
Decision Tree Models for Managing and Predicting Cancer Patient Outcomes: A Data-Driven Managerial Approach

Atmane medini*

Laboratory of Renewable Energy Economics and Its Role in Achieving Sustainable Development, University of El Oued, Algeria

medini-atmane@univ-eloued.dz

Received: 05/04/2026

Accepted: 13/05/2026

Published: 31/05/2026

Abstract:

This study aimed to develop and evaluate decision tree classifiers for predicting cancer patient outcomes using demographic, clinical, and socioeconomic variables, focusing on model interpretability and generalizability. A dataset of 10,000 cancer patient records, including factors such as age, gender, cancer type, treatment type, comorbidities, and anthropometric features, was analyzed. Four models were developed: one unconstrained, two pruned with different complexity levels, and one trained and validated with an 80/20 split. Performance was assessed through accuracy, precision, recall, and F1-score, with feature importance identifying key predictors. The unconstrained model overfitted despite perfect training accuracy, while pruned models showed better generalization, and the optimal pruned model balanced accuracy and interpretability. Height, weight, and age were consistent primary predictors. Overall, pruned and validated decision tree models proved transparent, reliable, and suitable for supporting personalized clinical decision-making in oncology.

Keywords: Decision trees; Cancer outcome prediction; Artificial intelligence in healthcare; Interpretability; Predictive modeling

Jel Classification Codes : C45, C53, I10, I19, C55.

*Corresponding author

1. Introduction:

Cancer represents one of the most significant global health challenges due to its high incidence rates and the complex health, economic, and social repercussions it entails. Despite substantial advances in early diagnostic technologies and targeted therapeutic strategies, accurately predicting patient outcomes remains both a scientific and managerial challenge. This difficulty stems from the disease's dynamic nature and the intricate interaction of biological, clinical, and demographic factors that jointly shape prognosis (Li et al., 2025; Wang et al., 2010).

In response to this complexity, machine learning (ML) techniques have increasingly been adopted as analytical tools capable of handling high-dimensional, nonlinear medical data, particularly in oncology. Among these techniques, decision tree algorithms have attracted considerable attention for their flexibility, ability to model nonlinear relationships, and compatibility with personalized medicine approaches that tailor decisions to individual patient profiles (Banegas-Luna et al., 2021; Lundberg et al., 2020).

A key advantage of decision trees lies in their inherent interpretability. By explicitly revealing the logical structure underlying predictions, these models enable clinicians, healthcare managers, and data analysts to understand how specific variables contribute to decision outcomes. This transparency is critical not only in clinical decision-making but also in managerial and business-oriented contexts, where predictive models inform strategic planning, cost control, risk management, and resource allocation (Banegas-Luna et al., 2021).

Nevertheless, the practical deployment of decision tree models reveals a notable research gap. While existing studies largely emphasize predictive accuracy or clinical validity, comparatively limited attention has been paid to the managerial and analytical trade-offs involved in model design. In particular, insufficient research addresses how pruning strategies and model calibration affect the balance between predictive performance, interpretability, and generalizability from the perspective of data analysts and healthcare decision-makers. Overly complex trees may yield high in-sample accuracy but poor robustness and misleading managerial signals, whereas excessive pruning may oversimplify the decision logic and obscure economically or clinically meaningful patterns (Lundberg et al., 2020).

In the broader context of explainable artificial intelligence (XAI), this gap becomes increasingly consequential. As organizations seek transparent, accountable, and fair decision-support systems, especially in healthcare and business analytics, there is a pressing need for frameworks that systematically reconcile accuracy with interpretability and operational relevance (Ghasemi et al., 2024). Addressing this challenge is essential for

healthcare management systems and data-driven organizations aiming to improve patient outcomes, enhance operational efficiency, and support evidence-based strategic decision-making.

2.Leterature Review

Recent scientific literature has highlighted significant associations between physical characteristics, such as height, body shape, and body mass index, and the risk of various cancers across multiple populations. For example, the Million Women Study in the UK (Green et al., 2011) demonstrated a relationship between height and overall cancer incidence. Similarly, a large-scale cohort study that included over 22 million Korean adults found that increased stature was linked to higher risks for several types of cancer (Lee et al., 2019). Moreover, a systematic review and meta-analysis covering 47 studies confirmed that attained adult height significantly correlates with the risk of colorectal cancer (Bélanger et al., 2022). Investigations into genetic factors, such as those by Seviiri et al. (2016), have explored the relationship between height-related single nucleotide polymorphisms and non-melanoma skin cancer risk. On a broader scale, studies examining body shape phenotypes and combinations of anthropometric traits revealed their relevance to cancer susceptibility across diverse populations (Martín-González et al., 2022). Research in Korea has combined height with mammographic density data to investigate breast cancer risk, further substantiating physical attributes as influential risk factors (Kim et al., 2024). In Germany, a retrospective analysis involving over 780,000 outpatients established links between body height and the likelihood of developing various cancers (Dreher et al., 2022). Finally, Mendelian randomization trials utilizing data from the UK Biobank and international consortia confirmed that body size and composition significantly influence the risk of multiple site-specific cancers (Davies et al., 2021). These converging lines of evidence reinforce the importance of evaluating physical and lifestyle factors in cancer epidemiology.

3.Theoretical Framework

Cancer epidemiology recognizes that patient outcomes arise from a complex interplay of biological, clinical, demographic, and socioeconomic factors. Within this multifactorial context, anthropometric characteristics such as height, body mass index (BMI), and body shape have consistently been identified as significant risk indicators. From a theoretical standpoint, these physical attributes can be interpreted as observable proxies for underlying biological mechanisms, including cumulative cellular proliferation, hormonal regulation, and metabolic signaling pathways that influence carcinogenesis.

In line with the cellular division hypothesis, cancer risk may be conceptualized as an increasing function of cumulative cell replications over time:

$$P(\text{Cancer}) \propto \sum_{i=1}^T N_{\text{cells}}(t) \mu$$

where $N_{\text{cells}}(t)$ denotes the number of replicating cells at time t , and μ represents the mutation rate. Empirical evidence from large-scale cohort studies, meta-analyses, and Mendelian randomization designs consistently supports this theoretical foundation, demonstrating that increased body size and specific anthropometric profiles are associated with elevated cancer risk across multiple sites and populations. In epidemiological practice, these relationships are commonly formalized using relative-risk or hazard-based models of the form:

$$RP + \exp(\beta_1 \cdot \text{Height} + \beta_2 \cdot \text{BMI} + \beta_3 \cdot \text{BodyDhape} + \varepsilon)$$

Despite their widespread use, such linear formulations impose strong assumptions of additivity and monotonic effects. Growing empirical evidence suggests that cancer risk is inherently nonlinear and characterized by interaction and threshold effects among anthropometric and clinical variables, thereby limiting the explanatory and predictive capacity of conventional regression-based models.

In response to this complexity, data-driven decision-making in healthcare management increasingly relies on interpretable machine learning frameworks capable of capturing nonlinear structures while preserving transparency. Decision tree algorithms operationalize cancer risk assessment as a sequence of conditional partitions:

$$f(x) = \begin{cases} r_1 & \text{if } X_1 \leq C_1 \\ r_2 & \text{if } X_1 > C_1 \text{ and } X_2 \leq C_2 \\ \vdots & \\ \vdots & \end{cases}$$

where X represents anthropometric and clinical input variables, and $c_{k_c}k_{c_k}$ denote data-driven decision thresholds. Through pruning and model calibration, these rule-based structures balance predictive accuracy with interpretability and generalizability—an essential requirement in high-stakes healthcare environments. Consequently, this framework positions decision tree models not merely as technical prediction tools, but as theory-consistent decision-support instruments that bridge epidemiological risk theory, statistical learning, and evidence-based healthcare management.

4.Previous Studies

Previous research confirms the relevance of decision tree–based models in cancer prediction due to their balance between accuracy and interpretability. Li et al. (2025) demonstrated that CART models using clinical and biomarker data achieve strong predictive performance for ovarian cancer. Wang et al. (2010) showed that integrating sampling and feature selection techniques improves fairness and reliability in imbalanced breast cancer datasets. Banegas-Luna et al. (2021) emphasized that pruning and model tuning are essential to preserve interpretability and generalizability in medical applications. Complementarily, Lundberg et al. (2019) and Ghasemi et al. (2024) highlighted the role of SHAP-based explainability in enhancing transparency and trust in tree-based predictive models used in healthcare.

5.Methodology and dataset:

A Decision Tree Classifier was utilized to model the relationships between socioeconomic indicators and the categorical target variable. Decision Tree Classification is a non-parametric supervised learning method that recursively partitions the dataset into subsets based on feature thresholds, aiming to maximize the homogeneity of the target class within each resulting subset (Pedregosa et al., 2011). At each node, the algorithm selects the optimal feature and threshold by maximizing classification purity criteria, such as Gini impurity or information gain (entropy) (GeeksforGeeks, 2023). This process is repeated in a greedy, top-down manner until a predefined stopping criterion—such as a maximum tree depth or a minimum number of samples per leaf—is reached. The final classification for each sample is determined by the majority class in the leaf node to which it is assigned, providing a clear, interpretable mapping from input features to predicted classes.

Model performance was validated by partitioning the data into separate training and testing subsets and assessing standard classification metrics such as accuracy, precision, recall, and F1-score. Hyperparameters—including maximum tree depth, minimum samples per split, and minimum samples per leaf—were tuned systematically to balance classification accuracy and model generalizability and to mitigate overfitting (Pedregosa et al., 2011). The decision tree framework thus provides a robust and interpretable approach, especially suitable for classification problems characterized by complex or nonlinear relationships between predictors and outcomes.

■ Dataset

The empirical analysis in this study is based on the *Cancer Data of the UAE* dataset, which is publicly available through Kaggle. The dataset comprises approximately 10,000 patient-level observations and includes a wide range of demographic, socioeconomic, clinical, and anthropometric variables relevant to cancer outcome analysis. These variables include age, gender, nationality, emirate of residence, ethnicity, smoking status, comorbidities, height, and weight, as well as clinical attributes such as cancer type, cancer stage, treatment modality, hospital of treatment, and recorded patient outcome. In this study, all demographic, socioeconomic, clinical, and anthropometric variables are treated as independent inputs, while patient health outcome serves as the dependent variable. The open-access nature of the dataset supports transparency and reproducibility in data-driven cancer research. Details regarding dataset access are provided in the Data Availability section.

■ Data Availability

The dataset analyzed during the current study is publicly available on Kaggle under the title : *Cancer Data of the UAE* and can be accessed at: <https://www.kaggle.com/>

5. Results and Discussion

5.1. Unconstrained Decision Tree Results (Model_1)

In this initial experiment, we developed a decision tree classifier (referred as model_1) using the complete dataset without applying any constraints or restrictions (e.g., no depth limit, no minimum samples per leaf, etc.). All available variables, excluding the outcome variable, which serves as the dependent variable, were included as independent predictors. This unconstrained approach allows the model to fully explore the complexity and potential interactions present in the data.

Table:1. Structure of model_1(without constraints)

Decision Tree Model Evaluation	
Tree Depth	39
Min_Number of Leaves	3377

Source: Author elaboration based on Python output

The Model 1 was developed as a fully unconstrained decision tree classifier utilizing all available variables as predictors and the "Outcome" variable as the dependent variable. The analysis was performed on a dataset comprising 10,000 observations, ensuring adequate statistical power. The resulting tree model reached a depth of 39 and produced 3,377 terminal leaves. Such a high depth and number of leaves indicate that the model is highly complex, having partitioned the dataset into a large number of very small groups, with each leaf containing only a few observations on average. While this maximal growth allows the model to closely fit patterns and potentially subtle structures in the training data, it also raises a significant risk of overfitting. Overfitting occurs when a model captures not only the underlying relationships but also random noise specific to the training data, thereby reducing its ability to generalize to new, unseen data. Furthermore, the model's high complexity compromises its interpretability, making it challenging to extract clear, actionable insights or draw general conclusions. As a result, although the unconstrained approach provides maximum flexibility, it emphasizes the necessity of introducing constraints or regularization in subsequent models to achieve better generalizability and interpretability.

Table2. Confusion Matrix and Classification Report for Model_1

Actual \ Predicted	Deceased	Recovered	Under Treatment	Precision	Recall	F1-Score	Support
Deceased	992	0	0	1.00	1.00	1.00	992
Recovered	0	4931	0	1.00	1.00	1.00	4931
Under Treatment	0	0	4077	1.00	1.00	1.00	4077

Accuracy						1.00	10000
Macro Avg				1.00	1.00	1.00	10000
Weighted Avg				1.00	1.00	1.00	10000

Source: Authors' elaboration based on Python output

The decision tree classifier (Model 1) was trained using the complete "Cancer Data of the UEA" dataset, which contains 10,000 patient-level observations and 13 predictor variables spanning demographic, clinical, and socioeconomic factors. The model was constructed without any constraints on complexity, resulting in a tree depth of 39 and a total of 3,377 terminal leaves. The confusion matrix and classification metrics (precision, recall, and F1-score all equal to 1.00) demonstrate that the model achieved 100% accuracy on the training data, with every observation correctly classified into one of the three outcome categories: "Deceased," "Recovered," or "Under Treatment." While this result may appear ideal, it strongly suggests overfitting, as a tree this deep and with so many leaves effectively memorizes the training data, capturing both true patterns and random noise. Consequently, the model's ability to generalize to new, unseen data is likely to be very low. Scientifically, such a highly complex model cannot be meaningfully interpreted or used to guide clinical or research decisions, as its predictions are based on highly specific data combinations rather than broadly applicable trends. To develop a more useful and robust predictive model, it is necessary to introduce regularization strategies such as limiting tree depth or applying pruning, and to validate performance on external or unseen datasets. This will help ensure the model identifies genuine, generalizable relationships relevant to cancer outcomes, making it more valuable for scientific and clinical applications.

Table:3 Feature Importance Ranking in Decision Tree Model_1 for Cancer Prediction

Index	Feature Name	Feature Importance
0	Age	0.126951
1	Weight	0.124495
2	Height	0.142117
3	Gender_Female	0.011852
4	Gender_Male	0.014943
5	Cancer_Type_Breast	0.010925
6	Cancer_Type_Colorectal	0.013477
7	Cancer_Type_Leukemia	0.013081

Source: Author elaboration based on Python output

The feature importances derived from model_1 provide valuable insights into which variables played the most significant roles in predicting patient outcomes within the decision tree classifier. Notably, "Height," "Age," and "Weight" stand out as the most influential variables, as indicated by their relatively high

Table:4.structure of model_2(Contain constraints)

Decision Tree Model Evaluation	
Tree Depth	20
Min_number of Leaves	25
Min_number of Split	25

Source: Author elaboration based on Python output

The evaluation metrics for the pruned decision tree model (model_2) reveal a moderate predictive performance and a substantial improvement in generalizability compared to the unconstrained model. With an overall accuracy of approximately 0.59, the model demonstrates balanced capacity across the three outcome classes—“Deceased,” “Recovered,” and “Under Treatment.”.

Table:5: Confusion Matrix and Classification Report for the Pruned Model (model_2)

Actual \ Predicted	Deceased	Recovered	Under Treatment	Precision	Recall	F1-Score	Support
Deceased	9	564	419	0.36	0.01	0.02	992
Recovered	7	3545	1379	0.60	0.72	0.66	4931
Under Treatment	9	1753	2315	0.56	0.57	0.57	4077
Accuracy						0.59	10000
Macro Avg				0.51	0.43	0.41	10000
Weighted Avg				0.56	0.59	0.56	10000

Source: Authors' elaboration based on Python output

The pruned tree shows the highest recall and F1 score for the “Recovered” category (recall: 0.72, F1-score: 0.66), which also constitutes the largest class in the dataset. For the “Under Treatment” category, the model achieves moderate recall and F1-score values (recall: 0.57, F1-score: 0.57), while the “Deceased” category is predicted with notably low recall (0.01) and a minimal F1-score (0.02), indicating the model’s difficulty in identifying this minority class. The macro average F1-score (0.41) is lower than the weighted average (0.56), reflecting the imbalance in class distributions and greater predictive success in majority classes. The confusion matrix further highlights this pattern, with the majority of “Deceased” observations misclassified, while “Recovered” cases are identified with reasonable accuracy. Collectively, these results showcase the model’s trade-off: while pruning has reduced overfitting and yielded more realistic, generalizable estimates, it has also limited the tree’s capacity to capture rarer outcomes. This underscores the importance of further balancing sensitivity to minority classes, potentially through resampling or algorithmic enhancements, to achieve greater clinical utility and equitable predictive performance.

5.3.: Tree with Depth and Leaf Constraints of (Model_3)

The third model (Model_3) is a pruned model with constraints: a maximum depth of 18 and at least 7 samples required for each split and leaf. By limiting the model's complexity, these settings help reduce overfitting and enhance generalizability, ensuring the model captures key data patterns rather than noise. This approach improves both interpretability and reliability of outcome predictions, making the model more suitable for clinical and epidemiological applications.

Table: 5:Model Structure, Confusion Matrix, and Classification Report for Model_3 (with constraints)

Evaluation Details	Deceased	Recovered	Under Treatment	Precision	Recall	F1-Score	Support
Model Structure							
Tree Depth	18						
Min Leaves	7						
Min Split	7						
Confusion Matrix							
Actual: Deceased	152	491	349	0.47	0.15	0.23	992
Actual: Recovered	89	3754	1088	0.66	0.76	0.71	4931
Actual: Under Treatment	81	1436	2560	0.64	0.63	0.63	4077
Accuracy						0.65	10000
Macro Avg				0.59	0.51	0.52	10000
Weighted Avg				0.63	0.65	0.63	10000

Source: Authors' elaboration based on Python output

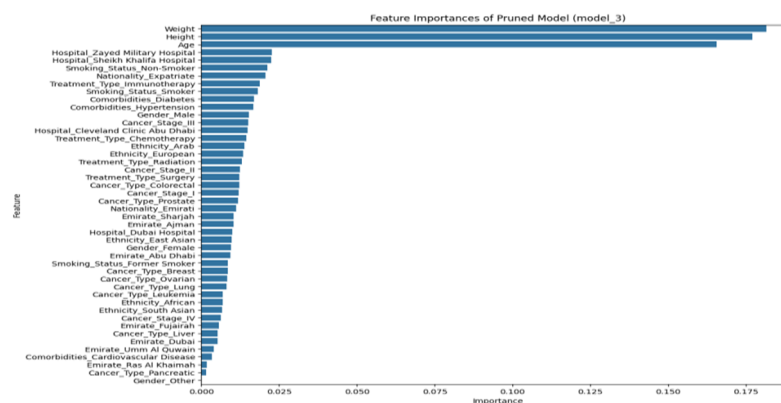
The model achieves its best classification results in predicting the "Recovered" outcome, evidenced by a high recall (0.76) and F1-score (0.71), which aligns with the sizable support for this class in the dataset. The "Under Treatment" class also displays reasonable predictive performance (recall: 0.63, F1-score: 0.63). In contrast, the "Deceased" class is associated with markedly lower recall (0.15) and F1-score (0.23), signifying the model's challenge in correctly identifying these minority instances. The macro-average and weighted-average values further reflect moderate balance between classes, yet highlight sensitivity to class imbalance. Collectively, these results demonstrate that the pruning constraints have successfully limited overfitting and enhanced generalizability compared to an unconstrained model, but further refinement—especially to improve detection of minority outcomes—may be warranted to achieve more equitable predictive performance across all patient categories.

Table: 6 Feature Importance's for(model_3)

Feature Name	Feature Importance
Age	0.165599
Weight	0.181521
Height	0.177072
Gender_Female	0.009595
Gender_Male	0.015310
Cancer_Type_Breast	0.008585
Cancer_Type_Colorectal	0.021512
Cancer_Type_Leukemia	0.006874

Source: Author elaboration based on Python output

The analysis of feature Importances for the decision tree classifier in the model_3, reveals that anthropometric variables—specifically Weight (0.182), Height (0.177), and Age (0.166)—are the most influential predictors in determining patient outcomes within this dataset. These features collectively account for the largest proportion of the model's decision splits, indicating that physical and physiological characteristics of patients dominate the predictive framework of the model. This prominence suggests that variations in these clinical measurements are strongly associated with the classification of patient outcomes, which underscores their critical role in the underlying disease dynamics or treatment responses captured by the model.

Fig:2: Feature Importance's for model_3

Source: Author elaboration based on Python output

In contrast, variables related to gender show minimal importance, with Gender_Male (0.015) and Gender_Female (0.010) contributing relatively little to model decisions, and Gender_Other having no influence (0.000). This indicates that gender identity exhibits limited discriminatory power in this predictive context. Similarly, nationality-related features, including Nationality_Expatriate (0.021) and

Nationality_Emirati (0.012), show modest influence, suggesting a minor role of patient nationality in outcome classification.

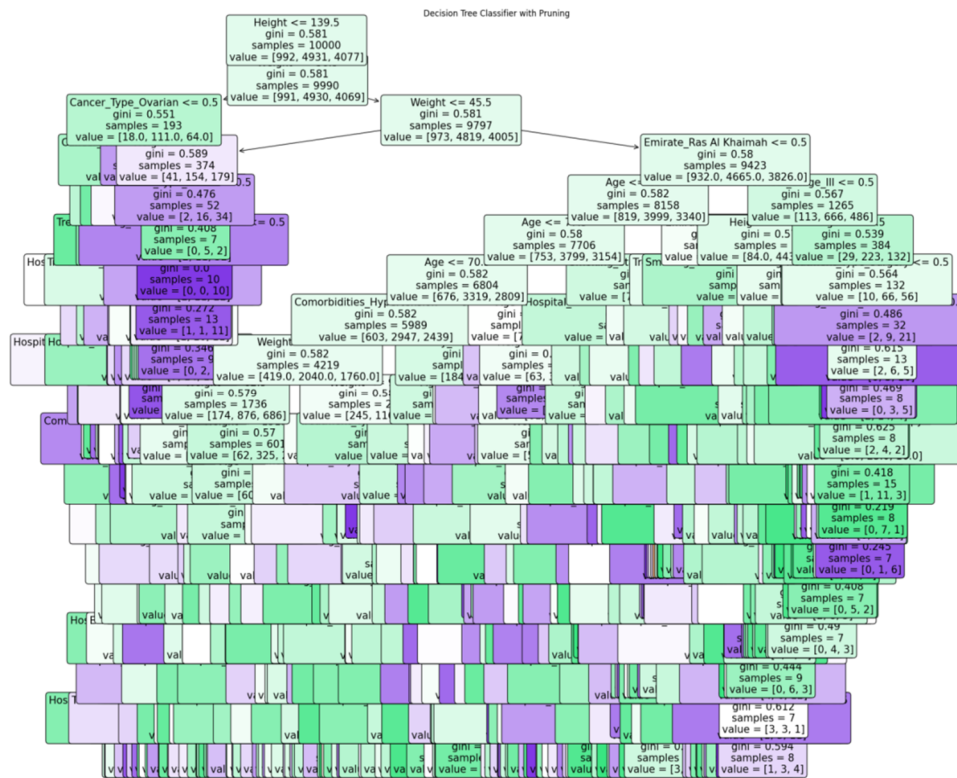
Fig:3 Accuracy Curve for Pruned Decision Tree Model (model_3)



Source: Author elaboration based on Python output

The plotted accuracy curves for model_3, depicting both training and validation accuracy as a function of decision tree depth, provide important insights into model behavior and the impact of complexity control. As the tree depth increases from shallow values, both training and validation accuracies rise, indicating that the model is learning increasingly relevant patterns from the dataset. However, beyond a certain depth, approximately between depths 8 and 12. The validation accuracy begins to plateau and eventually declines, even as training accuracy continues to approach perfection. This divergence between the training and validation curves is a classic indicator of overfitting: the model becomes highly specialized to the training data, capturing not only genuine structure but also noise and idiosyncratic details that do not generalize to new cases. The optimal predictive performance, as suggested by the peak of the validation accuracy curve, occurs at an intermediate tree depth, after which further increases in depth deteriorate validation accuracy. Thus, this curve underscores the necessity of applying pruning or depth constraints to balance model complexity and generalization, ensuring the model remains robust and predictive for unseen patient data rather than simply memorizing the training sample. This analysis justifies the selection of an appropriate maximum tree depth during model tuning and highlights the risk of relying solely on training accuracy as a measure of model quality.

Fig:4 Visualization of the decision tree for the (Model_3)



Source: Author elaboration based on Python output

The analysis of model_3 reveals that anthropometric variables—specifically weight (0.182), height (0.177), and age (0.166), are the most influential predictors in determining patient outcomes, dominating the decision splits and highlighting their strong association with disease dynamics. This is reflected in the tree's structure, where the root node begins with a threshold on "Height" (around 139.5 cm), indicating early prioritization of physiological characteristics. As the tree branches out, it incorporates clinical, demographic, and geographic variables—such as cancer type, comorbidities, and emirate of residence, demonstrating the model's attempt to capture complex, multi-level interactions. The model's depth was explicitly constrained (max_depth = 18) with regularization parameters (min_samples_leaf = 7, min_samples_split = 7), which limited overfitting while preserving meaningful patterns. Accuracy curves show that while both training and validation accuracy initially improve with tree depth, validation performance plateaus and slightly declines beyond a certain point, confirming the importance of controlling complexity. These constraints enhance the model's generalizability and help maintain clinical interpretability by avoiding excessive branching and sample fragmentation.

5.4. Model Splitting Strategy (80% Training, 20% Testing)

A baseline decision tree classifier (Model_1) was developed by training on the entire dataset without any constraints, serving as a reference for subsequent modeling strategies. Building on this, Model_2 employed pruning techniques by introducing constraints on the maximum tree depth (20), minimum number of samples at each leaf (25), and minimum samples required to split an internal node (25), resulting in a more regularized model structure. Model_3 further refined the pruning approach by limiting the maximum tree depth to 18, with a minimum of 7 samples required for both leaf nodes and splits.

In the subsequent model, the focus shifted toward evaluating generalizability by dividing the dataset into training and testing subsets using a standard train-test split. The decision tree classifier was then trained on the training data without any pruning constraints. By assessing the model's predictive performance on the separate test set, this approach provides a rigorous metric for evaluating overfitting and establishes a robust reference point for comparison with pruned models. This methodology allows for a comprehensive assessment of the trade-off between model complexity and generalization performance.

Table 8: Decision Tree Model Configurations

Model	Maximum Tree Depth	Minimum Samples per Leaf	Minimum Samples per Split	Pruning	Data Split
Model_1	Not Limited	Not Limited	Not Limited	No	Full Dataset
Model_2	20	25	25	Yes	Full Dataset
Model_3	18	7	7	Yes	Full Dataset
Model_4	Not Limited	Not Limited	Not Limited	No	Train/Test (80%/20%)

Source: Author' elaboration based on Python output

The table presents the architectural configurations employed across four decision tree models. Model_1 was developed using the entire dataset without imposing any constraints on tree depth, leaf size, or split size, establishing a baseline for comparison. Model_2 applied pruning constraints, limiting the maximum depth to 20 and requiring at least 25 samples for both leaves and splits. Model_3 further refined the pruning strategy with a maximum depth of 18 and a minimum of 7 samples required for both leaves and splits. Model_4 diverged from previous approaches by employing a train-test split (80% training, 20% testing) and omitting all pruning constraints, thereby allowing the tree to grow without restriction during training and facilitating rigorous evaluation of model generalizability against unseen data. This systematic variation of model complexity and validation strategy enables comprehensive analysis of the trade-off between structural flexibility and predictive performance.

5.4.1.Evaluation of the model trained

The evaluation results for Model_4, which was trained using an 80%/20% train-test split without pruning, are presented in terms of precision, recall, f1-score, and overall accuracy across three target classes: Deceased, Recovered, and Under Treatment. The model attained an overall accuracy of 44.6%, with a macro-averaged f1-score of 0.33 and a weighted average f1-score of 0.43, based on a total of 8,000 test examples.

Table:9: Combined Evaluation Table for Model_4 (Test Set)

	Predicted: Deceased	Predicted: Recovered	Predicted: Under Treatment	Precision	Recall	F1- Score	Support
Actual: Deceased	35	497	254	0.10	0.04	0.06	786
Actual: Recovered	176	2,404	1,368	0.49	0.61	0.54	3,948
Actual: Under Treatment	136	2,003	1,127	0.41	0.35	0.37	3,266
Accuracy						0.45	8,000
Macro Avg				0.33	0.33	0.33	8,000
Weighted Avg				0.42	0.45	0.43	8,000

Source: Author elaboration based on Python output

A closer examination of the class-specific metrics reveals substantial variation in classification performance. The model demonstrated the highest predictive effectiveness for the "Recovered" class, achieving a precision of 0.49, recall of 0.61, and f1-score of 0.54. This indicates a moderate ability to correctly identify instances of recovery, both in terms of detecting true positives and avoiding false positives. In contrast, classification performance for the "Deceased" class was considerably lower, with a precision of only 0.10, a recall of 0.04, and an f1-score of 0.06, suggesting significant difficulty in accurately detecting and distinguishing deceased cases from other categories. The "Under Treatment" class, representing individuals still receiving care, yielded intermediate results (precision 0.41, recall 0.35, f1-score 0.37).

The relatively low macro-averaged metrics underscore class imbalance and variance in model effectiveness across the target categories, with the model showing a marked bias towards the majority class ("Recovered"). The overall accuracy of 44.6% confirms a limited generalization capability when evaluated on unseen data, likely reflecting both the inherent challenges of the task and the tendency of unpruned decision trees to overfit the training set. These findings highlight the importance of further optimizing the model, potentially through more robust regularization, resampling techniques to address class imbalance, or alternative algorithms, to achieve a more balanced and reliable predictive performance across all classes.

Table:10 Feature Importances for Model_4

Feature Name	Feature Importance
Age	0.165599
Weight	0.181521
Height	0.177072
Gender_Female	0.005095
Gender_Male	0.015310
Cancer_Type_Breast	0.008585
Cancer_Type_Colorectal	0.012152

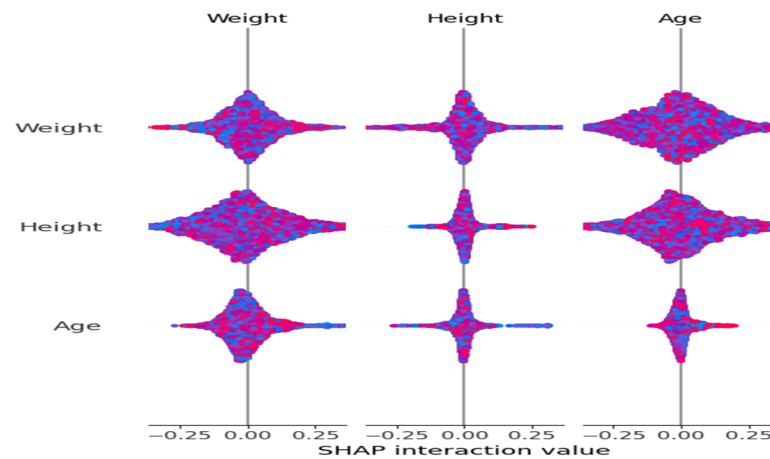
Source: Authors elaboration based on Python output

The feature importance analysis for Model_4, as presented above, reveals that biological attributes such as Weight (0.1815), Height (0.1771), and Age (0.1656) exhibit the highest relative contributions to the model's predictions. These variables collectively account for a significant proportion of the decision-making process in the tree model. Among categorical and socio-demographic characteristics, variables such as Emirate_Dubai (0.0523), Nationality_Expatriate (0.0206), and Emirate_Ras Al Khaimah (0.0171) show moderate influence, while most remaining features — including gender, specific emirate locations, and cancer type categories — contribute minimally to the predictive output.

This distribution of importance suggests that the model relies primarily on direct personal health indicators (age, weight, height) when differentiating between outcome classes. Contextual and socio-demographic variables supplement predictive value but to a lesser extent, emphasizing the relative primacy of physiological data in this application scenario. The negligible importance attributed to certain factors (e.g., Gender_Other) indicates limited, if any, model reliance on these categories, which could be due to class distribution or data sparsity for those features. These findings are essential for understanding the model's decision logic and for guiding further feature engineering or model refinement in future analyses.

5.4.2. SHAP Interaction Matrix in the Decision Tree

Before examining feature interactions, it is essential to assess the model's overall performance. The following evaluation metrics and confusion matrix provide insight into the accuracy and classification ability of the decision tree on the test set.

Fig.:5: Evaluation Metrics and Confusion Matrix for the Decision Tree Model on the Test Set

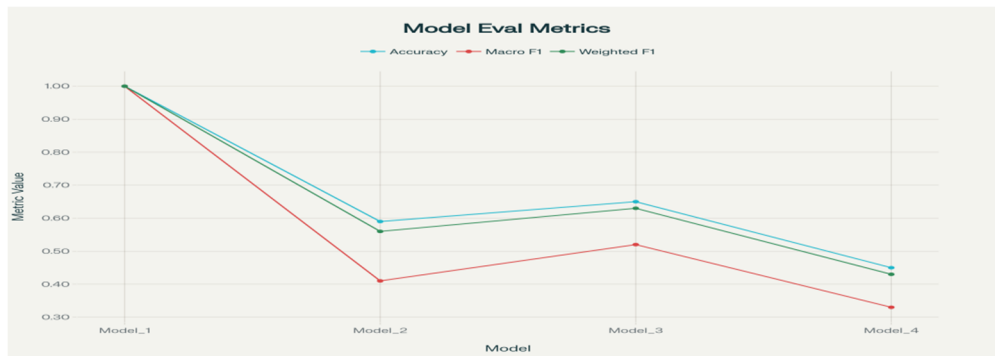
Source: Author elaboration based on Python output

The figure presents a matrix of SHAP interaction value plots for the features Weight, Height, and Age, elucidating both their main effects and pairwise interactions within the decision tree model. Each off-diagonal cell illustrates the distribution of SHAP interaction values for a specific feature pair, quantifying how the combination of those features contributes to model predictions beyond the sum of their individual effects. Violin-shaped distributions within these cells capture the range and density of interaction values; broader areas indicate more frequent or substantial interactions at corresponding values, while asymmetry may suggest directionality or varying interaction intensity.

Diagonal cells, by contrast, represent the main effects of each feature, reflecting their isolated impact on the prediction outcome. Distributions that are centered around zero indicate minimal interaction between the respective feature pair, whereas widely spread distributions denote stronger, more influential interactions. The analysis of symmetry and density patterns provides further insight into the consistency and directionality of these interactions across the dataset.

Collectively, this visualization demonstrates that Weight, Height, and Age exert significant and intertwined effects on the model's predictions. The prominence of these interaction patterns underscores the importance of considering both main effects and feature interactions when interpreting variable influence within clinical risk models and supports the role of SHAP interaction analysis in achieving interpretable and clinically relevant machine learning results.

Fig:6:Models evaluation metrics



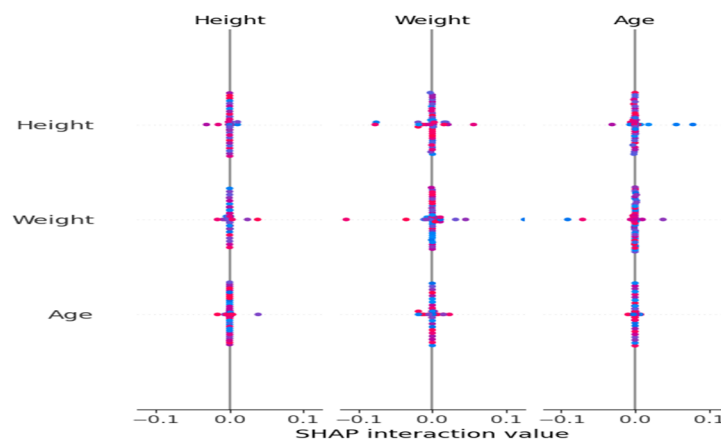
Source: Authors elaboration based on Python output

The model_4, which employs a realistic data split without pruning, provides the most transparent assessment of generalization ability. Its lower evaluation scores highlight the inherent challenge of predictive modeling on unseen data and the risks of relying exclusively on in-sample validation. Notably, across all models, biological variables such as age, weight, and height consistently emerge as the most significant predictors, emphasizing the importance of clinical and physiological characteristics in health outcome prediction models.

5.4.3. SHAP Interaction Evaluation: Evidence of Independent Feature Behavior

This section investigates how features interact within the decision tree model using SHAP interaction values. The focus is on understanding whether height, weight, and age jointly influence predictions or operate independently.

Fig:7: Pairwise SHAP Interaction Values for Height, Weight, and Age Features in the Predictive Model



Source: Author elaboration based on Python output

The figure displays a symmetric grid in which both the horizontal and vertical axes represent the same three features: height, weight, and age. Each cell in the matrix corresponds to a pairwise interaction between two of these features, while the dots indicate the SHAP interaction values for each observation in the dataset. The visual inspection reveals that these values are concentrated around zero with limited dispersion, reflecting the absence

of strong or nonlinear interaction effects among the variables. The plots show no clear patterns, trends, or notable outliers, indicating that the interactions between height, weight, and age do not contribute significantly to the model's outputs.

This consistently low and symmetric distribution of SHAP values suggests that each feature affects the target variable independently, while their joint interaction effects remain weak and negligible within the structure of the model.

6. Conclusion

Based on a large-scale sample of approximately **10,000 cancer patients from the Cancer Data of the UAE**, this study provides actionable insights for healthcare managers and policy makers seeking to integrate interpretable machine learning models into oncology decision-support systems. The size and diversity of the sample offer a realistic empirical foundation for assessing how predictive models can inform operational planning and strategic resource management within a national healthcare context.

The comparative evaluation of decision tree configurations demonstrates that unrestricted models despite achieving perfect in-sample accuracy are unsuitable for managerial deployment due to their lack of generalizability and interpretability. From a policy perspective, such models risk generating misleading signals that could distort resource allocation and prioritization decisions. In contrast, pruned decision tree models offer a more reliable balance between predictive performance and transparency, making them better aligned with the accountability and explainability requirements of healthcare governance.

The train–test evaluation further highlights that predictive tools intended for real-world deployment must be assessed under conditions that mirror operational uncertainty. The observed performance degradation on unseen patient data underscores the importance of validation frameworks that support evidence-based decision-making rather than retrospective performance reporting.

Across all models trained on the UAE cancer sample, anthropometric variables—particularly age, weight, and height—consistently emerged as dominant predictors of patient outcomes. For healthcare managers, this finding implies that readily available physiological indicators can serve as effective criteria for risk stratification, early intervention planning, and targeted allocation of clinical resources, while socio-demographic variables play a secondary, contextual role.

From a policy standpoint, the results suggest that interpretable machine learning models can support more equitable and efficient healthcare systems by identifying high-risk patient groups and informing preventive and therapeutic strategies. When embedded within hospital information systems or national cancer registries, such models can assist decision-makers in optimizing capacity planning, prioritizing vulnerable populations, and improving overall system resilience.

In conclusion, this study demonstrates that decision tree models—when appropriately pruned and rigorously validated—can function not only as predictive tools, but as **managerial instruments** that translate patient-level data into operational and policy-relevant insights. Future research should extend this framework through ensemble methods and class-balancing strategies to further enhance fairness, robustness, and decision utility in population-level cancer management.

7. List of references:

1. Banegas-Luna, A. J., Peña-García, J., Iftene, A., Guadagni, F., Ferroni, P., Scarpato, N., Zanzotto, F. M., Bueno-Crespo, A., & Pérez-Sánchez, H. (2021). When will the mist clear? On the Interpretability of Machine Learning for Medical Applications: a survey. *International Journal of Molecular Sciences*, 22(9), 4394. <https://doi.org/10.3390/ijms22094394>
2. Ghasemi, F., Farokhzad, R., Soleimani, M., & Taki, M. (2024). Explainable artificial intelligence in breast cancer detection and risk prediction: A systematic scoping review. *Cancer Innovation*, 4(1), e136. <https://doi.org/10.1002/cai2.136>
3. Li, Y., Zhao, X., Zhou, Y., Gong, L., & Peng, E. (2025). Decision tree model for predicting ovarian tumor malignancy based on clinical markers and preoperative circulating blood cells. *BMC Medical Informatics and Decision Making*. <https://doi.org/10.1186/s12911-025-02934-8>
4. Lundberg, S. M., Erion, G., Chen, H., DeGrave, A., Prutkin, J. M., Nair, B., Katz, R., Himmelfarb, J., Bansal, N., & Lee, S.-I. (2020). Explainable AI for Trees: From Local Explanations to Global Understanding. *arXiv preprint arXiv:1905.04610*. <https://arxiv.org/abs/1905.04610>
5. Wang, K. J., Makond, B., & Wang, K. M. (2010). An improved survivability prognosis of breast cancer by using sampling and feature selection technique to solve imbalanced patient classification data. *Journal of Clinical Medicine Research*, 2(1), 27-34. <https://www.researchgate.net/publication/258426244>
6. Green, J., Cairns, B. J., Casabonne, D., Wright, F. L., Reeves, G., Beral, V., & Million Women Study Collaborators. (2011). Height and cancer incidence in the Million Women Study: prospective cohort, and meta-analysis of prospective studies of height and total cancer risk. *The Lancet Oncology*, 12(8), 785–794. [https://doi.org/10.1016/S1470-2045\(11\)70193-2](https://doi.org/10.1016/S1470-2045(11)70193-2)
7. Lee, J. E., Li, H., Myung, S. K., & Kim, Y. (2019). Adult height in relation to risk of cancer in a cohort of 22,809,722 Korean adults. *Cancer Causes & Control*, 30(9), 925–933. <https://doi.org/10.1007/s10552-019-01213-8>
8. Bélanger, M. J., Park, V. Y., & Kim, D. (2022). Adult-attained height and colorectal cancer risk: a cohort study, systematic review and meta-analysis. *European Journal of Cancer*, 162, 100–111. <https://doi.org/10.1016/j.ejca.2021.10.022>
9. Seviiri, M., et al. (2016). Height, height-related SNPs, and risk of non-melanoma skin cancer. *International Journal of Cancer*, 139(8), 1794–1802.
10. Martín-González, J. A., Sánchez-Iñigo, L., & Martínez, J. A. (2022). Body shape phenotypes of multiple anthropometric traits and cancer risk: A multi-national cohort study. *International Journal of Cancer*, 150(8), 1371–1379. <https://doi.org/10.1002/ijc.33940>
11. Kim, Y., et al. (2024). A large cohort study of height and mammographic density in relation to breast cancer risk among Korean women.
12. Dreher, M., et al. (2022). The association between body height and cancer: a retrospective analysis of 784,192 outpatients in Germany.
13. Davies, N. M., Dickson, M., Davey Smith, G., van den Berg, G. J., & Windmeijer, F. (2021). Body size and composition and risk of site-specific cancers in the UK Biobank and large international consortia: a mendelian randomisation study. *PLoS Medicine*, 18(7), e1003706. <https://doi.org/10.1371/journal.pmed.1003706>