

Machine Learning for Predicting Health Council Decision of Return-to-Work at t Months for Tuberculosis Patients

Yazid Yacouba Hambally^{1*}, Amadou Diabagaté², Hafizatou Sani Yanoussa³, Adama Coulibaly², Abdellah Azmani⁴

¹National High School of Architecture and Urban Planning, University of Bondoukou, Bondoukou, Côte d'Ivoire

²Faculty of Mathematics and Computer Science, University Félix Houphouët-Boigny, Abidjan, Cote d'Ivoire

³Emy Polyclinic, Abidjan, Cote d'Ivoire

⁴Faculty of Science and Technologies, Abdelmalek Essaâdi University, Tétouan, Morocco

Email: *yazid.hambally@gmail.com

How to cite this paper: Hambally, Y.Y., Diabagaté, A., Yanoussa, H.S., Coulibaly, A. and Azmani, A. (2025) Machine Learning for Predicting Health Council Decision of Return-to-Work at t Months for Tuberculosis Patients. *Journal of Computer and Communications*, 13, 160-174.

<https://doi.org/10.4236/jcc.2025.136011>

Received: May 23, 2025

Accepted: June 22, 2025

Published: June 25, 2025

Copyright © 2025 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Predicting the exact duration of sick leave in patients with tuberculosis remains challenging due to the heterogeneity of recovery trajectories. This study uses machine learning to estimate sick leave duration at t months ($t = 6, 9, 12$) by integrating post-treatment radiographic progression and key clinical and sociodemographic factors. This is a retrospective study of tuberculosis patients with documented sick leave duration (2019-2021) presented to the Health Council by the Pulmonary and Phthisiology Department of the Cocody University Hospital. The methodological approach of this study also differs from previous work by the identification of innovative predictive factors, the use of pulmonary sequelae data as a dynamic marker, the provision of individualized predictions that can be updated with new radiographs, and the comparison of the relative impact of variables such as Human Immunodeficiency Virus and type of employment. Also, the personalization of predictions through fine patient stratification and dynamic recommendations adjusting predictions according to treatment progress are major contributions. This approach guarantees a rigorous, clinically relevant, and actionable evaluation for decision-makers and would allow us to assess both the technical performance of the models used and their clinical interpretability, while highlighting the most predictive factors of the duration of work stoppage.

Keywords

Machine Learning, Tuberculosis, Return-to-Work, Health Council, Artificial Intelligence, Decision Support

1. Introduction

Tuberculosis (TB) is an infectious disease transmitted between humans, predominantly through the respiratory tract, caused by a mycobacterium tuberculosis complex [1] [2]. Tuberculosis remains a global public health problem [3] [4] and a major cause of prolonged sick leave, with durations ranging from 1 to 12 months, depending on the health system. Uncertainty surrounding the return to work leads to either early returns or unjustified sick leave. Current criteria are based primarily on subjective clinical assessments, often neglecting the dynamic evolution of radiographic lesions and socio-professional factors.

Each year, according to the World Health Organization (WHO), there are more than 8 million new cases of tuberculosis worldwide, 95% of which occur in developing countries with a high prevalence of Human Immunodeficiency Virus (HIV) infection [5] [6]. In 2018, approximately 10 million new cases of tuberculosis, including half a million cases of rifampicin-resistant tuberculosis (78% of which were multidrug-resistant tuberculosis), were reported by WHO. The disease burden varies from country to country, ranging from less than 5 to more than 500 new cases per 100,000 inhabitants per year, with the global average being approximately 130 new cases [7] [8].

Tuberculosis is a disease unevenly distributed throughout the world. Its incidence remains high in developing countries, including Côte d'Ivoire, despite the existence of antibiotics active against the causative organism and national control programs [9]-[11].

Tuberculosis therefore has a socio-professional and economic impact because it causes prolonged sick leave of six months, which can be renewed depending on the progression of the disease [12]. To limit the transmission and spread of the disease within the population of Côte d'Ivoire, individual and collective prevention measures have been implemented. Among these collective measures is the health counseling provided to civil servants with tuberculosis.

The Health Council is an advisory body to the cabinet of the Ministry of Public Health and the Fight against AIDS. It reviews and provides its opinion on requests submitted by civil servants and government employees with long-term illnesses for long-term sick leave, which is six months, renewable six times. In Côte d'Ivoire, very few studies [13] [14] have been conducted on tuberculosis patients presented to the Health Council and who were granted long-term sick leave. This work contributes to the study of tuberculosis patients receiving health counseling. While traditional models identify isolated risk factors such as HIV and multidrug-resistant TB, they struggle to integrate the complex interactions between comorbidities (diabetes), occupations, and response to treatment, as well as the temporal dimension (length of sick leave) and the heterogeneity of post-treatment radiological trajectories.

In this study, a machine learning solution will be developed to predict the duration of sick leave ($t = 6, 9, 12$ months) by combining seven key dimensions: sociodemographic factors, decision-making times, medical history, and decisions,

and by integrating radiographic changes into the analysis of post-treatment chest X-rays.

The overall objective is to contribute to improving public health systems for managing tuberculosis patients on long-term sick leave while highlighting the key role of the Health Council. Specific objectives include:

- 1) Identifying innovative predictive factors in the care of tuberculosis patients.
- 2) Determining the sectors of activity of tuberculosis patients.
- 3) Predicting the duration of sick leave for tuberculosis patients based on the determinants taken into account by the Health Council.

For the presentation of this work, the functioning of the Health Council in Côte d'Ivoire is highlighted first to demonstrate the institutional anchoring. A literature review is then carried out to position the originality of the study. Similarly, a presentation of the data and methods used for the prediction of the duration of stoppage makes the technical work concrete. Also, the results obtained highlight the actionable knowledge and the section on the discussion will make the link between the technical aspect and the real impact. Finally, the conclusion comes to inspire action.

2. How the Health Council Works in Côte d'Ivoire

The Health Council of Côte d'Ivoire is an advisory body to the Cabinet of the Ministry of Public Health and Population. The Health Council was established by Order No. 248/MSP/CAB on October 6, 1970. It meets ordinarily twice a month and, on an extraordinary basis, whenever necessary, upon the invitation of its President. The Health Council's decisions are final when the meeting is attended by at least two-thirds of its members. However, in the event of an extremely urgent medical evacuation, the decision of three full members of the council will be valid.

The Health Council of Côte d'Ivoire reviews and provides its opinion on requests submitted by civil servants and government employees regarding:

- 1) Sick leave of 15 days up to three months.
- 2) Long-term sick leave of six months, renewable six times.
- 3) Exceptional sick leave (workplace accident, occupational illness, etc.) for up to 60 months, *i.e.* 5 years (maximum duration), beyond which it is considered disability.
- 4) Convalescence leave.
- 5) Changes in administrative position due to illness.
- 6) Reviews of work-related accident files.
- 7) Reviews and opinions on requests for medical evacuations outside Côte d'Ivoire.

Under the Labor Code, any sick worker is required to inform his employer within a maximum of 72 hours or 3 working days from the date of the employee's absence. After this period, the employee is considered to have abandoned his job for up to 3 months. Dismissal procedures for abandonment of job are initiated after the 3 months without justification and the employee's file is sent to the disciplinary board for decisions on dismissal, reinstatement or sanction.

Tuberculosis is a notifiable disease, which is included on the list of long-term illnesses [15]. Long-term sick leave entitles the patient to six months of sick leave, renewable six times depending on the patient's health status [16]. A government employee who is ill is required to inform his organization of his health status within a maximum of 72 hours. Moreover, communication difficulties between the treating physician, the medical advisor, and the occupational physician were highlighted in a survey conducted in Belgium by Vanmeerbeek *et al.* [17] in 2014 on the transmission of information and interprofessional collaboration.

In another study conducted by Vannier [18] in 2017 on administrative procedures, 18.1% of the physicians surveyed did not appear to be aware of the management of long-term tuberculosis. This situation may be due to the fact that this request can be made by the specialist doctor who establishes the final diagnosis and/or begins the treatment and/or ensures the follow-up and monitoring of the patient.

According to the Côte d'Ivoire Labor Code, leaving 72 hours without reporting the reason for absence is considered abandonment of post, and dismissal procedures may be initiated after 3 months without supporting documentation. In the event of this unjustified absence, your organization is required to send you a letter, email, phone call, or face-to-face communication. If you do not receive any response, you will receive a formal notice letter in which the organization requests you to return to work or provide a reason for your absence. If, following this letter, you are able to justify your absence, you will no longer be subject to dismissal for abandonment of post. The Health Council provides a structured clinical framework to assess return-to-work readiness and professional constraints based on objective criteria such as X-rays, residual bacterial load, and occupations at risk of transmission. The Health Council's decisions combine both medical data and socio-professional factors such as type of employment and working conditions.

3. Literature Review

The literature review on predicting the duration of sick leave due to tuberculosis highlights three key areas: duration of sick leave; machine learning and statistical approaches; socioeconomic and occupational studies. More generally, two other areas worth mentioning are clinical rules and survival analysis, which address the duration of sick leave due to tuberculosis differently.

Regarding tuberculosis and the duration of sick leave, very few publications explicitly address this topic. This study [19] aims to predict the duration of anti-tuberculosis treatment in Malaysia, a critical public health issue, using an optimized machine learning approach. This study illustrates how AI can optimize infectious disease management in real-world settings, with direct impacts on health policies. The systematic review [20] explores the long-term consequences of Tuberculosis (TB) on lung function, linking epidemiological data to underlying pathophysiological mechanisms. It highlights the often-overlooked burden of chronic disabilities, calling for a holistic approach to its management.

Regarding machine learning and statistical approaches, only work related to model development was considered. The study [21] proposes an innovative approach for temporal prediction of tuberculosis incidence in Colombia using Artificial Neural Networks (ANNs). It aims to inform predictions of the duration of work stoppages. This article [22] presents machine learning models that allow employers to estimate the duration of work stoppages due to tuberculosis. The objective is to predict the risk of abandoning anti-tuberculosis treatment at different stages of the treatment pathway. Castillo-Chavez and Song [23] present an in-depth analysis of the mathematical models used to study the dynamics of tuberculosis and their applications in public health. This publication [24] explores how Predictive, Preventive, and Personalized Medicine (PPPM) approaches can improve TB management, particularly in the context of increasing antibiotic resistance and individual variability in treatment response. Artificial Intelligence (AI) and machine learning models are used to predict disease progression and treatment response. This study [25] proposes a mathematical model to analyze TB transmission dynamics by integrating the role of exogenous reinfection (new infection after recovery) and optimization strategies to improve control policies.

In terms of socioeconomic and occupational studies, most data come from countries with a high prevalence of tuberculosis. The study [26] aims to develop and validate a predictive score to identify TB patients at high risk of treatment interruption. This model combines clinical, socioeconomic, and behavioral variables with the aim of improving targeted interventions in clinical pharmacy. The study [27] quantifies the economic losses associated with premature deaths from tuberculosis in the WHO African Region (47 countries), focusing on the impact in terms of lost productivity for national economies. The study [28] confirms the urgency of multisectoral interventions to reduce treatment delays in Ethiopia, combining education, infrastructure improvement and universal health coverage.

Our approach differs from traditional methods, which are primarily based on clinical rules and survival analysis. Clinical rules have the advantages of being simple and widely adopted in analyzing the duration of sick leave, but are limited because they are not personalized and ignore the dynamic evolution of tuberculosis [29]-[31]. Survival analysis, on the other hand, identifies risk factors but assumes constant proportional hazards [32] [33].

Our study takes a different approach by using machine learning to achieve high accuracy and integrate complex data, but is limited by the need for large cohorts [19] [28]. No study that combines an African context, socio-professional data, and interpretability for non-experts has been identified on the prediction of the duration of sick leave due to tuberculosis.

4. Data and Methods Used for Prediction

4.1. Presentation of Data Used

The data used in this study come from a study conducted in the Pneumophthisiol-

ogy Department (PPH) of the Cocody University Hospital and on the premises of the Health Council. It took place from January 2017 to December 2019, a period of three years. The study targeted tuberculosis patients on long-term sick leave and collected from the Pneumophthisiology Department of the Cocody University Hospital. The medical records analyzed only concern those relating to tuberculosis patients presented for health counseling by the Pneumophthisiology Department of the Cocody University Hospital.

The following criteria were applied in the selection of tuberculosis patients:

- 1) All tuberculosis patients with medical reports, regardless of age and sex, who presented to the pulmonology department for health counseling were included.
- 2) All incomplete files (absence of health counseling form, absence of follow-up form at the corresponding Anti-Tuberculosis Center) were excluded.

The parameters studied were:

- 1) Sociodemographic characteristics.
- 2) Time between the start of treatment and the date of health counseling.
- 3) Medical history.
- 4) Form and type of disease.
- 5) HIV serology.
- 6) Radiographic appearance after treatment.
- 7) Health council decision after treatment.

All these data were collected from medical reports using a survey form. **Table 1** presents the potential predictive variables organized into four main categories, which are sociodemographic data (sex, age, profession, marital status, number of children), medical history (comorbidities, personal history, family history), clinical characteristics of tuberculosis (form of the disease, type of patient, HIV serology) and variables related to management (treatment-counseling time, post-treatment counseling decision).

Table 1. Variables and response methods used for data collection.

Variable	Response modalities
Sex	Female
	Male
Age	Integer value
Occupation	Health worker
	Customs officer
	Economist
	Teacher
	Military/police officer
	Others
Time between start of treatment and date of advice	Number of days

Continued

Number of children	Integer value
Existence of other antecedents	Yes
	No
Diabetes	Yes
	No
Sickle cell disease	Yes
	No
Marital Status	Single
	Married
	Widower
	Cohabitant
	Divorced
Personal background	High Blood Pressure (HBP)
	Ulcer
	Diabetes
	Sickle cell disease
	Others
Family history	Yes
	No
Form of the disease	Bacteriologically confirmed Pulmonary Tuberculosis (BPT+)
	Clinically diagnosed Pulmonary Tuberculosis (CPT-)
	Extrapulmonary Tuberculosis (EPT)
Patient type	New case
	Relapse
	Multidrug-Resistant Tuberculosis (MDR-TB)
	Resumption
	Failure
HIV serology	Positive
	Negative
	Not done
Health council decision after treatment	Previous activity
	Change of activity
	Others

The critical analysis of this modeling highlights continuous variables (age, number of children, treatment-advice time), which will allow correlation analyses and

categorical variables with high predictive potential (type of patient, form of the disease, decision of the council).

4.2. Data Preprocessing

This section details the data preprocessing steps for input variable. These steps are designed to ensure transparency, reproducibility, and scientific robustness.

1) Data Cleaning.

a) Duplicate records were removed based on key identification variables.

b) Missing values: Binary variables (e.g. Diabetes, Sickle cell disease) imputed with the mode or labeled as “Unknown” if missing > 10%; Continuous variables (e.g. Age, Time between treatment and advice) imputed using the median or KNN imputation; HIV serology “Not done” retained as an informative category.

2) Encoding of Categorical Variables.

a) One-hot encoding for multi-class categorical variables (e.g. Occupation, Marital Status, Disease Form).

b) Binary encoding for boolean-type variables (0 = No, 1 = Yes).

3) Normalization/standardization.

Numerical variables (Age, Number of children, Time to advice) were standardized using z-scores for algorithms sensitive to scale.

4.3. Variable Selection Criteria

This section details the criteria for input variable selection.

1) Expert and Clinical Judgment.

Variables selected based on literature and clinical relevance for tuberculosis and work reintegration.

2) Univariate Analysis.

a) Categorical variables analyzed with chi-square tests.

b) Continuous variables analyzed with ANOVA tests.

c) Variables with $p < 0.10$ were retained for further analysis.

3) Multicollinearity Check.

a) Highly correlated variables ($\rho > 0.85$) were examined, and one was removed to avoid redundancy.

b) Variance Inflation Factor (VIF) was also calculated to control multicollinearity in linear models.

4.4. Methodology

In the specific case of our regression approach aimed at predicting a continuous duration, the following indicators allow us to assess the accuracy of machine learning models in predicting the duration of sick leave [34].

MAE (Mean Absolute Error), which gives the average absolute error (more intuitive than RMSE), provides a concrete idea of the model’s margin of error in months of sick leave and allows us to compare the impact of different factors, such as those between MDR-TB (multidrug-resistant) patients and new cases, or the difference

in error between HIV+ and HIV-:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

where y_i is the actual value and \hat{y}_i the predicted value.

MSE is useful for identifying cases where the model is seriously wrong and has the particularity of heavily penalizing large errors:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

where y_i is the actual value and \hat{y}_i the predicted value.

RMSE (Root Mean Square Error), which measures the average error between predicted and actual values, allows us to evaluate the standard error in months of sick leave:

$$RMSE = \sqrt{MSE} \quad (3)$$

R^2 (Coefficient of Determination), which indicates the proportion of variance explained by the model, is adapted to our multivariate context:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4)$$

where y_i is the actual value, \hat{y}_i is the predicted value and \bar{y} is the average of the actual values.

These indicators make it possible to rigorously evaluate the clinical utility of the model while identifying targeted avenues for improvement.

5. Results

To evaluate the performance of a machine learning model for predicting tuberculosis-related sick leave using specified variables, several metrics and indicators can be used as predictive performance measures (see **Table 2**).

Table 2. Results of the algorithms used.

Model	MAE (%)	MSE (%)	RMSE (%)	R^2 (%)
Linear Regression	75.50	97.49	98.73	-58.40
Artificial Neural Network	77.31	68.59	82.82	-11.45
Random Forest Regressor	28.50	33.97	58.30	44.80
SVM Regressor	44.96	48.81	69.86	20.70
Decision Trees Regressor	20	37.14	60.94	39.65
K-Nearest Neighbors Regressor	29.74	32.23	56.80	47.60

To predict the Health Council's return-to-work decisions for tuberculosis patients, a diverse set of machine learning models were selected. These models were chosen to reflect a balance between interpretability and predictive power, aligning with both the clinical relevance of the task and the heterogeneous nature of the

data. Simple models (e.g. linear regression) provide transparency, while more complex ones (e.g. neural networks, SVM) capture non-linear relationships. The use of multiple model types also helps mitigate selection bias. All models were evaluated using standard regression metrics (MAE, MSE, RMSE, R^2) with k-fold cross-validation to ensure robust and reproducible comparisons. Our selection strategy aimed to provide a fair benchmark across model complexities while supporting practical use in healthcare decision-making. Future work may explore additional models such as gradient boosting, subject to computational and deployment feasibility.

The best model is the Random Forest Regressor [35], which stands out with the second lowest MAE (28.50%), the highest R^2 (coefficient of determination) 44.80% of the variance, and a competitive RMSE (58.30%), indicating moderate error dispersion. The worst model is Linear Regression, which shows unacceptable results with a negative R^2 (-58.40%), which is worse than a simple average; and high MAE/RMSEs that are unsuitable for the non-linearity of medical data.

The analysis of the MAE metric shows that the Decision Trees model [36] has the best performance with an average error of 20%. It is followed by the Random Forest, which presents a good overall compromise with a moderate error (28.50%). Finally, the SVM exhibits mediocre performance, with a risk of clinical underestimation.

The R^2 analysis shows that Random Forest is the most explanatory model with 44.80% of variance explained. However, K-NN, despite slightly better performance of 47.60%, recorded a higher MAE (29.74). The Linear Regression model [34] should be rejected as unsuitable for the complexity of the data, with an R^2 of 58.40%. The RMSE analysis shows that K-NN has low error dispersion (56.80%), Random Forest has stable performance (58.30%), and ANN generates significant errors in certain complex cases.

The models to be retained are therefore Random Forest, which offers a better stability/accuracy compromise, and Decision Trees, which offers optimal performance but risks overfitting. The models to be excluded are SVM [37] [38], which provides mediocre performance with no interpretative advantage, as well as Linear Regression and ANN [39], which generate unacceptable negative R^2 .

To demonstrate the robustness and generalizability of the predictive model, it is strongly recommended to conduct external validation using an independent dataset. This could involve:

- 1) Applying the model to data from another hospital, region, or country with a comparable health system.
- 2) Or retaining a separate portion of the original dataset (not used during training) as a truly independent test set.

Such validation is essential to assess the model's real-world performance, test its out-of-sample generalization capacity, and detect potential risks of overfitting. In the longer term, implementing a multi-site or temporal validation protocol would further reinforce the model's predictive value across diverse healthcare contexts.

6. Discussion

Our study provides several significant advances over existing studies on predicting tuberculosis-related sick leave.

The combination of clinical, social, and temporal dimensions could reveal subgroups at risk of prolonged sick leave. Similarly, few studies quantify the impact of the delay between treatment and counseling because a long delay could indicate systemic dysfunctions, correlated with longer sick leave duration [40] [41].

Patient stratification by taking into account history and post-treatment radiographic appearance allows for the identification of cases requiring prolonged sick leave and patients eligible for early resumption [42]. Furthermore, considering HIV serology as a key modulator is an approach that integrates HIV-positive patients who often have longer sick leave duration, but this variable is rarely cross-referenced with radiographic data or TB type.

Sociodemographic characteristics could guide the targeting of interventions through the use of differentiated protocols and programs [43]. Furthermore, post-treatment counseling allows for analysis of whether medical decisions are consistent with objective data such as radiography, which could reveal biases in practices and thus improve guidelines.

Most existing models are either clinical-biological (radiological scores) or socio-economic, unlike the approach presented in this work, which combines both machine learning and survival analysis to manage complex interactions and predict the probability of resumption at t months. Furthermore, the inclusion of post-treatment radiographic evolution is innovative because it captures therapeutic efficacy.

This study differs from existing studies in several aspects and approaches described in **Table 3**.

Table 3. Difference of aspects and approaches to this study from classical studies.

Aspects	Classical Studies	Approach
Socio-demographic data	Often absent or simplified	Systematic integration (age, profession, etc.)
Processing time advice	Rarely studied	Key variable for monitoring effectiveness
Post-treatment radiography	Used for diagnosis only	Explicitly linked to the duration of work stoppage

Existing analyses are often limited to medical criteria, omitting factors such as age, occupation, or living environment. These omissions can bias recommendations, particularly for manual workers or rural populations. We systematically integrate several sociodemographic variables as predictors. The time elapsed between diagnosis and consultation with the Health Council is rarely documented, even though it influences the perceived severity and the prescribed duration of sick leave. Identifying this time period as a marker of the system's effectiveness.

Chest X-ray is used only to confirm the initial diagnosis, with no established link to the duration of sick leave. We correlate residual lesions visible on imaging with the recommended duration of sick leave.

To enhance the generalizability of the results, it would be relevant to expand the study to a more diverse population, including workers from the private, informal, or rural sectors, as well as other national or subregional contexts. This would help assess the robustness and transferability of the machine learning model across varying socio-economic and healthcare environments. Additionally, a comparative analysis could be conducted by applying the model to datasets from other health systems to identify variables or configurations specific to the Ivorian context. Finally, incorporating a sensitivity or domain adaptation analysis would strengthen the model's applicability as a decision-support tool in broader or different settings.

7. Conclusions

Current guidelines from the WHO and pulmonary societies lack the granularity to recommend durations adapted to residual radiological severity, such as persistent cavitations and stable fibrosis, occupational risk, and interindividual variability in treatment response.

Accurate prediction of the duration of sick leave for tuberculosis is crucial to avoid premature returns or unnecessary prolonged sick leave, to adapt medical follow-ups and radiographic assessments according to risk profiles, and to establish evidence-based recommendations for medical advice.

Including the functioning of the Health Council would enrich this study by clinically validating the model's predictions, identifying institutional or algorithmic biases and opening avenues for more equitable health policies. This would position this work at the interface between explainable AI and collective medical decision-making, offering an innovative perspective.

Among the six algorithms tested, the Random Forest Regressor emerges as the optimal choice, demonstrating its ability to capture non-linear relationships between clinical and socio-professional characteristics. Similarly, the poor performance of linear models highlights the need for non-parametric approaches for this complex clinical problem. This analysis provides a rigorous framework to discuss the strengths/weaknesses of each approach in the specific biomedical context of tuberculosis.

Such an approach would fill a methodological gap by unifying dimensions often treated separately, such as more accurate and personalized prediction of sick leave duration, and policy recommendations based on multidimensional data, such as the optimization of health and professional resources to maximize impact. Potential for prospective validation on multicenter cohorts would be ideal.

Among the areas for improvement, it is worth mentioning the poor performance of linear models, which suggest complex interactions between variables. Hyperparameter tuning to optimize random forests could further improve R^2 .

This study is of interest on several levels. For occupational physicians, a predictive analysis with the variables used would help standardize sick leave durations and reduce inequalities. Similarly, it could impact healthcare economics by reducing costs associated with unjustified sick leave through evidence-based prediction.

One of the innovations of this work lies in its complementarity with the use of radiographs, particularly in contexts where infrastructure is lacking and where the impact of poverty can mask other factors. Traditional methods for assessing sick leave durations for tuberculosis often rely on isolated clinical criteria, neglecting key dimensions such as sociodemographic context or post-treatment follow-up. Our approach systematically integrates these variables for more accurate and personalized prediction.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Toujani, S., Ben Salah, N., Cherif, J., Mjid, M., Ouahchy, Y., Zakhama, H., *et al.* (2015) La primo-infection et la tuberculose pulmonaire. *Revue de Pneumologie Clinique*, **71**, 73-82. <https://doi.org/10.1016/j.pneumo.2015.02.001>
- [2] Glanz, K., Rimer, B.K. and Viswanath, K. (2008) *Health Behavior and Health Education: Theory, Research, and Practice*. John Wiley & Sons.
- [3] Boulahbal, F. and Chaulet, P. (2004) La tuberculose en Afrique épidémiologie et mesures de lutte. *Medecine Tropicale*, **64**, 224-228.
- [4] Nutbeam, D. and Harris, E. (2004) *Theory in a Nutshell: A Practical Guide to Health Promotion Theories*. McGraw-Hill.
- [5] Rapport de l'OMS (2009) *Lutte contre la tuberculose dans le monde—épidémiologie, stratégie, financement*. Principales Constatations.
- [6] World Health Organization (2020) *Global Tuberculosis Report 2020*. WHO.
- [7] World Health Organization (2019) *OMS Global Tuberculosis Report 2019*. WHO.
- [8] UNAIDS (2020) *Global HIV & AIDS Statistics—2020 Fact Sheet*. UNAIDS.
- [9] Stop TB Partnership (2018) *The Global Plan to End TB 2018-2022*. Stop TB Partnership.
- [10] Lönnroth, K., Jaramillo, E., Williams, B.G., Dye, C. and Raviglione, M. (2009) Drivers of Tuberculosis Epidemics: The Role of Risk Factors and Social Determinants. *Social Science & Medicine*, **68**, 2240-2246.
- [11] Courtwright, A. and Turner, A.N. (2010) Tuberculosis and Stigmatization: Pathways and Interventions. *Public Health Reports*, **125**, 34-42.
- [12] OMS—Quartier général et Régions, UICTMR and KNCV (2001) Révisions des définitions internationales pour la lutte contre la tuberculose. *International Journal of Tuberculosis and Lung Disease*, **5**, 213-215.
- [13] Mbe, K. (2016) *Caractéristiques des patients tuberculeux admis en pneumologie pour des conseils de santé*. Thèse de Médecine, Université Félix Houphouët-Boigny.
- [14] Coulibaly, D. (1984) *Les congés de maladie de longue durée des agents de l'Etat: Étude des cas présentés pour tuberculose au conseil de santé de janvier 1980 à décembre 1981*. Thèse de Médecine, Université Félix Houphouët-Boigny.

- [15] Décret n°96-198 du 7 mars 1996 portant sur le droit du travail—Conditions de suspension du contrat, pour maladie du travailleur en Côte d’Ivoire. *Journal Officiel de la République de Côte d’Ivoire*, No. 19, 433–434.
- [16] Arrêté n°248/MSP/CAB du 06 octobre 1970 portant sur la réorganisation du conseil de santé de la Côte d’Ivoire. *Journal Officiel de la République de Côte d’Ivoire*, Ministère de la Santé Publique (Cabinet).
- [17] Vanmeerbeek, M., Govers, P., Schippers, N., Rieppi, S., Mortelmans, K., Donceel, P. and Mairiaux, P. (2014) Les médecins ont-ils vraiment envie de communiquer entre eux? Étude “Partnership in medicine”. Prévention de la désinsertion professionnelle liée à l’incapacité de travail. <https://orbi.uliege.be/handle/2268/163943>
- [18] Vannier, O. (2017) Médecin généraliste et tuberculose: Enquête auprès des médecins généralistes de Côte d’Or sur la prise en charge d’une suspicion de tuberculose pulmonaire. Thèse de Médecine, Université de Bourgogne.
- [19] Balakrishnan, V., Ramanathan, G., Zhou, S. and Wong, C.K. (2023) Optimized Support Vector Regression Predicting Treatment Duration among Tuberculosis Patients in Malaysia. *Multimedia Tools and Applications*, **83**, 11831-11844. <https://doi.org/10.1007/s11042-023-16028-y>
- [20] Ravimohan, S., Kornfeld, H., Weissman, D. and Bisson, G.P. (2018) Tuberculosis and Lung Damage: From Epidemiology to Pathophysiology. *European Respiratory Review*, **27**, Article ID: 170077. <https://doi.org/10.1183/16000617.0077-2017>
- [21] Orjuela-Cañón, A.D., Jutinico, A.L., Duarte González, M.E., Awad García, C.E., Vergara, E. and Palencia, M.A. (2022) Time Series Forecasting for Tuberculosis Incidence Employing Neural Network Models. *Heliyon*, **8**, e09897. <https://doi.org/10.1016/j.heliyon.2022.e09897>
- [22] Chen, J., Jiang, Y., Li, Z., Zhang, M., Liu, L., Li, A., *et al.* (2024) Predictive Machine Learning Models for Anticipating Loss to Follow-Up in Tuberculosis Patients Throughout Anti-TB Treatment Journey. *Scientific Reports*, **14**, Article No. 24685. <https://doi.org/10.1038/s41598-024-74942-z>
- [23] Castillo-Chavez, C. and Song, B. (2004) Dynamical Models of Tuberculosis and Their Applications. *Mathematical Biosciences and Engineering*, **1**, 361-404. <https://doi.org/10.3934/mbe.2004.1.361>
- [24] Dohál, M., Porvazník, I., Solovič, I. and Mokřý, J. (2023) Advancing Tuberculosis Management: The Role of Predictive, Preventive, and Personalized Medicine. *Frontiers in Microbiology*, **14**, Article 1225438. <https://doi.org/10.3389/fmicb.2023.1225438>
- [25] Ochieng, F.O. (2024) Mathematical Modeling of Tuberculosis Transmission Dynamics with Reinfection and Optimal Control. *Engineering Reports*, **7**, e13068. <https://doi.org/10.1002/eng2.13068>
- [26] Oh, A.L., Makmor-Bakry, M., Islahudin, F., Ting, C.Y., Chan, S.K. and Tie, S.T. (2024) Development and Validation of a Predictive Scoring Model for Risk Stratification of Tuberculosis Treatment Interruption. *Research in Social and Administrative Pharmacy*, **20**, 1102-1109. <https://doi.org/10.1016/j.sapharm.2024.08.091>
- [27] Kirigia, J.M. and Muthuri, R.D.K. (2016) Productivity Losses Associated with Tuberculosis Deaths in the World Health Organization African Region. *Infectious Diseases of Poverty*, **5**, Article No. 43. <https://doi.org/10.1186/s40249-016-0138-5>
- [28] Fetensa, G., Wirtu, D., Etana, B., Wakuma, B., Tolossa, T., Gugsu, J., *et al.* (2024) Tuberculosis Treatment Delay and Contributing Factors within Tuberculosis Patients in Ethiopia: A Systematic Review and Meta-Analysis. *Heliyon*, **10**, e28699. <https://doi.org/10.1016/j.heliyon.2024.e28699>
- [29] Tomeny, E.M., Nightingale, R., Chinoko, B., Nikolaidis, G.F., Madan, J.J., Worrall, E., *et al.* (2022) TB Morbidity Estimates Overlook the Contribution of Post-TB Dis-

- bility: Evidence from Urban Malawi. *BMJ Global Health*, **7**, e007643. <https://doi.org/10.1136/bmjgh-2021-007643>
- [30] Menzies, N.A., Quaipe, M., Allwood, B.W., Byrne, A.L., Coussens, A.K., Harries, A.D., *et al.* (2021) Lifetime Burden of Disease Due to Incident Tuberculosis: A Global Re-appraisal Including Post-Tuberculosis Sequelae. *The Lancet Global Health*, **9**, e1679-e1687. [https://doi.org/10.1016/s2214-109x\(21\)00367-3](https://doi.org/10.1016/s2214-109x(21)00367-3)
- [31] van Kampen, S.C., Wanner, A., Edwards, M., Harries, A.D., Kirenga, B.J., Chakaya, J., *et al.* (2018) International Research and Guidelines on Post-Tuberculosis Chronic Lung Disorders: A Systematic Scoping Review. *BMJ Global Health*, **3**, e000745. <https://doi.org/10.1136/bmjgh-2018-000745>
- [32] Jabir, Y.N., Aniley, T.T., Bacha, R.H., Debusho, L.K., Chikako, T.U., Hagan, J.E., *et al.* (2022) Time to Death and Associated Factors among Tuberculosis Patients in South West Ethiopia: Application of Shared Frailty Model. *Diseases*, **10**, Article 51. <https://doi.org/10.3390/diseases10030051>
- [33] Teketelew, G., Medhin, G. and Fenta, T.G. (2022) Survival and Its Predictors among Tuberculosis Patients on Treatment in Selected Health Centers of Addis Ababa, Ethiopia: A Retrospective Cohort Study. *Open Journal of Preventive Medicine*, **12**, 223-238. <https://doi.org/10.4236/ojpm.2022.1210017>
- [34] Breiman, L., Friedman, J.H., Olshen, R.A. and Stone, C.J. (1986) Classification and Regression Trees. Wadsworth and Brooks/Cole.
- [35] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. <https://doi.org/10.1023/a:1010933404324>
- [36] Quinlan, J.R. (1986) Induction of Decision Trees. *Machine Learning*, **1**, 81-106. <https://doi.org/10.1007/bf00116251>
- [37] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. <https://doi.org/10.1007/BF00994018>
- [38] Schölkopf, B. and Smola, A.J. (2001) Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press. <https://doi.org/10.7551/mitpress/4175.001.0001>
- [39] Wu, Y. and Feng, J. (2017) Development and Application of Artificial Neural Network. *Wireless Personal Communications*, **102**, 1645-1656. <https://doi.org/10.1007/s11277-017-5224-x>
- [40] Yasin, P., Yimit, Y., Cai, X., Aimaiti, A., Sheng, W., Mamat, M., *et al.* (2024) Machine Learning-Enabled Prediction of Prolonged Length of Stay in Hospital after Surgery for Tuberculosis Spondylitis Patients with Unbalanced Data: A Novel Approach Using Explainable Artificial Intelligence (XAI). *European Journal of Medical Research*, **29**, Article No. 383. <https://doi.org/10.1186/s40001-024-01988-0>
- [41] Li, Y., Wang, B., Wen, L.M., Li, H.X., He, F., Wu, J., Gao, S. and Hou, D.L. (2023) Machine Learning and Radiomics for the Prediction of Multidrug Resistance in Cavitary Pulmonary Tuberculosis: A Multicentre Study. *European Radiology*, **33**, 391-400. <https://doi.org/10.1007/s00330-022-08997-9>
- [42] Kim, J.W., Bowman, K., Nazareth, J., Lee, J., Woltmann, G., Verma, R., *et al.* (2024) Pet-CT-Guided Characterisation of Progressive, Preclinical Tuberculosis Infection and Its Association with Low-Level Circulating Mycobacterium Tuberculosis DNA in Household Contacts in Leicester, UK: A Prospective Cohort Study. *The Lancet Microbe*, **5**, e119-e130. [https://doi.org/10.1016/s2666-5247\(23\)00289-6](https://doi.org/10.1016/s2666-5247(23)00289-6)
- [43] Ochieng, F.O. (2025) SEIRS Model for TB Transmission Dynamics Incorporating the Environment and Optimal Control. *BMC Infectious Diseases*, **25**, Article No. 490. <https://doi.org/10.1186/s12879-025-10710-2>