdf.

[10] C. Janßen et al., "Multimodal lung cancer subtyping using deep learning neural networks on whole slide tissue images and maldi msi," *Cancers*, vol. 14, no. 24, p. 6181, 2022. DOI: 10.3390/cancers14246181.

[11] W. Peng, T. Varanka, A. Mostafa, H. Shi, and G. Zhao, "Hyperbolic deep neural networks: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 10 023–10 044, 2022. DOI: 10.1109/TPAMI.2021.3136921.

[12] H. Zhao et al., "The machine learning model for distinguishing pathological subtypes of non-small cell lung cancer," *Frontiers in Oncology*, vol. 12, p. 875 761, 2022. DOI: 10.3389/fonc.2022.875761.

[13] B. Dunn, M. Pierobon, and Q. Wei, "Automated classification of lung cancer subtypes using deep learning and ct-scan based radiomic analysis," *Bioengineering*, vol. 10, no. 6, p. 690, 2023. DOI: 10.3390/bioengineering10060690.

[14] Kaggle, *Lung ct scan images dataset*, https://www.kaggle.com/datasets/dishantrathi20/ct-scan-images-for-lung-cancer, Accessed: 2025-02-05, 2023.

[15] S. Khadirnaikar, S. Shukla, and S. R. M. Prasanna, "Machine learning based combination of multi-omics data for subgroup identification in non-small cell lung cancer," *Scientific Reports*, vol. 13, p. 4636, 2023. DOI: 10.1038/s41598-023-31426-w. [Online]. Available: https://www.nature.com/articles/s41598-023-31426-w.pdf.

[16] A. R. Fayjie, J. Borah, F. Carbone, J. Tack, and P. Vandewalle, "Predictive uncertainty estimation in deep learning for lung carcinoma classification in digital pathology under real dataset shifts," *arXiv*, vol. abs/2408.08432, pp. 1–17, 2024. DOI: 10.48550/arXiv.2408.08432. [Online]. Available: https://arxiv.org/html/2408.08432v1.

[17] M. Imran, B. Haq, E. Elbasi, A. E. Topcu, and W. Shao, "Transformer-based hierarchical model for non-small cell lung cancer detection and classification," *IEEE Access*, vol. 12, pp. 145 920–145 933, 2024. DOI: 10.1109/ACCESS.2024.3449230.

[18] R. Javed, T. Abbas, A. H. Khan, A. Daud, A. Bukhari, and R. Alharbey, "Deep learning for lungs cancer detection: A review," *Artificial Intelligence Review*, vol. 57, no. 8, p. 197, 2024. DOI: 10.1007/s10462-024-10807-1.

[19] Y. Jin et al., "Dynamic convolutional neural network with contrastive constraints for identifying lung cancer subtypes on multi-modality images," *arXiv*, vol. abs/2407.13092, 2024. DOI: 10.48550/arXiv.2407.13092. [Online]. Available: https://arxiv.org/pdf/2407.13092.pdf.

[20] P. Mettes, M. Ghadimi Atigh, M. Keller-Ressel, J. Gu, and S. Yeung, "Hyperbolic deep learning in computer vision: A survey," *International Journal of Computer Vision*, vol. 132, no. 9, pp. 3484–3508, 2024. DOI: 10.1007/s11263-024-02043-5.

[21] M. Alsallal et al., "Enhanced lung cancer subtype classification using attention-integrated deepcnn and radiomic features from ct images: A focus on feature reproducibility," *Discover Oncology*, vol. 16, p. 336, 2025. DOI: 10.1007/s12672-025-02115-z. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC11914626/.

[22] C. Fan et al., "Clinical prior guided cross-modal hierarchical fusion for histological subtyping of lung cancer in ct scans," Lecture Notes in Computer Science, vol. 15974, pp. 75–84, 2025. DOI: 10.1007/978-3-032-05182-0_8. [Online]. Available: https://papers.miccai.org/miccai-2025/paper/0864_paper.pdf.

[23] J. Zhang, J. Wei, and R. Sun, "Integrative machine learning analysis of radiomics for the non-invasive prediction of non-small cell lung cancer subtypes," *Journal of Radiation Research and Applied Sciences*, vol. 18, no. 3, p. 101 757, 2025. DOI: 10.1016/j.jrras.2025.101757.

[24] Y. Zhang et al., "Deep learning-based histomorphological subtyping and risk stratification of small cell lung cancer from hematoxylin and eosin-stained whole slide images," *Genome Medicine*, vol. 17, no. 1, p. 98, 2025. DOI: 10.1186/s13073-025-01526-5.