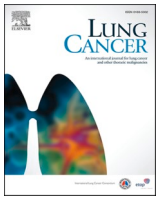
Lung Cancer 195 (2024) 107920

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan



Research Paper

Family history of cancer and lung cancer: Utility of big data and artificial intelligence for exploring the role of genetic risk

Virginia Calvo [a,*], Emetis Niazmand [b,c], Enric Carcereny [d], Delvys Rodriguez-Abreu [e], Manuel Cobo [f], Rafael López-Castro [g], María Guirado [h], Carlos Camps [i], Ana Laura Ortega [j], Reyes Bernabé [k], Bartomeu Massutí [l], Rosario Garcia-Campelo [m], Edel del Barco [n], José Luis González-Larriba [o], Joaquim Bosch-Barrera [p], Marta Martínez [a], María Torrente [a], María-Esther Vidal [b,c], Mariano Provencio [a,*]

[a] *Medical Oncology Department, Hospital Universitario Puerta de Hierro Majadahonda, Madrid, Spain*
[b] *TIB Leibniz Information Centre for Science and Technology, Hannover, Germany*
[c] *L3S Research, Hannover, Germany*
[d] *Medical Oncology Department, Catalan Institute of Oncology (ICO)-Badalona Badalona-Applied Research Group in Oncology (B-ARGO), Barcelona, Spain*
[e] *Medical Oncology Department, Hospital Insular de Gran Canaria, Las Palmas de Gran Canaria, Spain*
[f] *BIONAND, Unidad de Gestión Clínica Intercentros de Oncología Médica, Hospitales Universitarios Regional y Virgen de la Victoria, IBIMA, Málaga. Spain*
[g] *Medical Oncology Department, Hospital Clínico Universitario de Valladolid, Valladolid, Spain*
[h] *Medical Oncology Department, Hospital General Universitario de Elche, Elche, Spain*
[i] *Medical Oncology Departametnt, Hospital General Universitario de Valencia, Valencia, Spain*
[j] *Medical Oncology Department, Hospital Universitario de Jaen, Jaen, Spain*
[k] *Medical Oncology Department, Hospital Universitario Virgen del Rocío, Sevilla, Spain*
[l] *Medical Oncology Department, Hospital General Universitario Doctor Balmis, Alicante, Spain*
[m] *Medical Oncology Department, Hospital Universitario A Coruña, A Coruña, Spain*
[n] *Medical Oncology Department, Hospital Universitario de Salamanca-IBSAL, Salamanca, Spain*
[o] *Medical Oncology Department, Hospital Clínico San Carlos, Madrid, Spain*
[p] *Medical Oncology Department, Catalan Institute of Oncology, University Hospital Dr. Josep Trueta, Precision Oncology Group (OncoGIR-Pro), Girona Biomedical Research Institute (IDIBGI-CERCA), Girona, Spain*

ARTICLE INFO

Keywords:
Cancer registry
Family history
Lung cancer
Spain

ABSTRACT

Objectives: Lung Cancer (LC) is a multifactorial disease for which the role of genetic susceptibility has become increasingly relevant. Our aim was to use artificial intelligence (AI) to analyze differences between patients with LC based on family history of cancer (FHC).

Materials and methods: From August 2016 to June 2020 clinical information was obtained from Thoracic Tumors Registry (TTR), a nationwide database sponsored by the Spanish Lung Cancer Group. In addition to descriptive statistical analysis, an AI-assisted analysis was performed. The German Technical Information Library supported the merging of data from the electronic medical records and database of the TTR. The results of the AI-assisted analysis were reported using Knowledge Graph, Unified Schema and descriptive and predictive analyses.

Results: Analyses were performed in two phases: first, conventional statistical analysis including 11,684 patients of those 5,806 had FHC. Median overall survival (OS) for the global population was 23 months (CI 95 %:

Abbreviations: AEMPS, Spanish Agency for Medicines and Medical Devices; AI, artificial intelligence; ALK, anaplastic lymphoma kinase; BRAF, B-RAF proto-oncogene, serine/threonine kinase oncogene; CLARIFY, Cancer Long Survivor Artificial Intelligence Follow-up; DAG, directed acyclic graph; EGFR, epidermal growth factor receptor; EHR, Electronic Health Records; FCC, Familial Cancer Connectedness; FHC, family history of cancer; HER2, human epidermal growth factor receptor type 2; HUPHM, Hospital Universitario Puerta de Hierro-Majadahonda; IQR, interquartile range; KG, Knowledge Graph; KRAS, Kirsten rat sarcoma viral oncogene homolog; LC, lung cancer; NFCF, Normalized Familial Cancer Frequency; NFHC, no family history of cancer; NS, non-significant; NSCLC, non-small cell lung cancer; OR, Odds-Ratio; OWL, Web Ontology Language; PD-L1, programmed death-ligand 1; PS, performance status; ROS1, ROS proto-oncogene 1, receptor tyrosine kinase (ROS1); SCLC, small cell lung cancer; SD, standard deviation; SLCG, Spanish Lung Cancer Group; TIB, Technische Informationsbibliothek; TTR, Thoracic Tumors Registry.

* Corresponding authors at: Medical Oncology Department, Hospital Universitario Puerta de Hierro Majadahonda, Manuel de Falla 1, 28222 Majadahonda, Madrid, Spain.

E-mail addresses: vircalvo@hotmail.com (V. Calvo), mprovenciop@gmail.com (M. Provencio).

https://doi.org/10.1016/j.lungcan.2024.107920
Received 6 June 2024; Received in revised form 2 August 2024; Accepted 5 August 2024
Available online 9 August 2024

21.39–24.61) in patients with FHC versus 21 months (CI 95 %: 19.53–22.48) in patients without FHC (NFHC), p < 0.001. The second AI-assisted analysis included 5,788 patients of those 939 had FHC. 58.48 % of women with FHC had LC. 9.53 % of patients had an EGFR or HER2 mutation or ALK translocation and at least one relative with cancer. A family history of LC was associated with an increased risk of smoking-related LC. Non-smokers with a family history of LC were more likely to have an EGFR mutation in NSCLC. In Bayesian network analysis, 55 % of patients with a family history of LC and never-smokers had an EGFR mutation.
Conclusion: In our population, the incidence of LC in patients with a FHC is higher in women and younger patients. FHC is a risk factor and predictor of LC development, especially in people ≤ 50 years. These results were confirmed by conventional statistics and AI-assisted analysis.

1. Introduction

Lung cancer (LC) is one of the most important public health problems in the world due to its high incidence and mortality. According to Siegel et al, it is estimated that in 2024, it will be the second most common cancer in US men and women (11 % and 12 % of all estimated new cases respectively). Additionally, LC will be the deadliest cancer in both sexes, being responsible of 20–21 % of deaths [1]. In Spain, LC is the third most common cancer in men and women, was the leading cause of death in 2022 and it is estimated that the number of new cases in 2024 will reach 32,768 [2].

LC is classified into two types: small cell lung cancer (SCLC) accounts for about 15 % and non-small cell lung cancer (NSCLC) for 85 % [3]. It is a multifactorial disease, mainly associated with extrinsic or modifiable environmental factors. These include tobacco, environmental tobacco smoke, radon, occupation and certain leisure activities, environmental pollution, and diet [4]. There are also other intrinsic or non-modifiable factors such as age, gender, family, and personal history (previous respiratory diseases, and genetic susceptibility) [5]. Smoking is the main risk factor for LC. Tobacco is estimated to be responsible for around 79 % of LC cases in men and 47 % in women [6], being the most important modifiable risk factor. Smoking habits differ in different geographics regions and based on this the proportion of lung cancer attributable to tobacco smoking varies between countries, from > 80 % for men and women in the US [7] and UK [8] to 57.5 % for men and 13 % for women in China [9]. Passive smokers and those exposed to second-hand smoke also have a higher risk of LC [10]. Residential radon is considered the second most important risk factor for LC in smokers and the first in never-smokers [11].

Individuals may differ in their genetic susceptibility to environmental factors, including tobacco. This may also affect incidence [12]. In assessing genetic susceptibility, family history of cancer (FHC) is used as a surrogate for genetic risk [13,14].

The objective of our study is to explore the role of genetic risk in our population with FHC, for this purpose, we analyzed characteristics of patients with LC according to FHC in the Spanish Thoracic Tumors Registry (TTR). Given the emerging role of artificial intelligence (AI) in the collection and interpretation of clinical data, we implemented this framework to perform complementary analysis of the information collected.

2. Material and methods

2.1. Study design

From August 2016 to June 2020, clinical data was obtained from TTR. The TTR is an observational, prospective, and multicenter study, that includes patients with LC and other types of thoracic tumors, which was developed by the Spanish Lung Cancer Group (SLCG). It was approved by the Spanish Agency for Medicines and Medical Devices (AEMPS) in 2016 and is registered in the ClinicalTrials.gov database (NCT02941458). The TTR is an active database with continuous patient enrolment, for this reason, the analysis performed in two periods. Sociodemographic, epidemiological, clinical, molecular and treatment outcome variables were collected.

For the artificial intelligence-assisted analysis, we had the support from the collaborators of the Cancer Long Survivor Artificial Intelligence Follow-up (CLARIFY) of the European Union Horizon 2020 Research and Innovation Program. The research was conducted according to the Declaration of Helsinki. Protocol approval was obtained from the institutional ethical committee of the Hospital Universitario Puerta de Hierro-Majadahonda (HUPHM).

Informed consent was obtained from all the patients included in both analyses: statistical analyses and analyses using artificial intelligence (AI) assistance.

2.2. Patient population and inclusion criteria

Inclusion criteria were patients with lung cancer in active treatment or without treatment (palliative care only), with no gender or age restrictions. Patients with other tumor types and healthy volunteers were not included. Data collected included age, sex, race, family history of cancer (FHC), smoking status, Eastern Cooperative Oncology Group (ECOG) performance status, disease stage, histology and molecular markers tested.

FHC refers to the presence of cancer in one or more family members. It may include information on the type of cancer, the number of affected relatives, and the degree of relationship (first-, second-, and third-degree relatives).

Smoking status was defined as current smoker (an adult who has smoked 100 cigarettes in his or her lifetime and who currently smokes cigarettes), former smoker (an adult who has smoked at least 100 cigarettes in his or her lifetime but who has quit smoking more than one year ago) and never smoker (an adult who has never smoked, or who has smoked fewer than 100 cigarettes in his or her lifetime). Information on passive smoking, defined as the breathing in cigarette smoke from people smoking nearby, was also collected.

2.3. Statistical analysis

Conventional descriptive statistics analysis was performed. Quantitative data were summarized as mean, standard deviation (SD), median and interquartile range (IQR). Qualitative variables were reported as frequencies and percentages. The significance level was $\alpha = 0.05$. Univariate analysis was used to analyze the main objective. Multivariate analysis using Cox regression was used to determine associations. Overall survival (OS) was calculated using the Kaplan-Meier method.

2.4. Analysis using AI

Analysis supported by AI was represented using the Knowledge Graph (KG) and Unified Schema as well as descriptive and predictive analysis.

Technische Informationsbibliothek (TIB) in Germany merged data from Electronic Health Records (EHR) and database from HUPHM and the TTR of the SLCG.

2.4.1. The TTR Knowledge graph and unified Schema

TTR KG is structured data that represents factual knowledge about patients as entities and their relationships using a graph data model. Metadata of TTR dataset is part of the TTR KG, as well as taxonomies of entities, relationships, and classes. Ontologies and controlled are used to describe the meaning of relationships and to annotate entities in a consistent way in the KG. SDM-RDFizer is used to generate the TTR KG. SDM-RDFizer is a mapping rules interpreter that allows the transformation of data in different formats into KG [15]. Explaining the process of creating KG is beyond the scope of this paper. However, details can be found in Vidal et al [16].

TTR unified schema is used to describe the meaning of data collected and integrated into TTR KG. The integrated view of TTR unified schema and LC data in TTR KG provides relevant data to support the discovery of novel patterns that can enhance the understanding of patient conditions.

2.4.2. Descriptive analysis using AI

2.4.2.1. Normalized familial cancer frequency (NFCF). Characteristics of patients with LC are presented in terms of frequency. For each patient, it quantifies the normalized frequency of family members with cancer. A patient with a NFCF of 1.0 has the highest number of family members with cancer, while a of 0.0 indicates that the patient has no relatives with cancer.

2.4.2.2. Familial cancer connectedness (FCC). Is a measure to quantify the co-occurrence between the connections of patients with LC based on the values of their properties represented in TTR KG. FCC aims to provide a normalized ratio of the co-occurrence of the connections (or properties and their values) in which TTR patients participate.

2.4.2.3. Odds Ratio Analysis. Derived using logistic regression. Value less than 1.0 suggests no association, while the higher the value, the greater the likelihood of an association.

2.4.3. Predictive analysis using AI

2.4.3.1. Bayesian Networks. It was trained to detect the probability of mutations in TTR patients based on their characteristics. In addition, three families of structure learning methods are considered: (1) the constraint-based method, which discovers causal structure by conditional independent testing; (2) the score-based search method, which uses a search method to extend the nodes of the directed acyclic graph (DAG), and a score function to measure the goodness of the DAG; and (3) the hybrid method, which combines constraint-based and score-based methods. Causal structures were annotated by clinical collaborator experts. For the constraint-based method, the following algorithms are considered: Inferred Causation (IC) algorithm [17] which implemented in WEKA [18]. For the score-based method, the K2 score (a type of Bayesian score) is used as the score function. The structure searching algorithms include: the Tree Augmented Naive Bayes (TAN), the K2 searching algorithm, the Markov Chain Monte Carlo Model (MCMC) based searching algorithm, the Simulated Annealing (SA) based searching algorithm. The MCMC based algorithm implemented in the BNT toolbox [19] is utilized, as well as a score-based method implemented in WEKA. For the hybrid method, the Causal Minimum Message Length (CaMML) [20] algorithm, is utilized. Four different criteria are considered: (a) the Area Under the Curve (AUC), (b) Accuracy, (c) Precision, (d) Recall. All criteria are measured as the average performance under 10-fold cross validation. The Explainable Boosting Machine (EBM) [21] and SHAP (SHapley Additive exPlanation) [22] provide the explanation of the important features for the predictive tasks. EBM is a generalized additive model (GAM) with the form: g(E[y]) = β + Pfj(xj) + Pfij(xi,xj), where "g" is a link function adapting GAM for classification, and each "f" is a feature function that would be learned to measure the contribution of features to the final prediction. On the other hand, SHAP value resorts to cooperative game theory to measure feature importance. The Shapley value of feature "j" is computed by a function: $\phi_j = \sum_{S \subseteq N \setminus \{j\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f_x(S \cup \{j\}) - f_x(S$, where "N" is the set of all features, "S" includes all possible set of combination of other features except "j", and fx(S) = E[y(x)|xS] is the marginal expectation to "*y*" consider a set of features "S".

3. Results

3.1. Results of the statistical analysis

3.1.1. Baseline characteristics of patients

For the first statistical analysis, 12,351 patients diagnosed with lung cancer were included between August 2016 and June 2020 from 182 hospitals across Spain. 667 patients had unknown FHC and were excluded. Of the remaining 11,684 patients, 5,806 had a history of cancer and 5,878 did not. Baseline characteristics of patients by FHC in Table 1.

The median age at diagnosis in the global population was 64.31 years (range: 30–98). According to FHC and no family history of cancer (NFHC), the median age was 63.74 (standard deviation (SD) 9.98; 95 % CI: 63.48–63.99; $p < 0.001$) and 64.87 (SD 10.15; 95 % CI: 64.60–65.12; $p < 0.001$, respectively.

The percentage of patients younger than 50 years was 53.16 % in the FHC group compared to 46.84 % in the NFHC group ($\chi^2 = 4.93$; $p = 0.026$).

The percentage of women with LC was also higher in the FHC group: 28.92 vs 24.35 % ($p < 0.001$). There were more never smokers (12.30 %) and former smokers (47.41 %) in the NFHC group.

Differences by stage of disease, histological subtype or biomarker (Table 2).

3.1.2. Overall survival

Median overall survival (OS) for the global population was 23 months (CI 95 %: 21.39–24.61) in patients with FHC versus 21 months

Table 1
Baseline characteristics of patients by FHC.

Characteristics	Family history of cancer	No family history of cancer	χ^2
	N=5,806	N=5,878	*p* value
Sex			
Male	4,127 (71.08 %)	4,447 (75.65 %)	$\chi^2 = 31.34$
Female	1,679 (28.92 %)	1,431 (24.35 %)	$p < 0.001$
Race			
Caucasian	5,748 (99.0 %)	5,686 (96.73 %)	$\chi^2 = 79.23$
Other	58 (1.0 %)	192 (3.27 %)	$p < 0.001$
Smoking status			
Never smoker	667 (11.49 %)	723 (12.30 %)	
Former smoker	2,656 (45.75 %)	2,787 (47.41 %)	
Current smoker	2,454 (42.47 %)	2,310 (39.30 %)	$\chi^2 = 20.78$
Unknown	29 (0.29 %)	58 (0.99 %)	$p < 0.001$
Passive smoking			
Yes	1,084 (18.8%)	1,047 (17.97%)	
No	810 (14.04 %)	1,043 (17.9 %)	
Unknown	3,912 (67.16 %)	3,788 (64.13 %)	
ECOG PS			
PS 0–1	5,062 (87.19 %)	4,904 (83.43 %)	
PS≥2	736 (12.68 %)	969 (16.49 %)	
Unknown	8 (0.13 %)	5 (0.08 %)	*NS*

NS: non-significant; PS: performance status.

Table 2
Baseline characteristics of patients by FHC.

Characteristics	Family history of cancer	No family history of cancer	χ^2
	N = 5,806	N = 5,878	*p* value
Stage (NSCLC)			
I	483 (8.3%)	467 (7.9%)	
II	419 (7.2%)	414 (7%)	
III	1414 (24.3%)	1377 (23.4%)	
IV	2706 (46.6%)	2877 (48.95%)	
Limited stage (SCLC)	292 (5%)	232 (3.94%)	
Extensive stage (SCLC)	447 (7.7%)	461 (7.8%)	χ^2=24.24
Unknown	45 (0.9%)	50 (1.01%)	*p=0.187*
Histology			
Adenocarcinoma	3,098 (53.37%)	3,109 (52.89%)	
Adenosquamous carcinoma	52 (0.9%)	106 (1.8%)	
Squamous cell carcinoma	1,369 (23.58%)	1,442 (24.53%)	
Large cell carcinoma	148 (2.6%)	163 (2.8%)	
Sarcomatoid carcinoma	23 (0.4%)	13 (0.22%)	
NOS/undifferentiated	133 (2.3%)	126 (2.14%)	
Carcinoid tumors	33 (0.57%)	24 (0.41%)	
Small cell lung carcinoma	692 (11.77%)	734 (12.64%)	χ^2=32.45
Others	258 (4.51%)	161 (2.57%)	*p*=0.002
Biomarkers			
EGFR mutated	483 (17.44%)	382 (13.38%)	
ALK translocated	126 (5.54%)	155 (6.76%)	
ROS1 translocated	35 (3.29%)	30 (2.53%)	
KRAS mutated	53 (26.37%)	106 (29.12%)	
BRAF mutated	15 (3.96%)	20 (4.07%)	χ^2=87.91
PD-L1 positive	953 (54.15%)	1000 (52.66%)	*p*=0.281

NSCLC: non-small cell lung cancer; SCLC: small cell lung cancer.

(CI 95 %: 19.53–22.48) in patients NFHC, p < 0.001 (Fig. 1 supplementary).

For the multivariate analysis for survival (performed by Cox regression) according to histological type (NSCLC, SCLC) and tumor stage: in SCLC, the increase in tumor stage presented HR 2.958, p = 0.001 and for the presence of FHC HR 0.971, p = 0.821. On the other hand, in NSCLC, the increase in tumor stage presented HR 1.427, p = 0.001 and for the presence of FHC, HR 1.081, p = 0.123.

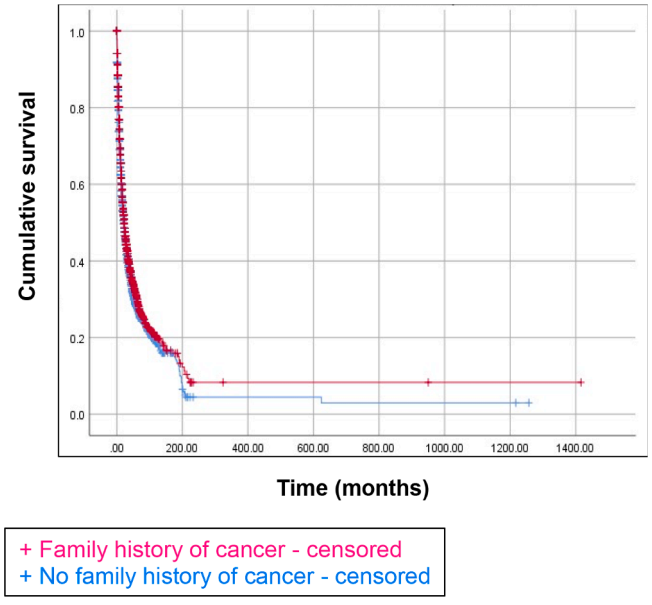


Fig. 1. Overall survival.

3.2. Results of analysis using AI

For the second analyses using AI assistance, we included 15,440 patients. 9,652 patients were excluded for an unknown FHC and reported no FHC. Of the 5,788 patients included, 939 had history of cancer and 4,849 did not.

3.2.1. TTR Knowledge graph and unified schema

Part of the TTR KG is shown in Fig. 2. Each generic concept of a type or category is defined as a class in OWL (Web Ontology Language) and is shown in Table 3 supplementary. The current version of TTR unified schema consists of 13 classes, 8 object properties, and 15 data type properties (Fig. 3).

In TTR KG, 7.51 % of patients have ALK translocation. Among these patients, 71.86 % are older than 50 years, while 29.84 % are younger than or equal to 50 years; 15.73 % are never-smokers and younger than or equal to 50 years. Also, 9.53 % of patients with ALK translocation or EGFR or HER2 mutations have at least one relative with cancer; however, 12.76 % of patients have at least one of these molecular biomarkers and one relative with breast, colorectal, ovarian, or bladder cancer.

3.2.2. Descriptive analysis using AI

3.2.2.1. Normalized familial cancer frequency (NFCF). 74.03 % and 25.07 % of population were male and female respectively. 37.88 % had at least one relative with cancer, 38.44 % had NFHC, and the rest were unknown. Regarding molecular biomarkers, 2 % had ALK translocation, 8.15 % EGFR mutation, and 0.05 % had both mutations. Finally, the distribution of patients by smoking habits was 41.01 % current smokers, 45.88 % former smokers, 11.62 % never smokers, and 1.25 % of patients had unknown smoking habits.

Figs. 4 and 5 in the Supplementary Appendix show the age distribution of patients and the correlation between the number of relatives with any type of cancer and the number of biomarkers.

Fig. 6 *and* Table 4 *(*supplementary*)* show the frequencies according: age, gender, smoking history, and molecular biomarkers. First, EGFR mutation and ALK translocation are more frequent in women. In patients with at least one relative with any type of cancer, the frequency of ALK translocation is higher than EGFR mutation. ALK translocation is more common in young female never-smokers, whereas EGFR mutation is more common in older female never-smokers. Additionally, EGFR mutation is more frequent in young male never-smokers and older male former smokers.

Fig. 7 *(*supplementary*)* shows the normalized frequency of TTR patients positive for molecular biomarkers by sex, age, and smoking habit.

The incidence of cancers by family degree is also shown in Table 5 *(*supplementary*)*. The relatives of the TTR lung cancer patients are first-degree relatives: father, mother, son and daughter; second-degree relatives: brother, sister, grandparents and third-degree relatives: uncles and aunts and great-grandparents.

Breast, lung, colorectal, head and neck, esophago-gastric and prostate cancers are the most common (major cancers). It was observed that cancers are more common in first-degree relatives. As expected, the major cancers were also more common in relatives.

LC is the most common cancer in close relatives of patients with LC (31.06 %), followed by breast (18.30 %) and colorectal (18.01 %). Fathers (9.96 %) and brothers (7.35 %) with LC are the most common relatives of male LC patients. Sisters (5.15 %) and mothers (4.58 %) with breast cancer are the most common relatives of male LC patients.

In female patients, fathers (3.02 %) and brothers (2.09 %) with LC are also more common, but sisters (2.68 %), aunts (1.96 %), and mothers (1.89 %) are the most common types of relatives with breast cancer. Genetic susceptibility to LC is higher in women (58.48 %) than in men (41.52 %) when considering molecular biomarkers ALK, EGFR, and HER2.

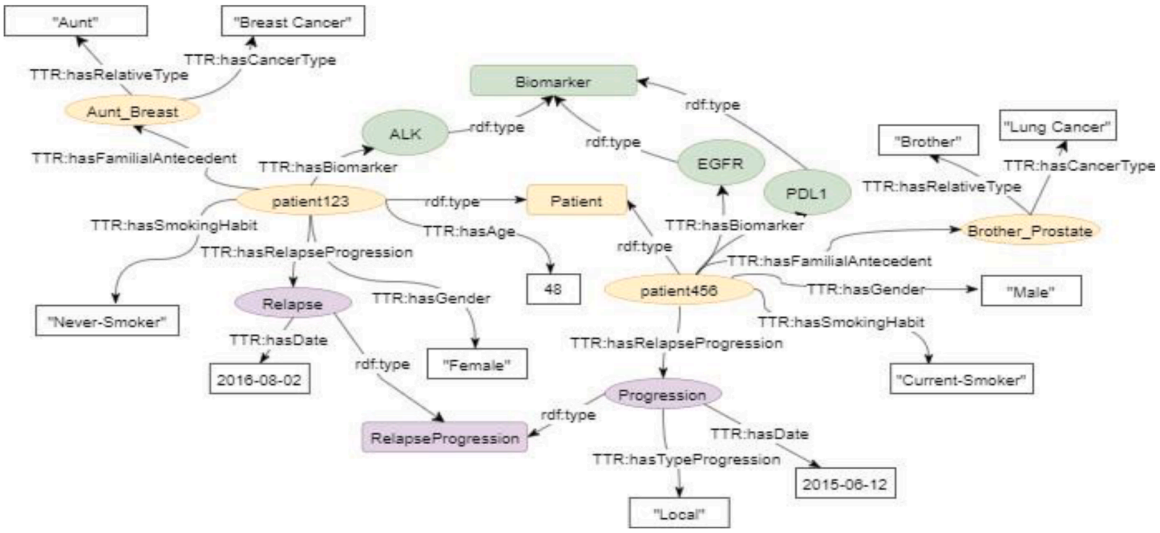


Fig. 2. Part of TTR KG.

Table 3
List of classes shown in Fig. 3 and their corresponding descriptions.

Class Name	Description of Classes
LCPatient	Lung cancer patients from the hospitals in Spain
Age	Patients categorized by age as younger $\leq$ 50 years and older $>$ 50 years
Gender	Shows both genders as male and female
Biomarker	A measurable indicator of a biological state or condition such as ALK, EGFR, KRAS, PDL1, ROS1, BRAF, HER2
SmokingHabits	Represents the patients' smoking habits as never smoker, former smoker, and current smoker
FamilialAntecedent	Represents the type of direct blood relatives of the patient who have suffered from any type of cancer
FamilyCancerType	Represents cancers suffered by relatives of the patient

3.2.2.2. Familial cancer connectedness (FCC). The population Pr of the TTR KG, consisting of female patients, aged 50 or less, and non-smokers. It has 116 LC patients; 57 with at least one relative with cancer. Of those, 12 have an ALK translocation and 6 of these have female second-degree relatives with breast cancer. The FCC measure captures the degree of connectedness of these 6 patients who share an aunt with breast cancer and ALK translocation.

FCC values calculated, in our population, with respect to gender, smoking habit and age are shown in Fig. 8 and Fig. 9 *(*supplementary*)*.

According to results shown in Fig. 9 LC women with ALK translocation who are young and never smokers often have relatives with colorectal, breast, uterine cervical, and skin cancers no melanoma. However, older female and never-smokers tend to have relatives with cervical, breast, head and neck, leukemia, colorectal, and lung cancers. On the other hand, young LC male with ALK translocation who have

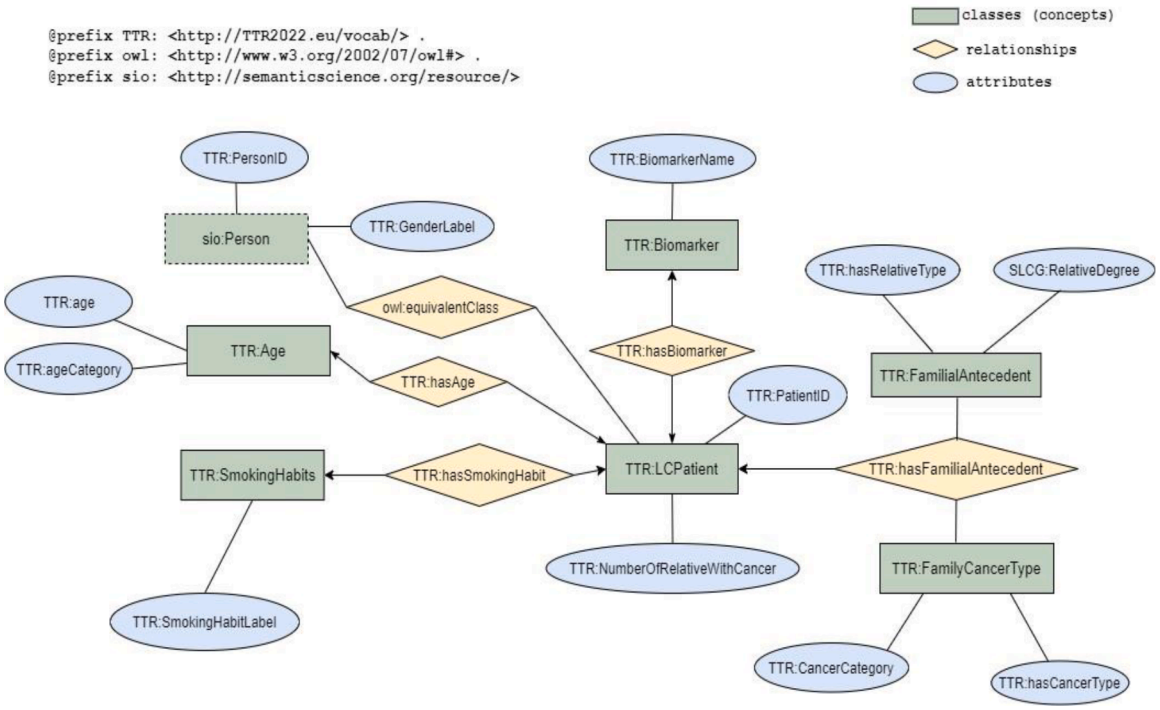
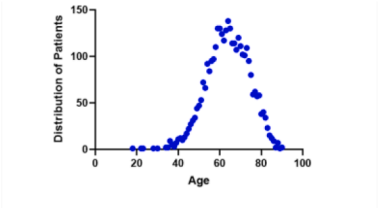
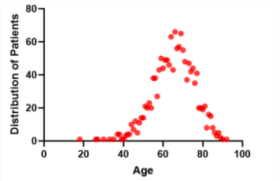


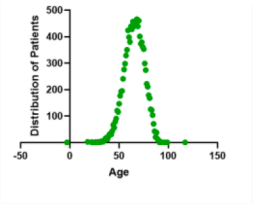
Fig. 3. Part of TTR ontological model (unified schema) covering the characteristics of lung cancer patients.



a) Patients who have at least one relative with one of the main cancers



b) Patient who have relatives with any other cancer, but not relatives with one of the main cancers



c) Patients who have no relatives with any type of cancers

Fig. 4. Distribution of patients by age.

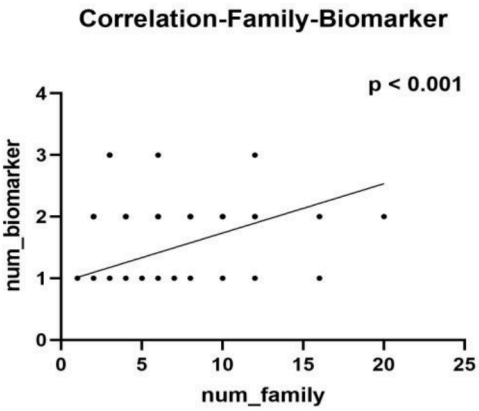


Fig. 5. Correlation between the number of relatives with any type of cancer and the number of biomarkers.

never smoked often have relatives with lymphoma, while older patients have relatives with lung and breast cancer.

Similarly, FCC values are shown for patients with EGFR mutation. Young female non-smokers often have relatives with lung, breast, and prostate cancer. However, when these patients are older, colorectal, and head and neck cancers are also observed. Finally, male with EGFR mutation, young never-smokers often have relatives with head and neck cancer, while lung, breast, colorectal, and prostate cancers are common in older never-smokers.

3.2.2.3. Odds-Ratio (OR) Analysis. As observed in Table 6 (supplementary), the highest OR (7.2042) was observed in young, never-smoker female, suggesting that these patients are more likely to have ALK translocation or EGFR mutation.

3.2.3. Predictive analysis

3.2.3.1. Bayesian Networks. Figs. 10 and 11 show that smoking history, gender, and stage are the 3 more relevant characteristics in both explanatory models (EBM and SHAP).

Furthermore, the local EBM interpretation showed that never smoking, female gender, and stage IA, accurately described patient with ALK translocation or EGFR mutation. On the other hand, local SHAP shows that never-smoking is the most relevant feature, also, the gender of affected relatives (male and female), LC stage (IA), and familial degree.

If the patient is a woman never-smoker, the probability of having ALK translocation or EGFR or HER2 mutations is 42 % versus 17 % of those with other alterations. If the patient is a man never-smoker, the probability of having ALK translocation or EGFR or HER2 mutations is 35 % versus 20 %.

With a family history of LC, 35 % and 40 % of patients are current and former smokers, respectively, while among never-smokers, 55 % have an EGFR mutation.

Considering LC in relatives of female patients, the probability of having only male relatives with LC is 77 %, while 16 % have only female relatives; 53 % and 34 % have only first- and second-degree relatives, respectively.

4. Discussion

Several studies have reported associations between a FHC and prognosis in LC. Lin H et al conducted a study of Chinese non-smoker patients and showed that familial aggregation of cancer or LC was associated with an increased risk of LC [23]. In another Chinese study of 4,491 patients with stage I-IV NSCLC, frequency of being female, never smoking, younger age, and adenocarcinoma histology was higher in patients with a FHC and was associated with a better prognosis in both early and advanced stage. A Spanish epidemiological study focusing on female with LC suggested significant associations between a FHC or LC and a favorable prognosis [24]. In this study, WORLD07, the authors found that 43 % of patients had a FHC, with a first-degree relative being the most common category of family member affected in just over half of patients. In our study, we observed a higher incidence of cancer in first-degree relatives. The median OS for the global population in WORLD07 was 24.0 (95 % CI: 22.1–26.0) months. The median OS was 23.0 (95 % CI: 20.7–25.3) months in patients with NFHC versus 25.3 (95 % CI: 21.9–28.8) months in patients with FHC (long-rank $p = 0.029$). These data were very similar to our data where median OS was 21 months (95 % CI: 19.53–22.48) months in patients with NFHC versus 23.0 (95 % CI: 21.39–24.61) months in patients with FHC, $p < 0.001$. Li N et al also reported that patients with NSCLC and a family history of LC had improved survival, but only in early-stage disease [25]. In contrast, Ganti AK et al reported that median survival was lower in patients with a

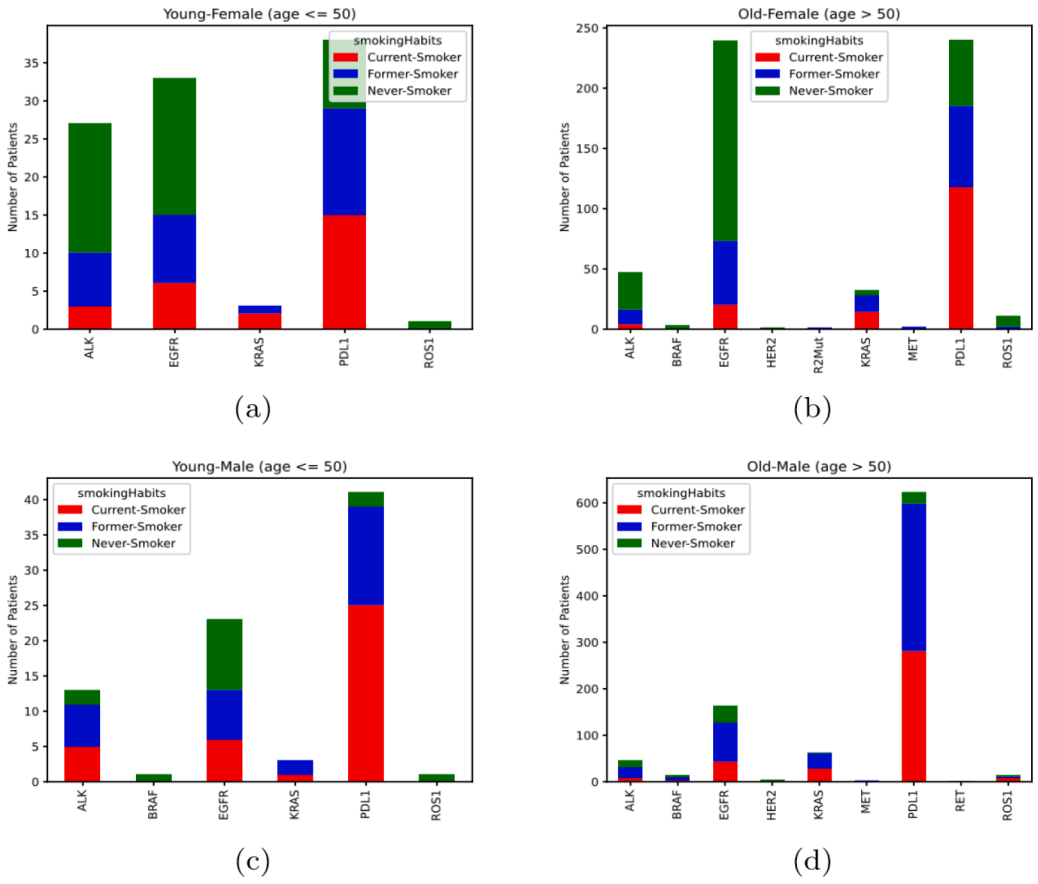


Fig. 6. Frequency of number of patients with at least one relative with any type of cancer by biomarker and patient smoking habit.

Table 4
Frequency of patients regarding their EGFR/ALK biomarkers, gender, family history of cancer and smoking habits.

		EGFR		ALK	
		Female	Male	Female	Male
		18.87 %	4.45 %	4.14 %	1.2 %
Patients with at least one relative with any type of cancer		39.00 %	38.00 %	46.95 %	43.00 %
	Never smoker	66.78 %	25.39 %	62.00 %	27.42 %
	Former smoker	21.64 %	49.22 %	25.97 %	51.00 %
	Current smoker	9.49 %	24.87 %	10.39 %	20.97 %

family history of LC than in those without (52 vs 58 months) [26].

In our study, conventional descriptive analysis and complementary analysis supported by AI, allowed us to identify patterns based on frequency, co-occurrence of variables, and causal associations. Kanwal M [27] observed that genetic susceptibility to LC is more likely in women than in men, and that breast, colorectal, bladder and ovarian cancers are often associated with an inherited predisposition. In our study, we confirmed these findings, the genetic susceptibility for LC is higher in women (58.48 %) than in men (41.52 %), whenever the molecular biomarkers: ALK, EGFR and HER2, but when PD-L1 is included, the distribution changes, 76.43 % are men and 46.95 % women. 9.53 % of the patients have ALK translocation or EGFR or HER2 mutation and have at least one relative with cancer. However, 12.76 % of patients have ALK translocation or EGFR or HER2 mutation, and have at least one relative with breast, colorectal, ovarian, or bladder cancer. Kanwal M et al also describe that breast, colorectal, bladder, and ovarian cancers are often associated with a hereditary predisposition in patients with LC. He Y et al investigate the association between FHC and EGFR mutation status in NSCLC population. In 538 patients with known EGFR mutation observed that those with FHC, especially of LC, might have a significantly higher incidence of EGFR mutation [28].

Evidence from the literature suggests that a family history of lung cancer is a predictor of an increased risk of developing the disease, particularly in people under the age of 50 [29,30]. In the TTR KG, 12.05 % of patients under 50 years have two or more affected relatives. At a cut-off age of 60 years, 41.42 % of patients meet this criterion.

Similarly, among never smokers, the risk of lung cancer has been shown to be increased in those with a family history of lung cancer [31]. We know that the incidence of familial LC is associated with an increased risk of LC due to tobacco. The association between an individuaĺs smoking habits and those of their parents or siblings is well documented [32]. Smokers with a family history of LC in a first-degree relative are much more likely to develop LC. Some studies have investigated the contribution of genetic factors to the risk of lung cancer; Jonsson S et al investigated this in the Icelandic population [33]. Their results confirmed the relevance of genetic predisposition in the development of lung cancer, especially in patients with early-stage disease. On the other hand, researchers from the International Lung Cancer Consortium evaluated nearly 50,000 LC cases and controls and their family histories [34]. Smoking is a common risk factor in other familial studies of LC, and it is usually not possible to make an appropriate adjustment for smoking. It is a limitation of our study that we do not have information on the dose and duration of the smoking habit, as well as on other environmental factors such as radon gas. This would have allowed us to reach better conclusions on the impact of this risk factor. Even if we do not have this data, nor do we have data on the smoking habits of the relatives with cancer, or information on other patients who

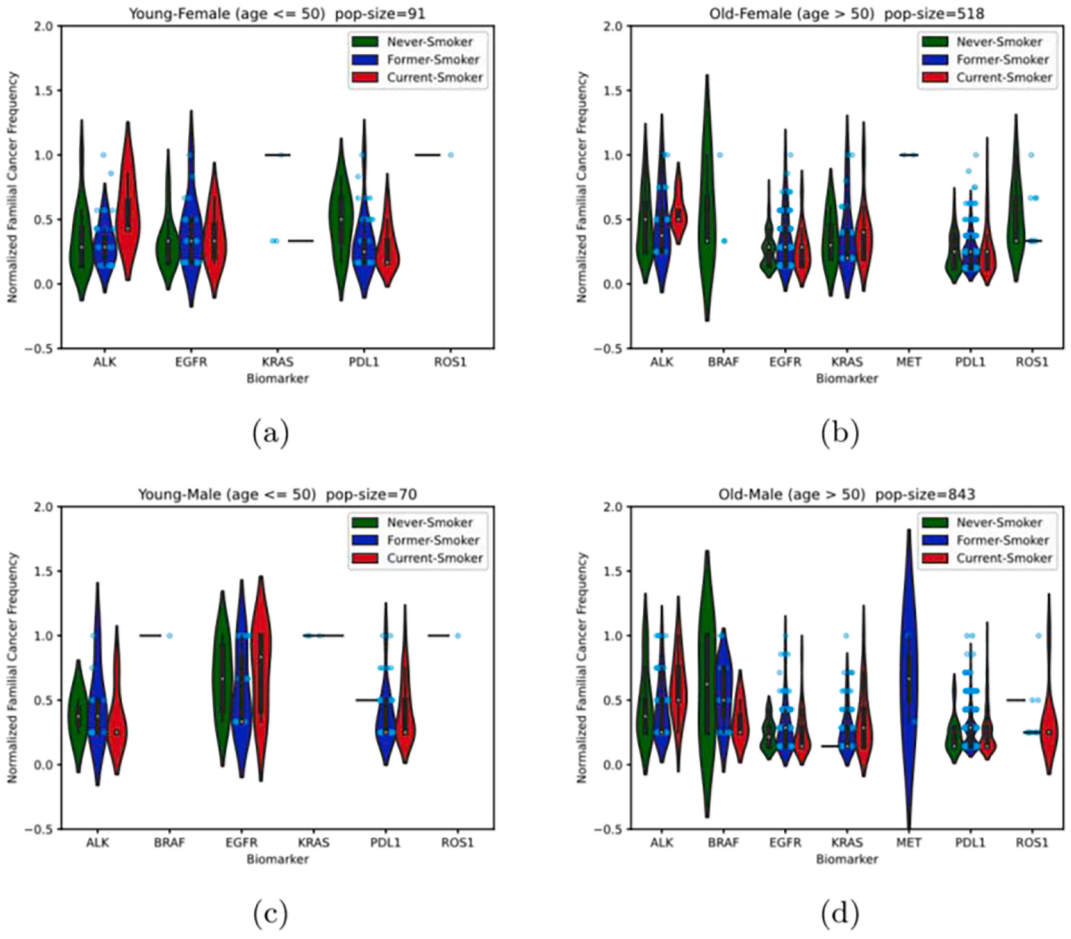


Fig. 7. Normalized frequency of TTR patients positive for molecular biomarkers by sex and smoking habits.

Table 5
Incidence of cancers by family degree.

Familial Degree	Cancer Type Category	Frequency
First-degree	Major Cancers	54.03 %
	Other Cancers	43.20 %
Second-degree	Major Cancers	17.30 %
	Other Cancers	10.90 %
Third-degree	Major Cancers	4.14 %
	Other Cancers	2.66 %
Unknown degree	Major Cancers	1.56 %
	Other Cancers	0.93 %

do not have LC, Bayesian network inference reveals related patterns. It can be inferred that 75 % of patients who have relatives with LC are current or former smokers. This observation may conceal the association between the smoking habits of relatives with LC. Kanwal M [27] observe that familial LC is associated with an increased risk of LC due to tobacco. We observed that with a family history of LC, 35 % and 40 % of patients are current and former smokers, respectively. In contrast, patients with a history of LC who are never smokers are more likely to have an EGFR mutation. Bayesian network inference confirmed in our study that 55 % of patients with a history of LC and never smokers have an EGFR mutation.

In addition, gender appears to be a relevant characteristic in patients with LC and relatives with LC. Isla D et al in the WORLD07 project showed that among Spanish women with LC, those with FHC or family history of LC were more likely to be current smokers than those without [24]. The Bayesian network did not confirm that gender is a relevant characteristic in patients with LC and relatives with LC. Looking at LC in relatives of female patients, the probability of having only male relatives with LC is 77 %, while 16 % have only female relatives. Furthermore, in this population, 53 % and 34 % had first- and second-degree relatives, respectively. The number of relatives with cancer was observed to be an important characteristic.

In relation to other characteristics such as stage of disease at diagnosis, Matakidou A [29] observed that people with a FHC were more likely to be diagnosed with early-stage disease.

Some limitations of this study should be noted. Firstly, the retrospective nature of the study and the performance of the analysis in two phases, considering that the TTR is an active database with continuous patient enrolment, the number of patients for each analysis was different. Secondly, the analysis using AI highlights some of the limitations of this technology, so the number of patients analyzed varied according to the data requested. Thirdly, as we have already mentioned, another limitation is the lack of knowledge of some data that could affect some of the results, such as data on smoking habits; we do not have information on dose and duration of smoking habits, as well as other environmental factors such as radon gas, occupation and certain free time activities, environmental pollution and diet, among others.

Despite these limitations, our study has strengths, such as the large number of patients included from different regions of Spain, the large amount of epidemiological, clinical and molecular data analyzed, and the promising role of AI in the collection and interpretation of clinical data.

These studies are very important because we need to identify subpopulations for genetic testing, and we need to improve the selection of candidates for lung cancer screening. There are already studies proposing new screening criteria to identify these populations with a high

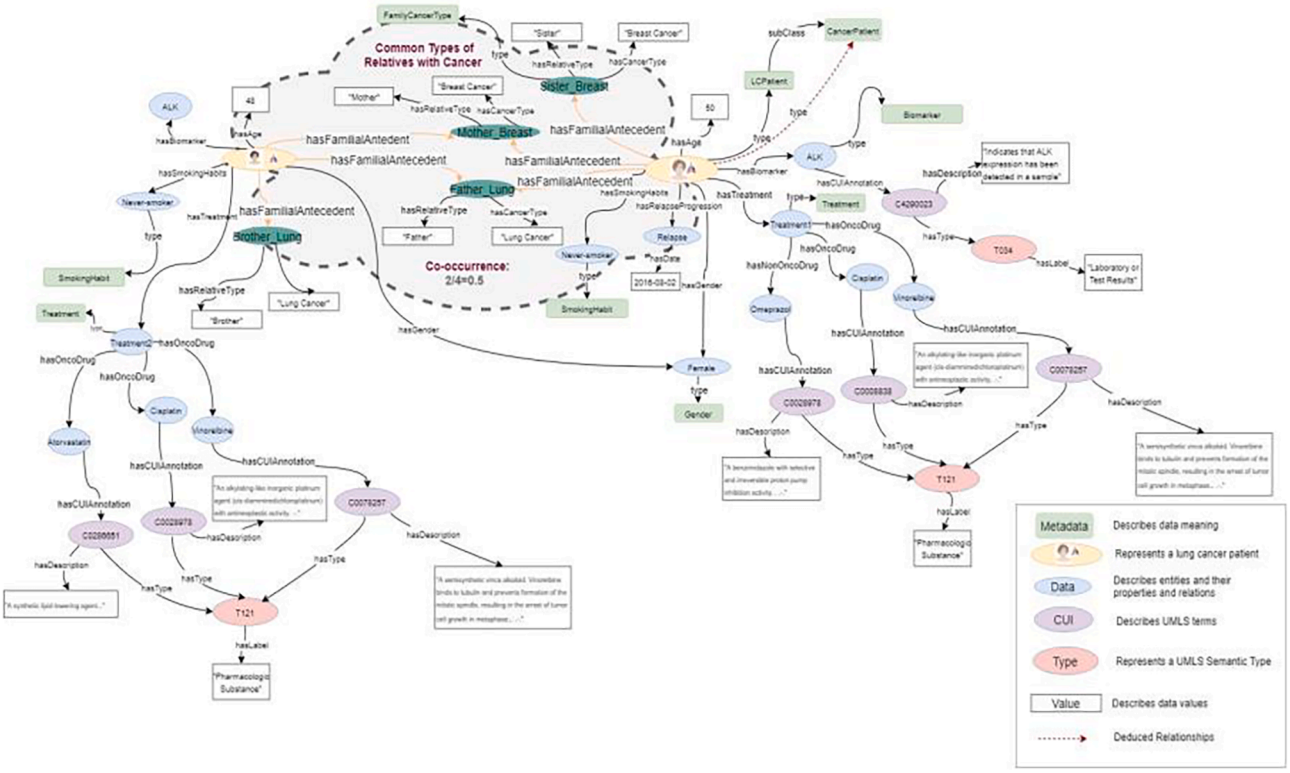


Fig. 8. Familial Cancer Connectedness (FCC).

prevalence of pathogenic germline variants (PGVs) that may facilitate the assessment of genetic cancer risk. Genetic susceptibility to LC is mainly mediated by pathogenic or likely PGVs that are specifically associated with cancer risk. However, determining an appropriate population for genetic testing remains uncertain. In the GERMLUNG study, selection criteria for enrolment included individuals with histologically confirmed lung adenocarcinoma and one of the following conditions: A) family history of LC, it was defined as having one first-degree relative or two or more second-degree relatives with LC, regardless of the age at diagnosis. B) Age at diagnosis $\leq$ 50 years, or $\leq$ 60 years with a pack-year index $<$ 20. C) Presence of $\geq$ 1 actionable genomic alterations (EGFR, ALK, ROS1, KRAS, BRAF, MET exon 14 skipping, or RET) [35].

Compared to other malignancies, there are few studies that have investigated the role of FHC in LC patients. Furthermore, the results of these studies are very heterogeneous. Recently, a systematic review of the literature was conducted according to PRISMA guidelines, searching PubMed and Scopus databases to identify studies reporting on the role of FHC in lung cancer patients [36]. FAHIC-Lung is the first cross-sectional/prospective study specifically designed to identify patterns of FHC and within-family clusters of other risk factors, including smoking habit, to guide patients with NSCLC towards systematic genetic counselling. Recognizing the largely heterogeneous results of this systematic review and considering the clinical implications of detecting pathogenic germline variants (PGVs), the FAHIC-Lung study aims to identify patients potentially enriched for PGVs/likely PGVs to guide them to germline screening outside the research setting.

5. Conclusions

In our research, patients with a FHC were diagnosed at a younger age and were more likely to be women. Young lung cancer patients ($\leq$50 years old) had two or more affected relatives with cancer. FHC is a potential risk factor and predictor of an increased risk of developing LC, especially in people under the age of 50. This association is already well supported by several studies. We believe that any strategy to prevent LC should consider the role of FHC. AI-assistance analysis is an emerging tool that can help us to improve the collection and interpretation of clinical data and should be further explored and enhanced to be incorporated into routine practice.

6. Data sharing

The data analyzed and presented in this study will be available from the corresponding author upon reasonable request, provided that the request meets local ethical and research governance criteria after publication. Patient-level data will be anonymized, and study documents will be redacted to protect the privacy of study participants. A research proposal should be included, which will be evaluated by the Spanish Lung Cancer Group and the Ethics Committee for Clinical Investigation.

Funding

This work was supported by the European Union Horizon 2020 Research and Innovation Program [grant number 875160, CLARIFY Project].

CRediT authorship contribution statement

Virginia Calvo: Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Emetis Niazmand:** Writing – review & editing, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Enric Carcereny:** Writing – review & editing, Data curation. **Delvys Rodriguez-Abreu:** Writing – review & editing, Data curation. **Manuel Cobo:** Writing – review & editing, Data curation. **Rafael López-Castro:** Writing – review & editing, Data curation. **María Guirado:** Writing – review & editing, Data curation. **Carlos Camps:** Writing – review &

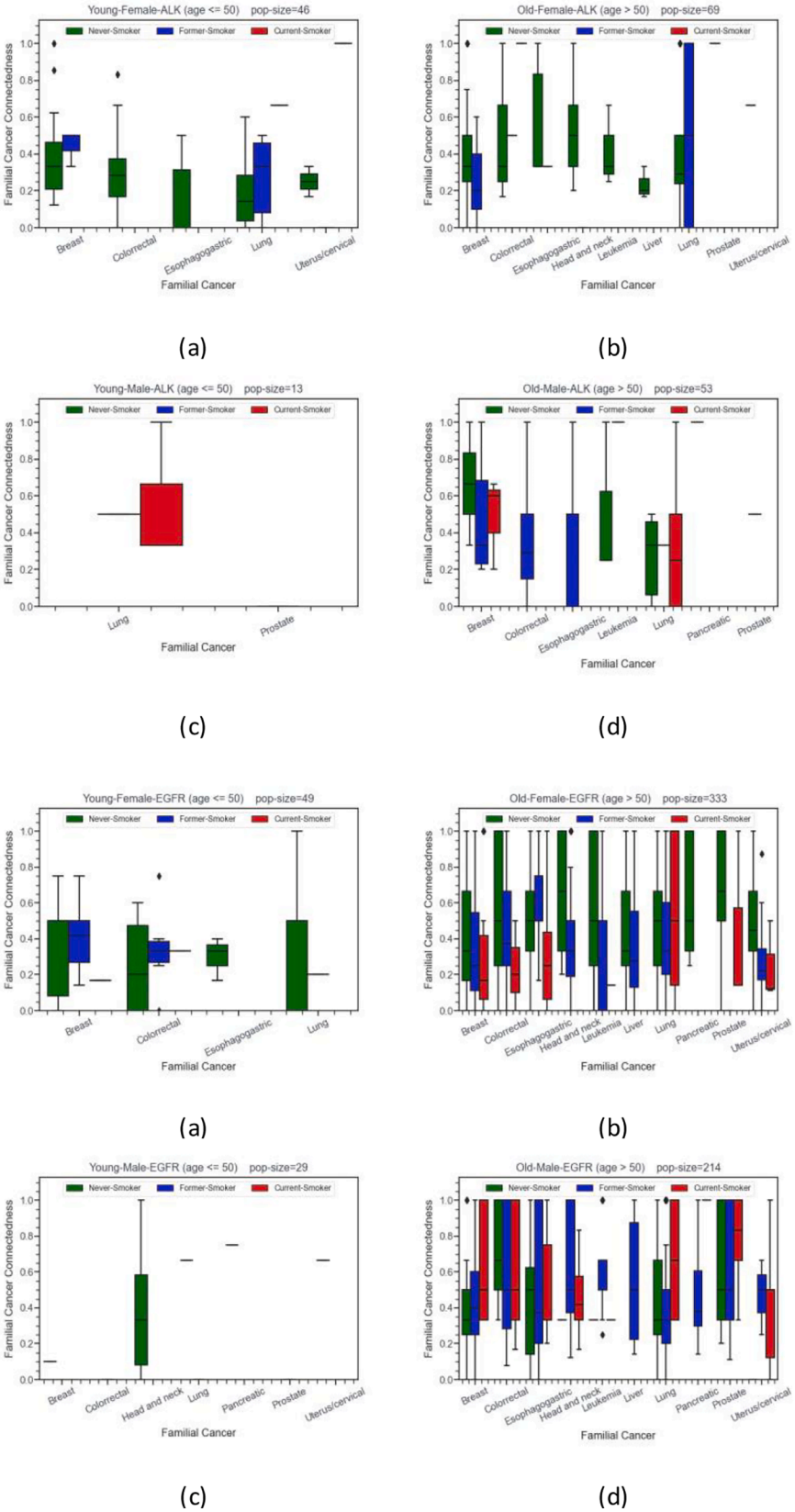


Fig. 9. Familial Cancer Connectedness (FCC) regarding major cancers in relatives and smoking habits of patients with ALK translocation and EGFR mutation in four categories.

Table 6
Odds-Ratio *P (ALK translocation or EGFR mutation) / P (other mutation)*.

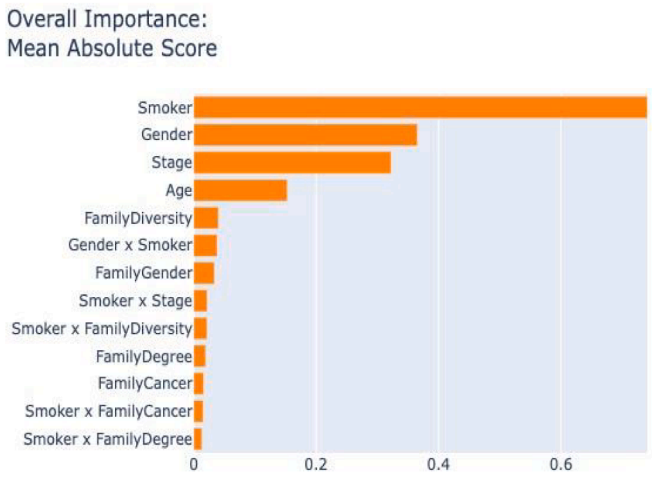
	Male		Female	
	Young	Old	Young	Old
Never smoker	3.3915	2.2082	7.2042	4.6908
Former smoker	0.4936	0.3214	1.0485	0.6827
Current smoker	0.2111	0.1375	0.4485	0.292

$p < 0.001$.

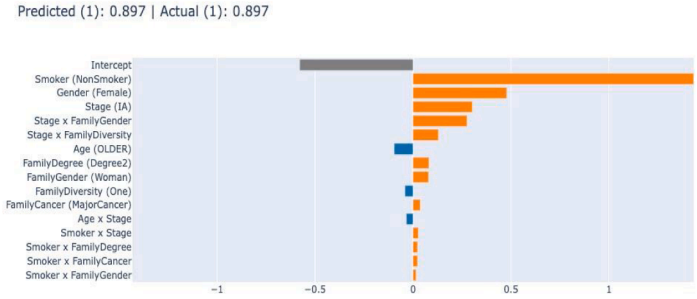
editing, Data curation. **Ana Laura Ortega:** Writing – review & editing, Data curation. **Reyes Bernabé:** Writing – review & editing, Data curation. **Bartomeu Massutí:** Writing – review & editing, Data curation. **Rosario Garcia-Campelo:** Writing – review & editing, Data curation. **Edel del Barco:** Writing – review & editing, Data curation. **José Luis González-Larriba:** Writing – review & editing, Data curation. **Joaquim Bosch-Barrera:** Writing – review & editing, Data curation. **Marta Martínez:** Writing – review & editing, Data curation. **María Torrente:** Writing – review & editing, Data curation. **María-Esther Vidal:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization. **Mariano Provencio:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Dr. Calvo reported receiving personal fees from Roche, Bristol Myers Squibb, Merck Sharp & Dohme, AstraZeneca, Takeda, Pfizer, Sanofi and AMGEN outside the submitted work; being a consultant or advisor and speakerś bureau for Roche, Bristol Myers Squibb/Celgene, Merck Sharp & Dohme, AstraZeneca, Takeda Sanofi and AMGEN and receiving travel, accommodations, and expenses from Takeda, Roche, Bristol Myers Squibb and Merck Sharp & Dohme. Dr. Carcereny reported receiving consulting fees from AstraZeneca, Novartis, Boehringer Ingelheim, Roche, Bristol Myers Squibb, Merck Sharp & Dohme; payment or honoraria from AstraZeneca, Roche, Boehringer Ingelheim, Bristol Myers Squibb, Pfizer and receiving travel, accommodations, and expenses from Roche, Merck Sharp & Dohme, Takeda, Bristol Myers Squibb and Pfizer. Dr. Cobo reported receiving consulting fees from Novartis, AstraZeneca, Boehringer-Ingelheim, Roche, Bristol Myers Squibb, Merck Sharp & Dohme, Lilly, Takeda, Pfizer, Kyowa, Sanofi, Jansen; payment or honoraria from Novartis, AstraZeneca, Boehringer-Ingelheim, Roche, Bristol Myers Squibb, Merck Sharp & Dohme, Lilly, Takeda, Kyowa, Pierre-Fabre, Novocure, Sanofi, Jansen and receiving travel, accommodations, and expenses from AstraZeneca, Boehringer-Ingelheim, Roche, Bristol Myers Squibb, Merck Sharp & Dohme, Lilly, Pierre-Fabre and

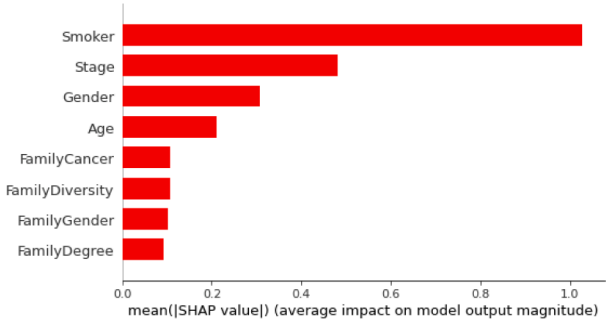


(a) EBM global feature explanation.

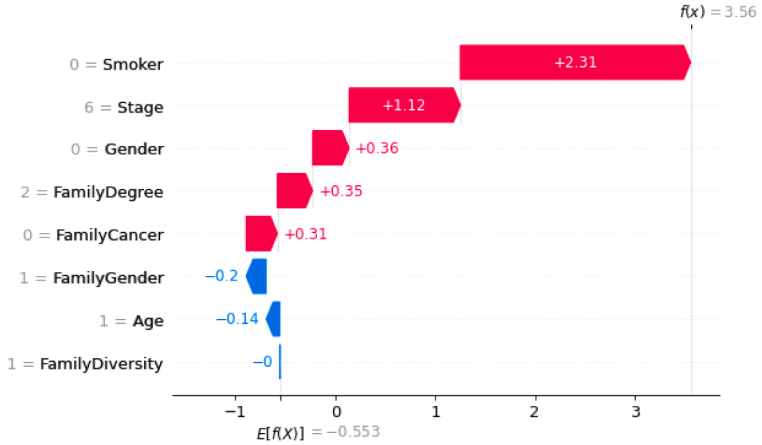


(b) EBM local feature explanation.

Fig. 10. EBM explanation model.



(a) SHAP global feature explanation.



(b) SHAP local feature explanation.

Fig. 11. SHAP explanation model.

being advisor for Gilead. Dr. López-Castro reported receiving consulting fees from Takeda, Roche, Pfizer, Novartis, Pierre-Fabre, Bristol Myers Squibb; payment or honoraria from Pfizer, Pierre-Fabre, Takeda, Roche, Novartis Jansen and receiving travel, accommodations, and expenses from Roche, Takeda, Novartis and Merck Sharp & Dohme. Dr. Bernabé reported receiving grant from Roche; payment or honoraria from Roche, Merck Sharp & Dohme, Pfizer, AMGEN, Takeda, Astrazeneca and receiving travel, accommodations, and expenses from Roche, Bristol Myers Squibb and being advisor for Takeda, Roche, Bristol Myers Squibb and Astrazeneca. Dr. Garcia-Campelo reported receiving consulting fees from AMGEN, AstraZeneca, Bayer, Bristol Myers Squibb, Boehringer Ingelheim, Roche, Janssen, Lilly, Merck Sharp & Dohme, Pfizer, Sanofi, Takeda, Pfizer; payment or honoraria from AMGEN, AstraZeneca, Bayer, Bristol Myers Squibb, Boehringer Ingelheim, Roche, Janssen, Lilly, Merck Sharp & Dohme, Pfizer, Sanofi, Takeda and receiving travel, accommodations, and expenses from AstraZeneca, Roche, Pfizer and being advisor for AstraZeneca. Dr. del Barco reported receiving travel, accommodations and expenses from Merck Sharp & Dohme. Dr. Bosch-Barrera reported receiving grants from Pfizer and Roche outside the submitted work; payment or honoraria from Astrazeneca, Pfizer, Merck Sharp & Dohme, Bristol Myers Squibb, Roche, Sanofi and receiving travel, accommodations, and expenses from Merck Sharp & Dohme, Roche and Takeda. Dr. Provencio reported receiving grants from Bristol Myers Squibb, Takeda, Roche, Pfizer, and Merck Sharp & Dohme outside the submitted work; being a consultant or advisor for Bristol Myers Squibb, Roche, Merck Sharp & Dohme, AstraZeneca, and Takeda; being on the speakers' bureau for Bristol Myers Squibb, Roche, AstraZeneca, and Merck Sharp & Dohme; receiving research funds from Pierre Fabre, Roche, Boehringer Ingelheim, and Bristol Myers Squibb; and receiving travel, accommodations, and expenses from Roche, Bristol Myers Squibb, and AstraZeneca. The remaining authors declare no conflict of interest.

Acknowledgments

The authors thank all investigators who enrolled patients in the TTR. We also thank the patients, their families, and all participating clinical teams, as well as the staff of the Spanish Lung Cancer Group for coordinating the study, especially Eva Pereira and María Fernández. This study was supported by the European Union Horizon 2020 Research and Innovation Program (H2020) under grant agreement n° 875160 CLARIFY project.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.lungcan.2024.107920.

References

[1] R.L. Siegel, A.N. Giaquinto, A. Jemal, Cancer statistics, 2024, CA Cancer J. Clin. 74 (1) (2024) 12–49, https://doi.org/10.3322/caac.21820.
[2] Las cifras del cáncer en España 2024. ISBN: 978-84-09-58445-1. Sociedad Española de Oncología Médica.
[3] R.S. Herbst, J.V. Heymach, S.M. Lippman, Lung cancer, N. Engl. J. Med. 359 (13) (2008) 1367–1380, https://doi.org/10.1056/NEJMra0802714.
[4] D. Loomis, Y. Grosse, B. Lauby-Secretan, F. El Ghissassi, V. Bouvard, L. Benbrahim-Tallaa, N. Guha, R. Baan, H. Mattock, K. Straif, International agency for research on cancer monograph working group IARC. The carcinogenicity of outdoor air pollution, Lancet Oncol. 14 (13) (2013) 1262–1263, https://doi.org/10.1016/s1470-2045(13)70487-x.
[5] J.A. Barta, C.A. Powell, J.P. Wisnivesky, Global epidemiology of lung cancer, Ann. Glob. Health. 85 (1) (2019) 8, https://doi.org/10.5334/aogh.2419.
[6] A.D. Lopez, C.D. Mathers, M. Ezzati, D.T. Jamison, C. Murray, Measuring the global burden of disease and risk factors, 1990-2001. In: Lopez AD, Mathers CD, Ezzati M,

et al. (Eds.), Global burden of disease and risk factors. Internet. World Bank, Washington (DC), pp. 1-13, 2006. (Chapter 1).
[7] F. Islami, A. Goding Sauer, K.D. Miller, R.L. Siegel, S.A. Fedewa, E.J. Jacobs, M. L. McCullough, A.V. Patel, J. Ma, I. Soerjomataram, W.D. Flanders, O.W. Brawley, S.M. Gapstur, A. Jemal, Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the United States, CA Cancer J. Clin. 68 (1) (2018) 31–54, https://doi.org/10.3322/caac.21440.
[8] K.F. Brown, H. Rumgay, C. Dunlop, M. Ryan, F. Quartly, A. Cox, A. Deas, L. Elliss-Brookes, A. Gavin, L. Hounsome, D. Huws, N. Ormiston-Smith, J. Shelton, C. White, D.M. Parkin, The fraction of cancer attributable to modifiable risk factors in England, Wales, Scotland, Northern Ireland, and the United Kingdom in 2015, Br. J. Cancer. 118 (8) (2018) 1130–1141, https://doi.org/10.1038/s41416-018-0029-6.
[9] F. Islami, W. Chen, X.Q. Yu, J. Lortet-Tieulent, R. Zheng, W.D. Flanders, C. Xia, M. J. Thun, S.M. Gapstur, M. Ezzati, A. Jemal, Cancer deaths and cases attributable to lifestyle factors and infections in China, 2013, Ann. Oncol. 28 (10) (2017) 2567–2574, https://doi.org/10.1093/annonc/mdx342.
[10] X. Ni, N. Xu, Q. Wang, Meta-analysis and systematic review in environmental tobacco smoke risk of female lung cancer by research type, Int. J. Environ. Res. Public Health. 15 (7) (2018) 1348, https://doi.org/10.3390/ijerph15071348.
[11] S. Environmental Protection Agency, 2012. A Citizeńs Guide to Radon: the Guide to Protecting Yourself and Your Family from Radon. U.S.: Environmental Protection Agency.
[12] P. Brennan, P. Hainaut, P. Boffetta, Genetics of lung-cancer susceptibility, Lancet Oncol. 12 (4) (2011) 399–408, https://doi.org/10.1016/S1470-2045(10)70126-1.
[13] American Society of Clinical Oncology, American Society of Clinical Oncology policy statement update: genetic testing for cancer susceptibility, J. Clin. Oncol. 21 (12) (2003) 2397–2406, https://doi.org/10.1200/JCO.2003.03.189.
[14] M.J. Khoury, R. Davis, M. Gwinn, M.L. Lindegren, P. Yoon, Do we need genomic research for the prevention of common diseases with environmental causes? Am. J. Epidemiol. 161 (9) (2005) 799–805, https://doi.org/10.1093/aje/kwi113.
[15] E. Iglesias, S. Jozashoori, D. Chaves-Fraga, D. Collarana, M. Vidal, Sdm-rdfizer: an RML interpreter for the efficient creation of RDF knowledge graphs, in: M. d'Aquin, S. Dietze, C. Hauff, E. Curry, P. Cudŕe-Mauroux (Eds.), CIKM '20, 19–23, Virtual Event, Ireland, 2020, pp. 3039–3046.
[16] M. Vidal, K.M. Endris, S. Jazashoori, A. Sakor, A. Rivas, Transforming heterogeneous data into knowledge for personalized treatments - a use case, Datenbank-Spektrum 19 (2) (2019) 95–106.
[17] T. Verma, J. Pearl, An algorithm for deciding if a set of observed independencies has a causal explanation, in: Uncertainty in Artificial Intelligence, Elsevier, 1992, pp. 323–330.
[18] R.R. Bouckaert, Bayesian network classifiers in WEKA for version 3-5-7, Artificial Intelligence Tools 11 (3) (2008) 369–387.
[19] K. Murphy, et al., The bayes net toolbox for matlab, Computing Science and Statistics 33 (2) (2001) 1024–1034.
[20] C. Wallace, K.B. Korb, H. Dai, Causal Discovery via Mml. in: ICML 96 (1996) 516–524.
[21] Y. Lou, R. Caruana, J. Gehrke, G. Hooker, Accurate intelligible models with pairwise interactions, in: In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2013, pp. 623–631.
[22] S.M. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, Advances in Neural Information Processing Systems 30 (2017).
[23] H. Lin, Y.S. Huang, H.H. Yan, X.N. Yang, W.Z. Zhong, H.W. Ye, J.J. Yang, Q. Zhou, Y.L. Wu, A family history of cancer and lung cancer risk in never-smokers: a clinic-based case-control study, Lung Cancer. 89 (2) (2015) 94–98, https://doi.org/10.1016/j.lungcan.2015.05.017.
[24] D. Isla, E. Felip, N. Viñolas, M. Provencio, M. Majem, A. Artal, I. Bover, P. Lianes, D.E. Las Peñas, S. Catot, D.E. Castro, A. Blasco, J. Terrasa, J.L. Gonzalez-Larriba, O. Juan, M. Dómine, R. Bernabe, P. Garrido, Lung cancer in women with a family history of cancer: the Spanish female-specific database WORLD07, Anticancer Res. 36 (12) (2016) 6647–6653, https://doi.org/10.21873/anticanres.11273.
[25] N. Li, K. Shao, Z. Chen, B. Qiu, Z. Wang, F. Tan, J. Wang, X. Tan, B. Li, M. Xiong, F. Zhou, J. He, The impact of positive cancer family history on the clinical features and outcome of patients with non-small cell lung cancer, Fam Cancer 10 (2) (2011) 331–336, https://doi.org/10.1007/s10689-010-9403-0.
[26] A.K. Ganti, F.R. Loberiza Jr, A. Kessinger, Association of positive family history with survival of patients with lung cancer, Lung Cancer. 63 (1) (2009) 136–139, https://doi.org/10.1016/j.lungcan.2008.04.008.
[27] M. Kanwal, X. Ding, Y. Cao, Familial risk for lung cancer, Oncol. Lett. 13 (2) (2017) 535–542, https://doi.org/10.3892/ol.2016.5518.
[28] Y. He, S. Li, S. Ren, W. Cai, X. Li, C. Zhao, J. Li, X. Chen, G. Gao, W. Li, F. Zhou, C. Zhou, Impact of family history of cancer on the incidence of mutation in epidermal growth factor receptor gene in non-small cell lung cancer patients, Lung Cancer. 81 (2) (2013) 162–166, https://doi.org/10.1016/j.lungcan.2013.05.004.
[29] A. Matakidou, T. Eisen, R.S. Houlston, Systematic review of the relationship between family history and lung cancer risk, Br. J. Cancer. 93 (7) (2005) 825–833, https://doi.org/10.1038/sj.bjc.6602769.
[30] J. Rachtan, A. Sokołowski, S. Niepsuj, B. Zemła, M. Zwierko, Familial lung cancer risk among women in Poland, Lung Cancer. 65 (2) (2009) 138–143, https://doi.org/10.1016/j.lungcan.2008.10.029.
[31] D.R. Brenner, R.J. Hung, M.S. Tsao, F.A. Shepherd, M.R. Johnston, S. Narod, W. Rubenstein, J.R. McLaughlin, Lung cancer risk in never-smokers: a population-based case-control study of epidemiologic risk factors. BMC Cancer. (14) (2010) 285, doi: 10.1186/1471-2407-10-285.
[32] E.J. Salber, B. Macmahon, Cigarette smoking among high school students related to social class and parenteral smoking habits, Am. J. Public Health Nations Health. 51 (12) (1961) 1780–1789, https://doi.org/10.2105/ajph.51.12.1780.
[33] S. Jonsson, U. Thorsteinsdottir, D.F. Gudbjartsson, H.H. Jonsson, K. Kristjansson, S. Arnason, V. Gudnason, H.J. Isaksson, J. Hallgrimsson, J.R. Gulcher, L. T. Amundadottir, A. Kong, K. Stefansson, Familial risk of lung carcinoma in the Icelandic population, JAMA. 292 (24) (2004) 2977–2983, https://doi.org/10.1001/jama.292.24.2977.
[34] M.L. Coté, M. Liu, S. Bonassi, M. Neri, A.G. Schwartz, D.C. Christiani, M.R. Spitz, J. E. Muscat, G. Rennert, K.K. Aben, A.S. Andrew, V. Bencko, H. Bickeböller, P. Boffetta, P. Brennan, H. Brenner, E.J. Duell, E. Fabianova, J.K. Field, L. Foretova, S. Friis, C.C. Harris, I. Holcatova, Y.C. Hong, D. Isla, V. Janout, L.A. Kiemeney, C. Kiyohara, Q. Lan, P. Lazarus, J. Lissowska, L. Le Marchand, D. Mates, K. Matsuo, J.I. Mayordomo, J.R. McLaughlin, H. Morgenstern, H. Müeller, I. Orlow, B.J. Park, M. Pinchev, O.Y. Raji, H.S. Rennert, P. Rudnai, A. Seow, I. Stucker, N. Szeszenia-Dabrowska, M. Dawn Teare, A. Tjønnelan, D. Ugolini, H.F. van der Heijden, E. Wichmann, J.K. Wiencke, P.J. Woll, P. Yang, D. Zaridze, Z.F. Zhang, C.J. Etzel, R.J. Hung, Increased risk of lung cancer in individuals with a family history of the disease: a pooled analysis from the International Lung Cancer Consortium, Eur. J. Cancer. 48 (13) (2012) 1957–1968, https://doi.org/10.1016/j.ejca.2012.01.038.
[35] O. Arrieta, E. Caballé-Pérez, N. Hernández-Pedro, E. Romero-Nuñez, J. Lucio-Lozada, C. Castillo-Ruiz, K. Acevedo-Castillo, R. María Álvarez-Gómez, C. Molina-Garay, M. Jiménez-Olivares, K. Carrillo-Sánchez, E. Cristina Mendoza-Caamal, A. F. Cardona, J. Remon, C. Alaez-Verson, Prevalence of pathogenic or likely pathogenic germline variants in cancer predisposition genes among selected patients with lung adenocarcinoma: the GERMLUNG study, Lung Cancer. 194 (2024) 107864, https://doi.org/10.1016/j.lungcan.2024.107864.
[36] F. Citarella, K. Takada, P. Cascetta, P. Crucitti, R. Petti, B. Vincenzi, G. Tonini, F. M. Venanzi, A. Bulotta, S. Oresti, C. Greco, S. Ramella, L. Crinò, A. Delmonte, R. Ferrara, M. Di Maio, F. Gurrieri, A. Cortellini, Clinical implications of the family history in patients with lung cancer: a systematic review of the literature and a new cross-sectional/prospective study design (FAHIC: lung), J. Transl. Med. 22 (1) (2024) 714, https://doi.org/10.1186/s12967-024-05538-4.