

"Development of a new strategy for early diagnosis and treatment of suspected lung cancer through artificial intelligence in Chile"

José Peña, Juan Cristóbal Morales, Carola Sánchez, Natalie Pinto, Rodrigo Villarroel, Claudio Montenegro, Juan Carlos Opazo, Richard Weber, Bruno Nervi, Sebastián Santana, Mariano Suárez.

About the project

Chilesincáncer Foundation -with the support of MSD Laboratory and in collaboration with the Pontifical Catholic University of Chile, the Faculty of Engineering at the University of Chile, the Southeast Metropolitan Health Service, the Complex Engineering Systems Institute, and Complejo Asistencial Dr. Sótero del Río Hospital- has developed a research project initiated in 2020. The project aims to create a new strategy for the early diagnosis and treatment of lung cancer using artificial intelligence in Chile.

This is an example of public-private partnership and collaboration to develop supportive tools aimed at improving the health of our country.

About Chilesincáncer

Chilesincáncer Foundation is a non-profit organization that was established in 2015 with the purpose of reducing the inequality of opportunities in the face of cancer. We work by forming public-private alliances to seek solutions and provide better opportunities for diagnosis and treatment of cancer for adults served in the public health system. For more information, visit www.chilesincancer.cl

Abstract

This document addresses the pressing issue of early detection of suspected lung cancer in Chile, where late-stage diagnoses and high mortality rates prevail. The authors introduce a novel digital tool powered by artificial intelligence (AI) to enhance the identification of potential lung cancer cases from computed tomography (CT) reports. By focusing on key sections of the reports and employing various machine learning models, including a balanced Random Forest, the tool achieved promising results in terms of accuracy and F1-score. When applied to 13,326 CT chest reports from the entire year 2022, it successfully identified 377 CTs of patients with suspected lung cancer previously undetected and not managed by the local lung cancer specialized team. This study underscores the potential of AI in early cancer detection and highlights the importance of its integration to the work of multidisciplinary teams. By increasing the number of patients timely referred for specialized management, the



proposed tool ("OncovigIA") offers a promising path towards improving lung cancer survival rates in Chile and beyond, with future prospects for its broader implementation, extending it to other cancer types and/or healthcare-related texts for continuous surveillance aiming at the early diagnosis and treatment of cancer.

Introduction

Worldwide, lung cancer is the most lethal cancer, resulting in 1.7 million deaths and killing more people than the next three deadliest cancers combined in the year 2020 [1]. In Chile, cancer leads in all-cause mortality and generates a tremendous social impact, with associated enormous costs for the healthcare system. In this context, lung cancer is among the five types of cancer with the highest risk burden, and ranks first in cancer mortality, being responsible for 3,550 deaths in 2020 [2]. Of note, the 5-year survival rate for lung cancer in Chile is substantially lower than in other countries in the region [3]. One of the main causes explaining this poor performance is the high smoking rates in the country, which are above the OECD average [4], but there are also delays in diagnosis, with the majority of patients being diagnosed with late-stage disease, no national screening or early detection strategies available and only a recent incorporation of lung cancer -starting in 2019- into the major public health program of "Explicit health guarantees" [31]. Chile's public healthcare insurance ("FONASA") and its associated healthcare facilities cover over 80% of the population; it disproportionately concentrates older people, with more comorbidities and adverse social determinants of health [5]. Furthermore, Chilean healthcare institutions and providers form complex networks that often present problems in data integration and case management, increasing the chance of late diagnoses, slow workup studies, and low adherence to treatments.

Lung cancer diagnosis presents a tremendous challenge, and a large proportion of patients receive their first cancer-related care when they are admitted to the emergency department with severe symptoms or cancer-related complications, such as pneumonia [6] or thromboembolism [7]; this situation partly explains their limited survival outcomes. In this context, the concept of early detection becomes extremely important, and one way to accomplish that, given the lack of a national lung cancer screening program, is the active management of pulmonary nodules. These are benign in 95% of cases (e.g. granulomas), but must be properly followed and eventually studied invasively to rule out lung cancer. In very general terms, their management takes into account (a) pre-test probability of cancer according to patient-related risk factors (e.g. smoking, pre-existing lung disease, etc.), (b) characteristics in the nodule itself on imaging studies (solid vs subsolid nodules, calcification pattern, etc.), (c) comorbidities, (d) lung function and (e) patient preferences. Options include getting follow-up imaging, a PET-CT study, obtaining a biopsy, and treating the lesion, either by surgical resection or radiosurgery (Stereotactic body radiation therapy, "SBRT"); the decision is ideally made on a case-by-case evaluation, by an interdisciplinary team specialized in lung cancer management [8].



Starting in 2020, a lung cancer diagnosis and management plan was implemented in one of the largest public hospitals in Chile ("Complejo Asistencial Dr. Sótero del Río"). In this model, a multidisciplinary team cares for patients throughout their diagnostic and treatment procedures and keeps tracking them throughout their recovery process, supported by an informatic system that allows patients' follow-up and provides a communication channel between professionals. To date, this model has already been implemented on 245 patients (80-100 per year approx.), subjectively improving the coordination of care between the different agents involved in their management. The core of the model is case management by a specialized nurse navigator who coordinates team activities and maintains contact with patients throughout the period of diagnostic studies and treatment. However, one of the main limitations of the widespread use of this model is that it still receives a low number of referrals by healthcare professionals, partly due to the lack of fluent communication between primary care centers and hospitals, and also because of the heavy workload faced by specialized medical personnel in charge of referring suspected cases (e.g. emergency physicians, radiologists, etc.).

Thus, here we present the development of a new digital tool for detecting suspected cases as early as possible, which integrates artificial intelligence (AI) into the previously implemented strategy of specialized lung cancer clinical case management by monitoring the reports of chest CTs. This tool seeks to establish a first experience of massive surveillance of lung nodules and suspected lung cancer connected to specialized lung cancer case management. As will be described below, in the mid-long term, we hope this experience could be replicated in the rest of the Chilean health system, with the aim of improving lung cancer outcomes through the early detection and timely treatment of suspected cases.

This document is organized as follows: Section 2 describes a short literature review of cancer detection techniques and cases where AI has been successfully used for this purpose. Section 3 presents the data we used and the different models we developed; we focused on machine learning models for the classification of chest CT reports, text representation, and treatment of imbalanced classes. Section 4 shows the results obtained with the algorithms we trained over our data. Section 5 presents a discussion of these results and their implications, while Section 6 concludes this report and hints at future work avenues.

Literature review

Clinical decision support systems (CDSS) allow medical teams to have specific information related to patients or their diseases, in order to improve care, and a recent systematic review of 24 studies evaluating CDSS in oncology practice concluded that there is a positive impact



of CDSS on quality of care [8]. Furthermore, there are successful experiences in which this type of system has improved the detection of lung cancer cases in CT image sets [28].

A relevant aspect of this project is the use of AI-oriented to health and the study of cancer in particular. [10] conclude that by integrating machine learning in diagnosis, treatment, and management, AI can "empower physicians to provide personalized care of the highest quality in an efficient manner that meets the demands of modern clinical practice". In Chile, these types of applications are still very incipient, although promising initiatives have been developed recently. In an interesting case, a Natural Language Processing algorithm to classify referrals that have guaranteed treatments by law was developed [9].

Al has been applied to lung cancer screening using CXR and chest CT since the 1960s [29]. In this scenario, AI has shown potential for improving the diagnosis of lung cancer through the analysis of computed tomography (CT) reports, and AI-assisted diagnostic systems for CT images have demonstrated considerable diagnostic accuracy for lung cancer diagnosis, making them valuable tools in the field of clinical diagnosis [12]. Also, several AI models have been developed and compared for their performance in lung nodule cancer detection. Potentially, these models could assist physicians by reducing their workload, optimizing hospital operational workflow, and providing more time to develop high-quality doctor-patient relationships [13].

Deep learning techniques have been extensively used for lung cancer detection and classification using CT images. Various deep learning models have been compared for their effectiveness in detecting and classifying lung cancer [14]. One notable AI tool developed by MIT researchers ("Sybil"), has shown promising results in predicting whether a person will develop lung cancer up to six years in advance, with an accuracy of 86% to 94% to predict the development of it in the year to come [15]. Furthermore, AI algorithms with high diagnostic accuracy have been found to improve radiologists' performance in detecting lung cancers on chest X-rays, and increase human acceptance of AI suggestions [16].

While AI has shown promising results in lung cancer detection and diagnosis, there are still challenges to overcome; some include the need for a more extensive validation of AI algorithms, balanced collaboration between men and machines, and the development of models that could adapt to differing clinical contexts [17].

Materials and methods

This report describes the development of an AI tool using medical records and chest CT reports from the hospital "Complejo Asistencial Dr. Sótero del Río" to detect suspected lung cancer or other pulmonary lesions (e.g. lung nodules) that need follow-up studies to rule it out. In the following, we first describe the data we had available in more detail; then we



present the AI methods we used, and finally, the results of the retrospective application of the tool for the detection of cases in a new data set covering an entire year (2022) of chest CT reports.

Data description

We first present the data available for this study. Then we address the particular problem of labeling we had to address due to low data quality.

Lung cancer at "Complejo Asistencial Dr. Sótero del Río"

"Complejo Asistencial Dr. Sótero del Río" is a large secondary hospital in the southeast urban area of Santiago, Chile's capital city. It provides healthcare for a large population, serving the densely populated communities of Puente Alto, La Florida, San Ramón, La Granja, La Pintana, San José de Maipo, and Pirque, covering over 1.5 million people. Table 1 shows the number of chest CTs obtained by year; of note, the reduced number performed in 2020 is explained because of the COVID-19 pandemic and its associated effects on healthcare provision. Table 2 shows some demographic characteristics of the population that underwent chest CT studies in 2021.

Year	Nº Chest CTs
2020	8.844
2021	12.261
2022	13.326

Table 1: CTs at Dr. Sótero del Río displayed by year

Tabla 2.	Ago and	cov dict	ibution fo	r nationte	who	wont through	igh n	choct (T in	2021
I able Z.	Age allu	sex uisu	IDULION 10	i Datients		went thio	igii d	CHESL	- 1 111	ZUZI
	U				-				-	-

Age (mean)	Sex
58,3	Male (51.1%)
60,5	Female (48.5%)

Broadly speaking, chest CT reports in this particular healthcare setting are structured in three main sections: clinical history and diagnostic hypothesis, study findings, and conclusions.

Clinical history: Summarizes patients' relevant clinical history, possible symptoms, and any recent relevant procedures or surgeries. It also includes the reason why the chest CT scan was obtained. This information is important for radiologists to interpret chest CT scan



findings and for eventual patient care or follow-up recommendations.

Findings: This section describes in detail specific findings on the chest CT scan images. Findings are typically divided into two categories: normal findings and abnormal findings. In general terms, the former are expected and do not indicate any medical problems, while on the contrary the latter are not expected and may indicate a medical problem. Radiologists describe abnormal findings in detail, including size, location, and specific features if applicable (Figure 1).

Conclusions: This section summarizes the most important findings from the chest CT study and eventually provides recommendations for patient care. Radiologists may also discuss differential diagnosis, which is a list of possible causes for abnormal findings, and could recommend follow-up tests or procedures if necessary.



Figure 1: Examples of abnormal chest CT findings and their descriptions by radiologists.

Round atelectasis	Lung nodule suspicious of malignancy
An oval subpleural lesion is observed in the left lower lobe, accompanied by curvilinear bands produced by the pulling of the bronchovascular bundles ("comet tail" sign).	There is a right upper lobe lung nodule with spiculated margins, measuring 11 mm and highly suspicious of lung cancer.
This patient does not need specialized case management nor follow-up.	This patient benefits from specialized case management in order to rule out (or treat) lung cancer.

Data labeling

Like the great majority of hospitals in Chile, Complejo Asistencial Dr. Sótero del Río does not have a single, centralized system for tracking all cases of lung cancer. This means there is no unique and accurate way to know how many people are diagnosed each year. Additionally, independent systems that do exist (e.g. pathology studies) are often incomplete or inaccurate. Furthermore, we are not only aiming at identifying definitive or confirmed lung cancers but also at finding people with lung nodules that need tomographic follow-up or other specific management to rule out the diagnosis.

One of the challenges of lung cancer is that many cases are not definitely confirmed (e.g. by biopsy). This is especially true in low- and middle-income countries, where access to cancer screening and diagnostic workups is limited. Manual labeling of chest CT reports can help identify under-represented cases of lung cancer by reviewing scans from patients whose cancer was not confirmed or adequately studied, and also cases in which the disease was



ruled out at a later stage but needed follow-up studies. This information could then be used to target cancer control programs in areas where lung cancer is most prevalent.

Hence and bearing in mind our target population, manual labeling of chest CT reports was necessary to create an accurate dataset for training our models. To be labeled as positive, reports had to describe either a definitely suspicious image (like a lung mass) or a single (or multiple) lung nodule(s) or other images in need of follow-up studies according to the radiology report (like a subsequent chest CT in 3-6 months). Of note, cases were considered "positive" for suspicion irrespective of whether there was another primary tumor described in the same study (e.g. a person being treated for colorectal cancer with a sigmoid tumor develops a suspicious lung nodule). Cases were considered negative in case of definitive metastatic disease to the lungs originating elsewhere. The concept behind this strategy was to exclusively detect lung cancer or its precursor lesions regardless of the simultaneous presence of other cancers but to avoid the detection of overt metastatic lung disease. Initially, we labeled 1,190 cases, of which 180 were considered suspicious and 1,010 were not leading to an imbalanced data set. All of these CT reports corresponded to 2021, so we could evaluate our model later on cases from 2020 and 2022. Table 3 shows the resulting data set.

Table 3: Available data	(CTs year 2021)
-------------------------	-----------------

Labeled	Suspected lung cancer	180
	Not suspected lung cancer	1,010
Not labeled		11,071
Total		12,261

Artificial Intelligence Approaches

In the following subsections, we present machine learning models we used, approaches for text representation, and techniques to treat the class imbalance problem. Finally, we briefly describe how we deploy the developed models.

Models

Decision Trees (DT): Decision trees (DTs) are a tree-like design architecture in which internal nodes represent features and branches represent decision functions. This flowchart-like structure makes DTs useful for decision-making. DTs are nonparametric methods, meaning that they do not make any assumptions about the underlying probability distribution of the data. This makes them robust to overfitting and allows them to be used with high-



dimensional data.

The DT algorithm first splits the data by selecting the best attribute using attribute selection measures (ASMs). The chosen attributes become decision nodes, and the data is split into subsets based on these attributes. The algorithm then recursively repeats this process on each child node until one of the stopping conditions is met, such as all tuples having the same attribute value.

Random Forest (RF): This algorithm works in two steps [18]:

1. Constructing many DT classifiers: In this phase, the RF algorithm constructs a large number of DT classifiers. Each DT classifier is constructed using a random subset of the training data and a random subset of the features. This ensures that each DT classifier is different from the others and that the RF model is not overfitting the training data.

2. Integrating these classifiers into voting: In this phase, the RF algorithm uses the DT classifiers to vote on the test samples. Each DT classifier votes for the class that it predicts for the test sample. The final decision for the test sample is made using the most dominant rule. This means that the class with the most votes is the class that is assigned to the test sample.

XGBoost: Gradient boosting is an ensemble learning algorithm that combines a set of weak learners to create a strong learner. XGBoost is known for its high accuracy and speed, and it is a popular choice for text classification tasks.

XGBoost works by iteratively adding weak learners to a model. Each weak learner is a decision tree that is trained to predict the residual errors of the previous weak learners. The residuals are the differences between the actual values and the predictions of the previous weak learners [19].

BERT (Bidirectional Encoder Representations from Transformers): This is a pre-trained language model that can be used for a variety of natural language processing tasks, including text classification [20]. BERT works by first encoding the input text into a sequence of vectors. These vectors represent the meaning of each word in the text, taking into account the context of the words around it. BERT can then be used to classify the text into a specific category by comparing the encoded text to a set of known categories.

LSTM (Long Short-Term Memory): This is a type of recurrent neural network (RNN) that is well-suited for text classification tasks. RNNs are able to learn long-term dependencies in sequences, which is essential for text classification because words in a sentence often have meaning in relation to other words in the sentence. LSTMs are a particular type of RNN that are able to overcome the vanishing gradient problem, which is a problem that can occur in RNNs when trying to learn long-term dependencies [21].

LSTMs can be used for text classification by first converting the text into a sequence of



numbers. This can be done using a technique called word embedding, which maps each word in the vocabulary to a unique vector of numbers. The LSTM network can then be trained on a dataset of labeled text. The labels for the text can be anything from sentiment (positive, negative, neutral) to topic (sports, politics, entertainment). The LSTM network will learn to associate the sequence of numbers representing the text with the correct label.

Text representation

Text representation is a necessary step in the machine learning process, as it bridges the gap between raw, unstructured text data and the learning algorithm. In this study, we evaluated two types of algorithms:

1. Traditional algorithms, such as Random Forests (RF) and Decision Trees (DTs), require a structured matrix to represent the text. For these algorithms, we represented the text in a matrix using the TF-IDF transformation.

2. Neural network algorithms (such as BERT and LSTM) learn their own representation of the text by creating a word embedding. This embedding can be pre-trained or not.

TF-IDF (Term Frequency - Inverse Document Frequency) is a statistical measure that is used to quantify the importance of a word in a document. It is a widely used technique for text representation in the task of text classification [22]. TF-IDF is calculated as follows:

Term Frequency (TF): This is the number of times a word appears in a document.

Inverse Document Frequency (IDF): This is the logarithm of the number of documents in a corpus divided by the number of documents that contain the word.

The TF-IDF score for a word is calculated by multiplying the TF score for the word by the IDF score for the word. The TF-IDF score for a word is higher if the word appears more frequently in the document and if the word appears less frequently in the corpus.

TF-IDF is a useful measure for text representation because it captures the importance of a word in a document without being influenced by the length of the document. This makes it a good choice for text classification tasks, where the goal is to identify the topic of a document.

Word Embeddings: Word embeddings are a type of representation for text that captures the meaning of words in a way that is useful for machine learning tasks. They are typically represented as vectors of real numbers, where the closeness of two words in the vector space reflects the similarity of their meanings [23].

Word embeddings can be used in text classification tasks by representing each word in a document as its corresponding embedding vector. This allows the machine learning algorithm to learn the relationships between words in the document, which can be used to



classify the document into a particular category.

Class imbalance

Our dataset is imbalanced, with 84.9% of the data belonging to the class "Not suspected lung cancer". Therefore, we performed multiple methods for tackling class imbalance.

First, we trained the RF, DT, and XGBoost models on the original dataset, which was not modified, and used a TF-IDF representation. Second, we used the *imbalanced-learn* library to train two modified versions of Random Forest: the Imbalance Random Forest and the Balanced Bagging Classifier.

Third, we proposed our own version for balancing the data. We performed an undersampling of 25% on the majority class, followed by an oversampling of the minority class, so that both classes had a 1:1 ratio. This technique was found to be better than the unbalanced version, and competitive with the *imbalanced-learn* library in terms of F1-score and Accuracy.

In parallel, we explored training models based on modern techniques such as Transformers. With the help of the Flair library, we fine-tuned a pre-trained BERT model. However, due to the imbalance in the dataset, the models mostly predicted the majority class, resulting in poor performance on the test set.

Deployment

We implemented a computational platform with our predictive models as a web application. The front-end and back-end were implemented using Streamlit in Python. The platform allows two types of inputs: (1) chest CT reports, as text descriptions, and (2) Spreadsheets with chest CT reports.

When a case is evaluated, the platform displays the probabilities of cancer risk and provides a recommendation. When a spreadsheet is evaluated, the platform reports the number of cases where cancer risk is detected and returns the same file with a new column indicating whether cancer risk was detected for that evaluation.

In the modeling process, we realized that clinicians focus their attention on the sections "Clinical history" and "Impressions" of the report written by radiologists. Only in case of doubts do they analyze the "Findings" section. Therefore, we replicated this process in a final version of the model that only considers those two parts for training. This version has less information about the report than the original version.

The version of the model that includes only those two parts outperforms the version of the model that includes the full report since it focuses on the most relevant information.



Therefore, a version of the model where the data is balanced by our process and the #Findings" part are excluded is the most accurate way to train these models. Table 4 summarizes the performance of the different combinations of (model, data) used for training these models. All the results reported are for the test set; the full dataset is summarized in Table 4.

Data	Model	Recall	Precision	F1-score	Accuracy
Not balanced / full report	Random Forest	0.48	0.84	0.77	0.89
Not balanced / full report	Decision Tree	0.48	0.91	0.78	0.90
Not balanced / full report	XGBoost	0.57	0.83	0.81	0.90
Not balanced / full report	Balanced Random Forest	0.93	0.34	0.62	0.66
Not balanced / full report	Balanced Bagging Classifier	0.93	0.43	0.71	0.76
Not balanced / full report	BERT	0.10	0.50	0.55	0.87
Balanced / full report	Random Forest	0.68	0.62	0.78	0.87
Balanced / full report	Decision Tree	0.59	0.51	0.72	0.82
Balanced / full report	XGBoost	0.73	0.70	0.82	0.89
Balanced / Clinical history + Impressions	Random Forest	0.75	0.73	0.84	0.90
Balanced / Clinical history + Impressions	Decision Tree	0.59	0.58	0.74	0.85
Balanced / Clinical history +Impressions	XGBoost	0.70	0.72	0.82	0.90

Table 4: Performance of the best model under different data treatments



As delineated in Table 4, the model exhibiting the highest F1-score and Accuracy aligns with the Random Forest model, which was trained using balanced data and exclusively relied upon the Clinical History and Impressions extracted from the chest CT reports. Notably, due to the inherent data imbalance within our problem, models trained on non-balanced data encompassing the complete report yielded an average F1-score of 0.7. Upon rebalancing the dataset and incorporating the entirety of the report, these models achieved an average F1-score of 0.77. Subsequently, following the rebalancing process and restricting the model inputs to solely the Clinical History and Impressions, the models achieved an average F1-score of 0.8. This outcome signifies a noteworthy improvement of 0.1 in F1-score through the application of our modeling approach.

Evaluation of CTs 2022

After training our model, we decided to validate the results on a new dataset of chest CT reports. As stated before, since we trained on chest CT scans from 2021, we decided to study the ones reported in 2022. To do this, we evaluated all the reports with our best model and handed the results to the specialized clinical team for review.

Our model detected 1,351 suspicious reports out of 13,326 (10.13%). So far the clinical team has evaluated 1209 (100.0%) cases, confirming 621 (51,36%) as correctly labeled by the algorithm for suspected lung cancer; of those, 81 (6,69%) cases were already confirmed lung cancer cases and were being treated accordingly. Interestingly, 377 (31,18%) of the correctly labeled cases were not being treated for lung cancer nor had been referred to the lung cancer team. Therefore, those cases are new findings from the lung cancer team's perspective and could potentially benefit from their specialized management. Figure 2 shows the results after the evaluation of the clinical team over 1209 cases reported as suspicious by the algorithm. De los casos de sospecha de cáncer ¿cuántos detectamos correctamente?



Figure 2: Outcome of the reviewed cases



Discussion

Our work is preceded by the growing relevance of data analysis technologies for the early detection of cancer [5]. On the other hand and regarding case management, multidisciplinary teams (MDT) are considered the gold standard of cancer care; they result in better clinical processes and outcomes for cancer patients, with evidence of improved survival and positive consequences for patients in multiple dimensions [24]. The model proposed in this project considers an AI tool to foster referral to such multidisciplinary teams.

From the clinical standpoint, it is important to highlight several areas. First of all, the preventive approach of this strategy, since it could not only provide surveillance for definitely suspected lung cancer but also for lung nodules that require subsequent studies to rule it out; this is particularly relevant in this context –very common at the regional level– where there is no national Lung Cancer Screening program and the overwhelming majority of lung cancer patients are diagnosed in advanced stages of their illness. Second, it should be noted that the model detected an elevated number of CTs to be studied or followed up, since although it roughly failed half of the time, we can safely assume there will be several hundred cases to manage each year (500-600 CTs approx). Likewise, we must note a significant proportion of cases wrongly labeled by the model had other primary cancers, which must be



taken into account for potential developments to come: theoretically, we could turn to a tool that detected "cancer" in a more general (or "sensitive" way, in diagnostic studies language), filtering subsequently according to their likely primary sites for their appropriate clinical management. Fourthly, this dataset gives us a very valuable opportunity to retrospectively evaluate the effect of specialized nursing management: We will be able to compare the trajectories and outcomes of people who were actually managed by the lung cancer team with those who weren't, assessing for example how many of them were appropriately followed to confirm or rule out lung cancer, how long it took for confirmed cancer patients to begin treatment, etc.. This is extremely important because it is plausible that the nursing case management model could not be sufficient to improve relevant outcomes in this lowresource context as a result of structural barriers that do not respond to patient navigation issues (e.g. lack of equipment, for example for diagnostic studies, such as MRI). Finally, the high number of cases -- and potential lung cancers-- detected demands a more systemic approach aimed at developing a strategy that could effectively enable their timely management, which should include the design of a currently non-existent service that could serve as an interface and promote patient education and empowerment, incorporating them more actively in the pursuit of appropriate suspected cancer management.

Assuming specialized case management delivers the desired outcomes in this context, we plan to scale this solution in a comprehensive process: OncovigIA will periodically deliver suspicious cases to the clinical team in order to initiate their management, with a specialized nurse in charge of the process of expert medical evaluations and diagnostic and treatment procedures. A traceability tool will aid the navigating nurse record and update the patient's current cancer stage while monitoring the necessary steps for each patient; these could include proper confirmation studies in cases of a very high probability of cancer (e.g. biopsy, surgery, etc.), but in the majority of cases identified as bearers of lung nodules in which lung cancer is not probable enough at the time of the index CT will lead to later follow-up studies (e.g. a CT 3-6 months after the original one). We expect this strategy to help allocate scarce resources, shorten the diagnosis and treatment times for patients with cancer, and hopefully improve their survival rate.

Figure 3: Proposed policy for early lung cancer detection, diagnosis, and treatment





At the same time, the implementation of the previously described process will provide additional information to effectively monitor the developed models: We will closely check the system's classification and react accordingly in the case of errors of both types retraining the respective models, particularly during the first months of its application. While the current models are classifying satisfactorily, we anticipate several changes over time that could harm their accuracy. Modifying the current protocol to describe CT scans would lead to different input vectors of the predictive models. In that case, standardization of the respective reports could ensure obtaining systematically adequate input vectors. In the case of permanent changes (new technology revealing novel findings in CT scans, improved protocols, etc.) the respective models could be retrained or alternative models could provide better results; see e.g. Table 4 where different models are compared. Monitoring the model's performance over time could also establish insights regarding the feature's importance, and provide feedback for the respective protocols or specific terminology radiologists should follow when describing the respective findings in CT scans. Furthermore, as described in the next paragraph, we expect this monitoring process will also feed new models to detect different types of cancer.

Extending this lung cancer detection tool to other types of cancer entails several challenges, but it is feasible with certain adjustments and considerations following the project's methodology. Firstly, clinical, radiological, and other relevant data must be collected for the specific cancer to be detected. These data may encompass medical reports, medical images, electronic health records, and other medical documents.



Each type of cancer can exhibit specific features in the data that are pertinent to detection, needing the development of new features. Additionally, machine learning models and algorithms for detecting new types of cancer must be adapted. This entails tuning hyperparameters, selecting relevant features, and training specific models. Collaboration with clinical oncology experts and other medical professionals is essential for the success of such projects. In the literature, machine learning stands out for its high effectiveness in predicting various types of cancer, including breast, brain, lung, liver, and prostate cancer [25]. Of note, we are fully aware of the widely known potential for deep learning techniques for imaging –like CT or magnetic resonance– throughout the entire cancer disease trajectory, including early detection, treatment response prediction, enhancement of radiation therapy planning, and others; in the same way and as stated throughout in this report, we agree with a recent review in the need for multidisciplinary engagement to ensure meeting an appropriate case use, a robust development and testing phase prior to its adoption into healthcare systems, and the requisite of a local ecosystem to promote growth of these initiatives [26]. Bearing all those factors in mind, despite the highly reported diagnostic accuracy of AI in the detection of lung cancer on imaging studies [27], we chose a text-based approach at this first stage for considering it more attractive given its potential applicability to the large number of documents that are produced in a healthcare system on a daily basis that could be actively surveilled, but also given the current state of the local healthcare system and its already ongoing models of care.

As a summary of our planned future work, we will apply OncovigIA to more cases, increasing the experience and thus contributing to its continuous improvement. We will also extend its scope for early detection and management of any kind of cancer present in CT scans and to other types of text contained in a wider set of documents or medical reports (e.g. biopsy reports, medical specialties referrals, outpatient visits records, endoscopy reports, etc.), increasing the sensitivity of continuous surveillance supported by AI and lessening the need for referrals or other bureaucratic procedures, always backed up by specialized case management and interdisciplinary teams. Of note, we recognize the need for developing the "patient side" of this strategy and will work on it integrating service design and potentially other disciplines with the objective of fostering patient education and empowerment. Finally, in another avenue for the scaling of this project and dissemination of tools to increase the timely diagnosis of cancer, we will offer OncovigIA to other public hospitals in Chile.

Conclusions

In this project, we present a novel system called OncovigIA, which uses AI for the early detection of suspected lung cancer in chest CT reports from a large public Hospital in Santiago de Chile to improve their timely referral to a specialized interdisciplinary team. Applying OncovigIA to 13,326 cases we obtained excellent results: Depending on the particular model, we could achieve an F1-score of 80%, and an accuracy of 90% for the early detection of



relevant cases. In practice and considering the clinical context in which it will be implemented, it could increase timely referrals for the early detection and treatment of lung cancer, hopefully improving the outcomes of this patient population. Finally, although we recognize several challenges (and anticipate others), we plan to scale this technology to other types of cancer and medical text documents, with the certain possibility of extending its implementation to other healthcare centers in the country.

Acknowledgment: We gratefully acknowledge financial support by Merck Sharp & Dohme (I.A.) LLC. (Chile), a subsidiary of Merck & Co., Inc., Rahway – New Jersey, USA.

References

[1] World Health Organization (2020), International Agency for Research on Cancer. Rates: cancers by population: All cancer; Globocan. Available at: <u>http://gco.iarc.fr/today/data/factsheets/populations/152-chile-fact-sheets.pdf</u>. Accessed Septermber 8, 2023.

[2] World Health Organization (2018), International Agency for Research on Cancer. Rates: cancers by population: Chile; Globocan. Available at: <u>http://gco.iarc.fr/today/data/factsheets/populations/152-chile-fact-sheets.pdf</u>. Accessed September 8, 2023.

[3] Allemani, C., Matsuda, T., Di Carlo, V., Harewood, R., Matz, M., Nikšić, M. et al. (2018). Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *The Lancet*, 391(10125), 1023-1075.

[4] OECD (2019), Health at a Glance: OECD Indicators, OECD Publishing, Paris, <u>https://doi.org/10.1787/4dd50c09-en</u>.

[5] Okma, K. G., Tenbensel, T., & Crivelli, L. (Eds.). (2020). *Health Reforms Across the World: The Experience of Twelve Small and Medium-sized Nations with Changing Their Healthcare Systems*. World Scientific.

[6] Valvani, A., Martin, A., Devarajan, A., & Chandy, D. (2019). Postobstructive pneumonia in lung cancer. Annals of Translational Medicine, 7(15).

[7] Vitale, C., D'Amato, M., Calabrò, P., Stanziola, A. A., Mormile, M., & Molino, A. (2015). Venous thromboembolism and lung cancer: a review. *Multidisciplinary respiratory medicine*, *10*(1), 1-9.
[8] Mazzone, P. J., & Lam, L. (2022). Evaluating the patient with a pulmonary nodule: a review. *Jama*, *327*(3), 264-273.

[9] Villena, F., Pérez, J., Lagos, R., & Dunstan, J. (2021). Supporting the classification of patients in public hospitals in Chile by designing, deploying and validating a system based on natural language processing. *BMC medical informatics and decision making*, *21*(1), 1-11.

[10] Goetz, L. H., & Schork, N. J. (2018). Personalized medicine: motivation, challenges, and progress. *Fertility and sterility*, *109*(6), 952-963.



[11] Chiu, H. Y., Chao, H. S., & Chen, Y. M. (2022). Application of artificial intelligence in lung cancer. *Cancers*, *14*(6), 1370.

[12] Liu, M., Wu, J., Wang, N., Zhang, X., Bai, Y., Guo, J., & Tao, K. (2023). The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis. *Plos one*, *18*(3), e0273445.

[13] Sathyakumar, K., Munoz, M., Singh, J., Hussain, N., Babu, B. A., & Babu, B. (2020). Automated lung cancer detection using artificial intelligence (AI) deep convolutional neural networks: A narrative literature review. *Cureus*, *12*(8).

[14] Thanoon, M. A., Zulkifley, M. A., Mohd Zainuri, M. A. A., & Abdani, S. R. (2023). A Review of Deep Learning Techniques for Lung Cancer Screening and Diagnosis Based on CT Images. *Diagnostics*, *13*(16), 2617.

[15] Zheng, X., He, B., Hu, Y., Ren, M., Chen, Z., Zhang, Z., & Li, G. (2022). Diagnostic accuracy of deep learning and radiomics in lung cancer staging: a systematic review and meta-analysis. *Frontiers in Public Health*, *10*, 938113.

[16] Lee, J. H., Hong, H., Nam, G., Hwang, E. J., & Park, C. M. (2023). Effect of Human-AI Interaction on Detection of Malignant Lung Nodules on Chest Radiographs. *Radiology*, *307*(5), e222976.

[17] Schreuder, A., Scholten, E. T., van Ginneken, B., & Jacobs, C. (2021). Artificial intelligence for detection and characterization of pulmonary nodules in lung cancer CT screening: ready for practice?. *Translational lung cancer research*, *10*(5), 2378.

[18] Breiman, L. (2001). Random forests. Machine learning, 45, 5-32.

[19] Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., & Zhou, T. (2015). Xgboost: extreme gradient boosting. R package version 0.4-2, 1(4), 1-4.

[20] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

[21] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[22] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information processing & management*, *24*(5), 513-523.

[23] Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C. C. J. (2019). Evaluating word embedding models: Methods and experimental results. *APSIPA transactions on signal and information processing*, *8*, e19.

[24] Taberna, M., Gil Moncayo, F., Jané-Salas, E., Antonio, M., Arribas, L., Vilajosana, E., & Mesía, R. (2020). The multidisciplinary team (MDT) approach and quality of care. *Frontiers in oncology*, *10*, 85.

[25] Zhang, B., Shi, H., & Wang, H. (2023). Machine Learning and AI in Cancer Prognosis, Prediction, and Treatment Selection: A Critical Approach. *Journal of Multidisciplinary Healthcare*, 1779-1791.

[26] Koh, D. M., Papanikolaou, N., Bick, U., Illing, R., Kahn Jr, C. E., Kalpathi-Cramer, J., & Prior, F. (2022). Artificial intelligence and machine learning in cancer imaging. *Communications Medicine*, 2(1), 133.

[27] Liu, M., Wu, J., Wang, N., Zhang, X., Bai, Y., Guo, J., & Tao, K. (2023). The value of artificial intelligence in the diagnosis of lung cancer: A systematic review and meta-analysis. *Plos one*, *18*(3), e0273445.



[28] Lanjewar, M. G., Panchbhai, K. G., & Charanarur, P. (2023). Lung cancer detection from CT scans using modified DenseNet with feature selection methods and ML classifiers. *Expert Systems with Applications*, 224, 119961.

[29] Jacobs, C., Setio, A. A., Scholten, E. T., Gerke, P. K., Bhattacharya, H., M. Hoesein, F. A., & Farahani, K. (2021). Deep learning for lung cancer detection on screening CT scans: results of a large-scale public competition and an observer study with 11 radiologists. *Radiology: Artificial Intelligence*, *3*(6), e210027.

[30] Mount Sinai. (2022). Lung Cancer Screening Dramatically Increases Long-term Survival Rate. Available in: <u>https://www.mountsinai.org/about/newsroom/2022/lung-cancer-screening-dramatically-increases-long-term-survival-rate</u>. Accessed September 8, 2023.

[31] Biblioteca del Congreso Nacional de Chile. (2014). Plan GES o AUGE. https://www.bcn.cl/portal/leyfacil/recurso/plan-ges-ex-auge. Accessed November 18, 2023

José Peña Durán	Education:
	2020 Certificate in Health Innovation. MIT / Center for Primary Care, Harvard Medical
	School.
	2016 Subspecialty Postgraduate Program in Medical Oncology. Pontifical Catholic
	University of Chile.
	2012 Postgraduate Program in Internal Medicine. Pontifical Catholic University of Chile.
	2008 Degree in Medicine and Surgery. Pontifical Catholic University of Chile.
	Current position:
	-Assistant Clinical Professor in the Department of Hematology and Oncology, Pontificia
	Universidad Católica de Chile (Pontifical Catholic University of Chile).
	-Advisor to the Chilean Ministry of Health in the Cancer Department, DIPRECE.
	Technical Advisor to the Coordination Unit of the Oncology Network in the Servicio de
	salud Metropolitano Sur Oriente (Southeast Metropolitan Health Service).
	-Director of the Epistemonikos Foundation.
Richard Weber	Education:
	1986: Diploma in Mathematics, RWTH Aachen University, Aachen, Germany
	1988: M.Sc. in Operations Research, RWTH Aachen University, Aachen, Germany
	1992: Ph.D. in Operations Research, RWTH Aachen University, Aachen, Germany
	Current position:
	10/2019 – present: Director Department of Industrial Engineering, FCFM, Universidad de
	Chile
	09/2018 – 09/2019: Director (S) Department of Graduate Studies, Universidad de Chile
	08/2014 – 06/2018: Director Graduate School, Faculty of Physical Sciences and
	Mathematics, Universidad de Chile
	1999 - present: Professor; Universidad de Chile, Department of Industrial Engineering.
Juan Cristóbal	Education:
Morales	Diploma in Leadership in Digital Transformation in Health / Harvard Medical School
	Master's in Health Administration. Pontifical Catholic University of Chile.
	Master's in Family Sciences. University of Santiago de Compostela. Spain.
	Fellowsnip in Coagulopathies. University Hospital La Fe. Valencia. Spain.
	Diploma in Ultrasonography in Coagulation Disorders. University of California. San Diego.
1	USA.

About the team & researchers



	Postgraduate Program in Adult Family Medicine. Pontifical Catholic University of Chile.
	Degree in Medicine and Surgery. Pontifical Catholic University of Chile.
	Current position:
	- Head of Digital Health Department. South East Metropolitan Health Service
	Directorate.
	- Coordinator of the Hemophilia and Coagulopathies Program, Dr. Sótero del Río
	Hospital.
	- Clinical activity in the Coagulopathies Unit at Dr. Sótero del Río Hospital and the -
	Department of Family Medicine, UC-Christus Health Network.
	 Researcher in the field of Coagulation Disorders.
Bruno Nervi	Education:
	2003 – 2006 Hematology-Oncology Fellowship Training Program, Washington University
	School of Medicine, USA.
	1999 – 2002 Hematology and Oncology Residency, Pontifical Catholic University of Chile,
	Chile.
	1996 – 1999 Internal Medicine Residency, Pontifical Catholic University of Chile, Chile.
	1989 – 1995 Medicine, Pontifical Catholic University of Chile, Chile.
	Current positions
	2022 – Director, Cancer Prevention and Control Conter (FONDAR 152220002)
	2022 – Director, Cancer Program, Pontifical Catholic University of Chile
	2022 – Brector, Cancer Frogram, Fontinear Catholic University of Chile.
	Chile
	2017 – Medical Director, Cancer Institute, UC-CHRISTUS Health Network.
	2016 – Associate Professor, Pontifical Catholic University of Chile.
	2016 – President of Chilesincáncer Foundation.
Natalie Pinto	Education:
	Nurse, Autonomous University of Chile.
	Current position:
	Nurse, Palliative Care Unit. Dr. Sótero del Río Hospital, South East Metropolitan Health
	Service – Puente Alto.
Carola Sánchez	Education:
	2021 - Industrial Civil Engineering, Faculty of Physical and Mathematical Sciences,
	University of Chile.
	2019 - Semester exchange at FEUP, Faculty of Engineering, University of Porto, Portugal.
	Current Position
	Project Engineer - Digital Health Department - South East Metropolitan Health Service -
	Chincal mormatics Project Links Onit.
luan Carlos Onazo	Education
Juan Carlos Opa20	2019 - Postgraduate Program in Internal Medicine, Pontifical Catholic University of Chile
	2015 - Degree in Medicine and Surgery, Pontifical Catholic University of Chile.
	Current Position:
	-Adjunct Instructor in the Department of Internal Medicine at Pontifical Catholic University
	of Chile.
	- Medical Staff at Sótero del Río Hospital Medicine Service.
	- Medical Staff at Arturo López Pérez Oncology Foundation.
Sebastián Santana	Education:
	2021- today - PhD Student - Engineering Systems, University of Chile.
	- Research topics: Deep Reinforcement Learning for sequential decision making,
	applied to problems of optimal control and operations research



- Visiting Student Research Collaborator (3 months). University of Seville (Spain,
2022)
Supervisor: Prof. Emilio Carrizosa
2018- Ingeniero Civil Industrial, University of Chile.
Current position:
 Head of Advanced Analytics & Co-founder Dic 2019 – Today Deep Metrics Santiago, Chile Our promise is to achieve a deep and meaningful impact on our client's performance with our data and technology-driven solutions.
Lecturer March 2019 - Today
Industrial Engineering, University of Chile Santiago, Chile
• Lecturer of Graduate Diploma in Business Intelligence, Credit Risk, and Digital Marketing & Analytics (MBA).
 I teach Python in computer labs mixing theory, practice, and programming, covering the main aspects of a data science project.
Lecturer June 2019 - Today
U-Negocios — Faculty of Economic and Business, University of Chile Santiago, Chile.
Lecturer in co-created courses for companies with a focus on developing skills for data analysis and programming.