IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 14, No. 4, August 2025, pp. 2899~2908
ISSN: 2252-8938, DOI: 10.11591/ijai.v14.i4.pp2899-2908

Prediction of side effects of drug resistant tuberculosis drugs using multi-label random forest

Siti Syahidatul Helma[1,4], Wisnu Ananta Kusuma[1], Mushthofa Mushthofa[1], Diah Handayani[2,3]
[1]Department of Computer Science, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor, Bogor, Indonesia
[2]Department of Pulmonology and Respiratory Medicine, Faculty of Medicine, Universitas Indonesia, Depok, Indonesia
[3]Central Friendship General Hospital, Jakarta, Indonesia
[4]Department of Information Technology, Caltex Riau Polytechnic, Pekanbaru, Indonesia

Article Info

Article history:

Received Jan 5, 2024
Revised Apr 8, 2025
Accepted Jun 8, 2025

Keywords:

Drug side effects
Drug-resistant tuberculosis
Multi-label classification
Tree-based learning
Tuberculosis

ABSTRACT

Drug-resistant tuberculosis (DR-TB) has become a concern because anti-tuberculosis drugs (ATD) used to treat it can cause side effects in patients. This study aimed to predict the potential side effects of ATD using a multi-label classification approach with a random forest (RF) algorithm. This study used 660 medical record data, including the 14 ATD treatments prescribed to the patients and the six side effects experienced by patients. The model was trained using the best parameters based on the hyperparameter tuning process. The results show that the RF multi-label algorithm can be an alternative for building ATD side effect prediction models because it produces the most optimal performance value compared to the decision tree (DT) and extreme gradient boosting (XGBoost). The area under the curve (AUC) score of all RF multi-label models is above 0.8, which means that all RF multi-label models are considered acceptable and applicable for ATD side effect prediction. In addition, eight features influenced the models based on the average feature importance score of the RF models. This study is expected to help predict the side effects of ATD used to treat DR-TB based on ATD treatment and determine the most promising tree-based machine learning algorithm for predicting ATD side effects.

Corresponding Author:

Wisnu Ananta Kusuma
Department of Computer Science, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor
Bogor Regency, West Java 16680, Indonesia
Email: ananta@apps.ipb.ac.id

1. INTRODUCTION

Tuberculosis (TB) has been declared by the World Health Organization (WHO) as one of the deadliest diseases in the world and a threat to public health. It has become a high priority challenging to treat TB, especially drug-resistant tuberculosis (DR-TB) [1]. DR-TB causes problems in terms of tolerance to anti-tuberculosis drugs (ATD), and it was predicted that there were 465,000 DR-TB cases in 2019 [2]. The treatment of DR-TB is complicated because it is necessary to monitor and manage side effects properly, which affects patient adherence [3], [4]. Each individual has various responses in the DR-TB treatment therapy process with ATD. However, treatment cannot be stopped because of reactions from the drugs consumed. During DR-TB treatment therapy, it is necessary to monitor side effects because all ATD can potentially cause various side effects in patients [5].

Regular clinical monitoring, laboratory analysis, and a multidisciplinary approach are essential for the follow-up treatment of DR-TB patients [6]. Computational approaches can be used to monitor the side effects of drugs, including their predictions. This study used DR-TB data from the Indonesian TB

information system (SITB). A previous study by Zhao *et al*. [7] on drug side effect prediction was published using five types of heterogeneous drug information. This study limited the side effect prediction problem to binary classification using random forest (RF) to determine whether a drug and its side effects were associated. If the resulting prediction is positive, the drug is considered to have a side effect and vice versa. Authors in [8], [9] modelled the side effect prediction problem as a multi-label classification task because each drug can potentially have more than one side effect. Multi-label learning provides a potential solution if each sample in the dataset has more than one label [10].

The study by Kouchaki *et al*. [10] on TB drug resistance classification and mutation ranking with multi-label RF produced better performance than single-label RF, where multi-label RF can improve the performance of conventional clinical methods by 18.10% compared to single-label RF, which is only 0.91%. Research by Zhao *et al.* [7], which used a binary classification approach with the RF algorithm to predict drug side effects, did not yet reflect the multiple side effects that the drug might have. Therefore, this study used a multi-label classification approach to build an ATD side effects prediction model using RF [10]. We also used decision tree (DT) and extreme gradient boosting (XGBoost). These algorithms are widely used because they can produce realistic outputs and are easy to interpret and intuitive [11], [12]. This study aimed to predict the side effects of ATD based on an ATD treatment regimen with a multi-label problem transformation approach using the RF algorithm compared with DT and XGBoost. The results of this study are expected to help identify the side effects of DR-TB drugs and determine the most suitable and accurate tree-based machine learning algorithms for predicting ATD side effects.

2. METHOD

The research was divided into several main stages, as Figure 1 depicts. The first stage was the collection of research data that formed the basis of the analysis. The collected data is processed through a preprocessing stage to ensure data quality. The data modeling and hyperparameter tuning stages are performed using three tree-based learning algorithms. The last stage includes calculating and analyzing model performance to obtain optimal results.

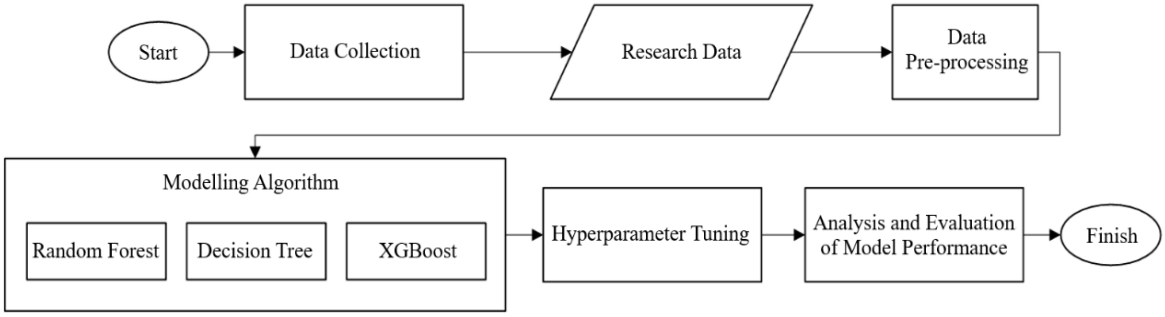


Figure 1. Research methodology

2.1. Research data

This study used secondary data from SITB and the medical records of DR-TB patients from the Persahabatan National Respiratory Referral Hospital in Jakarta from January 2018 to December 2021. This study was approved by the Hospital and Ethics Committee of the Faculty of Medicine, University of Indonesia (ethics number: KET-1326/UN2/F1/ETIK/PPM.00.02/2022). The dataset is divided into two parts: each drug given represent data features consist of clofazimine (Cfz); bedaquiline (Bdq); kanamycin (Km); levofloxacin (Lfx); moxifloxacin (Mfx); linezolid (Lzd); cycloserine (Cs); ethambutol (E); isoniazid (H); ethionamide (Eto); delamanid (Dlm); pyrazinamide (Z); P-aminosalicylic acid (PAS); and also streptomycin (S) and each side effect represent data labels or classes consist of gastrointestinal; neuropsychiatric; cardiovascular; musculoskeletal; anemia; and other side effects.

2.2. Data preprocessing

Data preprocessing aims to form a numerical feature vector to be used as input data for machine learning models [13]. The preprocessing flow is illustrated in Figure 2. The existing data were then cleaned by referring to the results obtained during the exploratory data analysis stage. The data transformation stage

was then performed to obtain numerical values using the one-hot encoding technique, which produced a binary data array for each feature and label. Binary data consisted of a value of 1, indicating that the patient received each drug and had side effects, and a value of 0, indicating that the patient did not receive the drug and had no side effects.

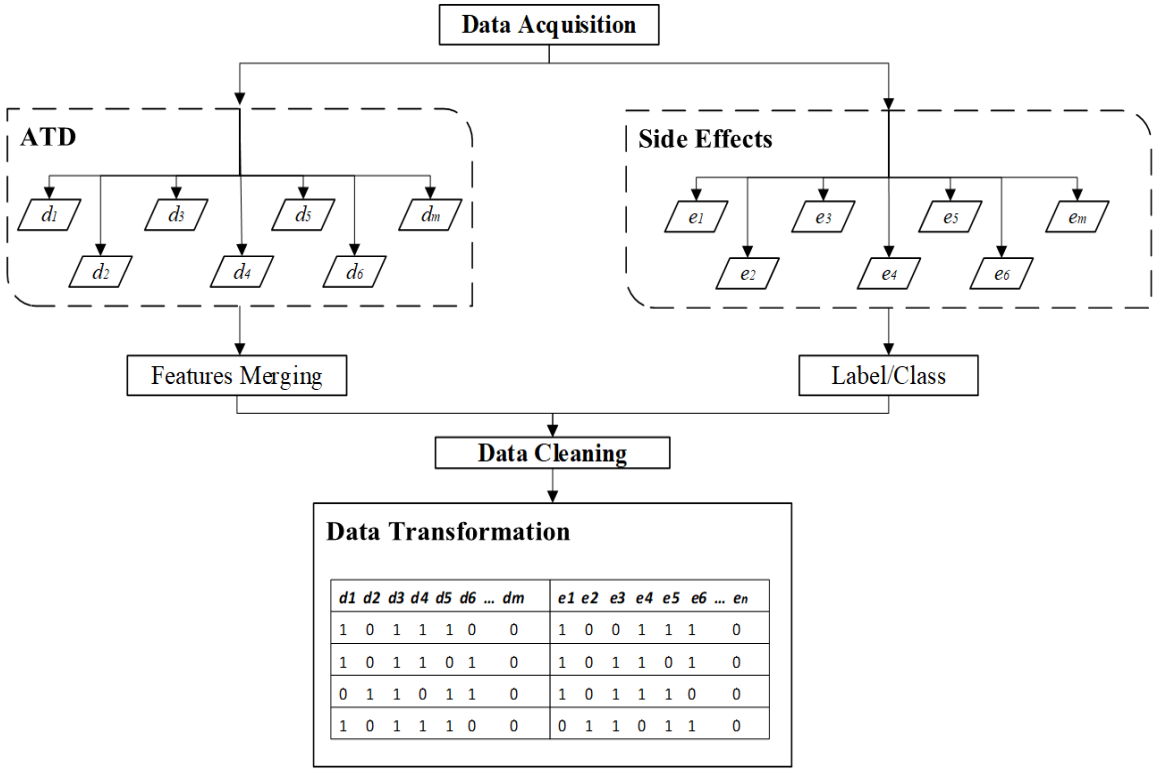


Figure 2. Flow of preprocessing to obtain features and labels

2.3. Multi-label random forest modeling

Multilabel classification is a part of supervised learning that aims to map each dataset into more than one label or class [14]. There are several techniques for multi-label classification using problem transformation methods, namely, classifier chain (CC), binary relevance (BR), and label powerset (LP) [15]. In this study, given a set of m ATD, where $D = \{d_1, \dots, d_m\}$, with d_i, $i = 1, \dots, m$, and there is a set of n side effect labels, where $E = \{e_1, \dots, e_n\}$, with $e_j, j = 1, \dots, n$. Each record in D is associated with one or more side effect labels in L.

RF is an ensemble classification method based on building multiple independent DT classifiers on different subsets of a dataset by considering the combination of each classification output to improve the final prediction performance [16]. The RF model can be extended to study and predict side effects of TB drugs by considering a combined score (Gini index). The process of calculating the Gini index, as shown in (1), is performed by calculating each DT for each pair (f, x) of feature (f) and value x (feature value) with label y (side effects) at node (t) [10].

$$\text{Gini index } GI_J(t, f, x) = \sum_{y \in Y} GI_y(t, f, x) \quad (1)$$

The y variable represents the number of labels or side effects in the analysis. The GI_J and GI_y are the combined and per-label Gini indices, respectively. These two indices play an important role in calculating the importance of features. The calculation is calculated by averaging the impurity reduction associated with each feature.

To improve the final prediction performance, Figure 3 shows that RF combines the output of each selected independent tree using voting or majority techniques. The dataset was divided into training data that were used for training and model validation, and test data that were used to test the model. The training process was conducted using a cross-validation technique with k=5 to evaluate the performance of the model [17].

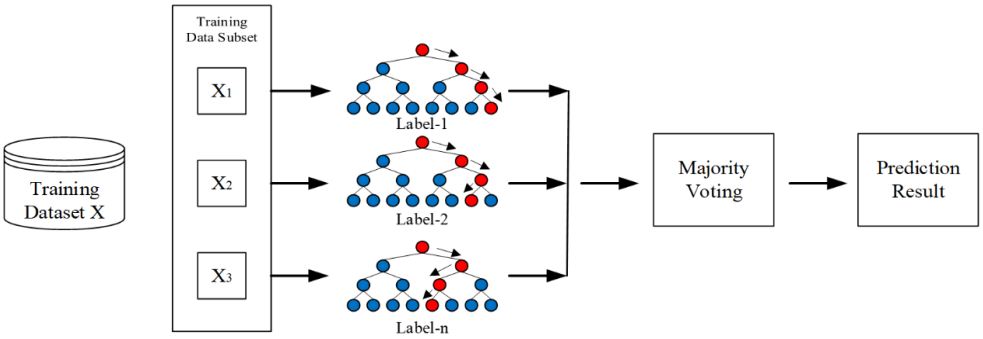


Figure 3. Illustration of the RF model

2.4. Hyperparameter tuning

This study uses a grid search technique to perform hyperparameter tuning. Grid search works by trying all possible combinations based on predefined parameter values. This process aims to select optimal parameters that support maximum model performance [18]. Several parameters are required to build models. A list of tested hyperparameters is provided in https://drive.google.com/file/d/1BhXOypkGJAXr4hIiJXlhLMfJxwwiU2w7/view.

2.5. Model evaluation

The parameters used to calculate model performance were based on accuracy, recall, precision, F-score, Fβ, and Hamming loss values. Accuracy is a metric used to measure the correctness ratio [19]. p is the number of training data to be tested, and Y_i and $h_{(xi)}$are the actual and predicted data labels, respectively. The intersection of Y_i with $h_{(xi)}$ represents the correct predicted label, and the union of Y_i with $h_{(xi)}$ represents the combination of the predicted label and actual label [20] as in (2).

$$Accuracy = \frac{1}{p}\sum_{i=1}^{p}\frac{|Y_I \cap h(x_i)|}{|Y_I \cap h(x_i)|} \quad (2)$$

Recall is the accuracy of the model in predicting positive classes by minimizing mispredicted positive data, and precision is the accuracy of the model in predicting positive classes by minimizing mispredicted negative data, where n is the number of test data to be tested, Y_i, Z_i, and the intersection of Y_i and Z_i are the actual data label, the predicted data label, and the intersection of Y_i with Z_i indicates the label correctly predicted by the model [20]. The recall and precision in (3) and (4), respectively, are as follows:

$$Recall = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_I \cap Z_i|}{|Y_I|} \quad (3)$$

$$Precision = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_I \cap Z_i|}{|Z_I|} \quad (4)$$

As shown in (5), the F1-score is a weighted average comparison of the precision and recall values used to measure the overall minority class performance [21], where L is the label in the multi-label model. The F1-score considers each output label [14]. The performance is good if it exhibits a high average class value [15].

$$F1 - score = \frac{1}{L} \times \frac{2 \times precision \times recall}{precision + recall} \quad (5)$$

F_β is a metric that measures the weighted harmonic mean of recall or precision. The value of β determines the weights of the recall and precision. If the recall value is greater, β>1. Conversely, if the precision value is greater, β<1 [22]. In the case of side effect prediction, minimizing the false-negative value is considered more important than minimizing the false-positive value. Therefore, more weight was given to recall using β=2 to emphasize the importance of fewer false-negative occurrences. The F_β formula is given by (6) [23].

$$F_\beta = \frac{1}{n}\sum_{i=1}^{n}\left[(1+\beta^2)\frac{2|Y_i \cap Z_i|}{\beta^2|Y_i| + |Z_i|}\right] \quad (6)$$

The Hamming loss (7) calculates how many labels should not belong to an instance but are predicted to belong to that instance or vice versa. The fewer the mispredicted labels, the smaller the Hamming loss value, which indicates a better performance of the multi-label learning model [24].

$$Hamming\ Loss = \frac{1}{ml}\sum_{i=1}^{m}\sum_{j=1}^{l}[\![h_{ij} \neq y_{ij}]\!] \quad (7)$$

2.6. Analysis of important features of the model

The overall feature importance is calculated based on the decrease in impurity of the nodes in the model. This decrease is weighted according to the probability of reaching a particular node. The impurity value of the node was then reduced until it reached tree level. The higher the feature score, the greater the feature importance in the RF model [25].

3. RESULTS AND DISCUSSION

3.1. Data preprocessing

We combined two datasets based on SITB and side effect data from medical records, there were 660 subjects eligible for this study, with a mean age of 43 years (CI: 41.791, 44.018), and 368 (55.8%) were male. The 14 drugs were used in combination as a regimen. Each drug was inputted as a feature, and the side effects were classified into six types of side effects.

3.2. Multi-label modeling

Multi-label modeling was applied to the RF, DT, and XGBoost algorithms. The initial modeling uses the default parameters of each algorithm. Then, modeling is applied using the parameters that have been determined. At this stage, modeling uses 80% of the data from the total dataset. The optimal parameters for each multi-label model are listed in Table 1.

Based on the tuning results, we found that each RF multi-label model had different sets of optimal parameters. The difference in parameters lies in the value of n_estimators RF, which is 50 in RF-CC and 25 in RF-BR and RF-LP, respectively. The max_features parameter is sqrt in RF-CC and RF-BR and none in RF-LP. The min_samples_split parameter is 2 for RF-CC and RF-BR and 10 for RF-LP. Another difference is in the max_leaf_nodes parameter, which is three for RF-CC and RF-BR, and six for RF-LP. The only difference in the best parameters in the DT model is the min_samples_leaf parameter, which is 5 in DT-CC and DT-BR and 4 in DT-LP, and the best parameters are the same in all XGBoost models.

Table 1. List of best parameters in RF, DT, and XGBoost multi-label models

Classifier	Hyperparameter	Multi-label model		
		Classifier chain	Binary relevance	Label powerset
RF	n_estimators	50	25	25
	max_depth	none	none	none
	min_samples split	2	2	10
	max_features	sqrt	sqrt	none
	max_leaf_nodes	3	3	6
	class weight	none	none	none
DT	max_features	sqrt	sqrt	sqrt
	max_depth	none	none	none
	min_samples_leaf	5	5	4
XGBoost	max_depth	5	5	5
	min_child_weight	1	1	1
	subsample	0.1	0.1	0.1
	colsample_bytree	1	1	1
	gamma	0	0	0
	learning rate	0.2	0.2	0.2

3.3. Model performance evaluation

The best parameters obtained are applied to each type of multi-label model. This process is to ensure optimal model performance. The results of training the models with the best parameters are compared across the board. Table 2 shows the performance values for each model based on the predefined metrics.

Based on the performance of model training as shown in Table 2, the best accuracy was found in the RF-CC model, and the lowest accuracy was found in the DT-CC model. The precision of all models was similar, demonstrating their ability to produce low false-positives. The DT-BR model had the highest precision, indicating that it had the highest probability of correctly predicting positive side effects (side effects) against adverse side effects (side effects do not occur). Recall indicates the capability to minimize positive labels (side effects) that are incorrectly predicted as negative labels. The best recall was found in the RF-LP model, indicating that it has the highest probability of correctly predicting side effects by measuring the percentage of side effects that occur (true-positives) from all side effects that occur (true positives+false negatives). In comparison, the XGBoost-BR model had the lowest recall, which indicates that it has a reasonably poor ability to predict compared with the other models.

Table 2. Model performance with best parameters

Multi-label	Classifier	Multi-label model performance (%) Accuracy	Precision	Recall	F1 score	F_β	Hamming loss
CC	RF	74.46	72.17	87.64	79.14	84.03	25.53
	DT	72.25	71.39	83.67	76.94	80.81	27.75
	XGBoost	73.61	72.13	85.34	78.12	82.28	26.39
BR	RF	74.40	72.08	87.70	79.11	84.04	25.60
	DT	73.96	73.11	83.77	78.05	81.38	26.04
	XGBoost	72.70	72.32	82.24	76.89	79.99	27.30
LP	RF	74.12	70.89	90.30	79.41	85.60	25.88
	DT	72.95	70.87	87.20	78.08	83.27	27.05
	XGBoost	73.42	71.20	87.35	78.43	83.54	26.58

Based on the performance metrics, the RF model outperformed the other models with the exception of precision. RF had the lowest precision in the RF-LP model but also produced the highest recall among the other models. The F1-score is used to measure the model's ability, which combines precision and recall to overcome false-positives and false-negatives [26]. The best F1-score was obtained using the RF-LP model. This study found the best Hamming loss in RF-CC, indicating that the model correctly predicted each side effect as an actual label [27]. In predicting side effects, low false-negative values are more likely than false-positive values; therefore, the F_β metric uses β=2 to give more weight to recall [28], where the model is concerned with reducing false-negatives rather than false-positive errors. In this study, the best F_β value was also found in the RF-LP model, which means that it can minimize false-negative values.

The best accuracy and Hamming loss were obtained for RF-CC. The best precision was observed for DT-BR, whereas the lowest precision was observed for DT-LP. The best recall, F1-score, and F_β were found in RF-LP. The RF model was the best overall and optimal because it produced the best evaluation value among the models. Further analysis was conducted by evaluating the receiver operating characteristic (ROC) curve for each RF model illustrated in Figure 4. The area under the curve (AUC) values from the ROC curve of each RF model were not significantly different, and all three had scores >0.8. It can be concluded that the overall RF multi-label models are considered good or excellent [29] and can be applied for further ATD side effect prediction.

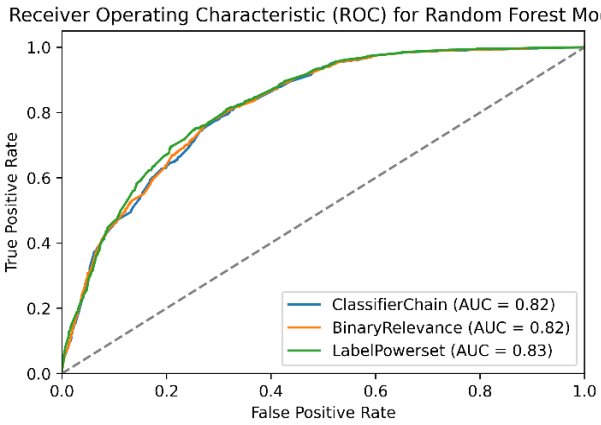


Figure 4. ROC curve of multi-label RF model

3.4. Prediction of ATD side effects with RF model

The RF multi-label model was used to predict The ATD side effects. Predictions were performed on 20% of the test data taken from the total dataset. This test data evaluated the model's ability to recognize ATD side effects. The model prediction performance results were thoroughly analyzed. The prediction performances are presented in Table 3 to illustrate the evaluation results.

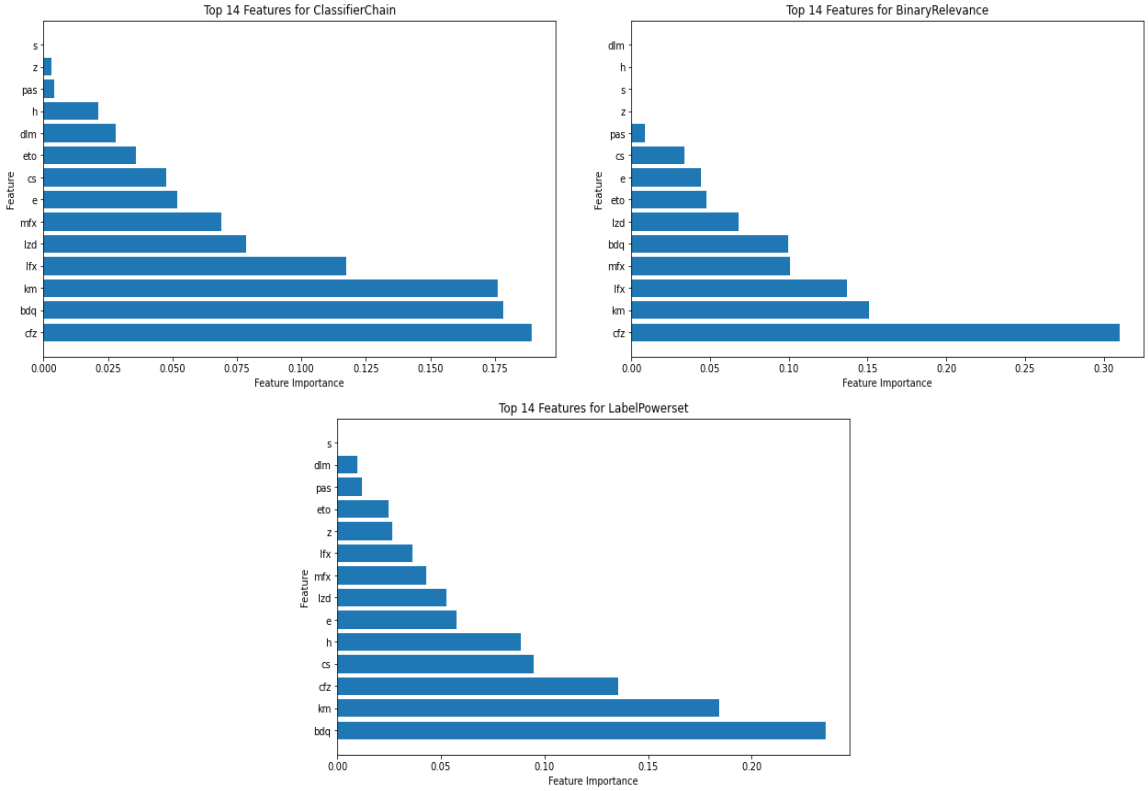
The performance of the ATD side effect prediction results using RF-CC and RF-BR produced the same values for all the metrics. The ATD side effect prediction results of the RF-LP were slightly higher than those of the other RF models for the F1-score, recall, and F_β metrics, and the accuracy, precision, and Hamming loss metrics were slightly lower than those of the other RF multi-label models. The performance results of the ATD side effect prediction in the multi-label RF-LP testing model are similar to those of the previous RF-LP training model, which excels in recall, F1-score, and F_β.

Table 3. Side effect prediction performance with RF multi-label model

Side effect prediction performance with RF multi-label model						
Multi-label	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	F_β (%)	Hamming loss (%)
CC	73.49	71.59	86.30	78.26	82.90	26.52
BR	73.49	71.59	86.30	78.26	82.90	26.52
LP	72.98	70.00	89.50	78.56	84.78	27.02

3.5. Analysis of important features of the RF multi-label model

The essential features of the ATD side effect prediction model were analyzed by applying the feature_importance_ function in the RF sklearn package in Python. In the ATD side effect prediction model, feature analysis can be used to determine the level of importance or influence of features (ATD side effects). Figure 5 shows the feature importance of each RF multi-label model. Based on the average feature importance score of the three RF multi-label models, eight features had the highest level of impact with an average score value greater than or equal to 0.1, namely Cfz, Bdq, and Km with a score of 0.2, as well as Lfx, Mfx, Lzd, Cs, and E with an average score of 0.1. The features that have a low level of impact with an average score below 0.1 are H, Eto, Dlm, Z, PAS, and S.



Description:
Cfz = Clofazimine; Bdq = Bedaquiline; Km = Kanamycin; Lfx = Levofloxacin; Mfx = Moxifloxacin; Lzd = linezolid; Cs = Cycloserine; E = Ethambutol; H = Isoniazid; Eto = Ethionamide; Dlm = Delamanid; Z = Pyrazinamide; Pas = P-Aminosalicylic Acid; and S = Streptomycin

Figure 5. Feature importance of RF multi-label model

Based on feature important, Cfz, Bdq, Km were drugs that have the highest important score in the RF-CC and RF-LP model, and also Lfx in RF-BR model. These findings were consistent with the finding that Cfz causes skin discoloration and gastrointestinal disorders [30], [31]. Bdq is associated with QTc interval prolongation in cardiovascular [32], neurological and gastrointestinal disorders [33]. Another severe side effect is ototoxicity caused by Km, which could lead to hearing loss. In this study, km also had high feature importance; even in the new guidelines, this drug was no longer used [34]. Long-term Lfx treatment has many side effects, including paralysis involving tendons, muscles, joints, nerves, and neuropsychiatric disorders, hepatotoxicity and cardiovascular disorders through QTc prolongation, and phototoxic reactions such as skin redness and severe bullous eruptions [35], [36]. Mfx can cause visual disturbances in the form of uveitis (inflammation of the uveal layer) [37], liver disorders such as acute liver failure (acute liver injury)

[38], and metabolic disorders such as hypoglycemia and hyperglycemia [35]. Since the treatment of DR-TB must consist of four to five drugs, all side effects could occur due to drug-drug interactions (DDI) and can be analyzed well using the RF model.

In this study, we found that Mfx, Lzd, Cs, H, E, Eto, Dlm, Z, PAS, and S were equal to or lower than 0.1, even though many studies found that all these drugs had a risk of some side effects. Therefore, we can consider them an alternative when designing a regimen to reduce side effects [39]. For example, Dlm can cause QTc interval prolongation, anorexia, malaise, gastritis or gastric ulcer, anemia, and psychiatric disorders. However, this study found that Dlm had a low-feature important score. Regarding the efficacy of Dlm, we should consider using this drug as an alternative to avoid severe or multiple side effects of the regimen. Since side effects can also arise due to DDI, we had to consider using these prediction models since the RF algorithm could simulate the drug interaction process. The limitation of this study was that we did not analyze any demographic, clinical, and comorbidity that might influence side effects. Future research endeavors could incorporate a more nuanced examination of these demographic and clinical factors, allowing for a more holistic understanding of the interplay between various variables and their potential impact on side effects.

4. CONCLUSION

Monitoring and early identification of the adverse effects of DR-TB drugs are essential to support the successful treatment of DR-TB. We created ATD side effect prediction models using tree-based learning algorithms, where each given drug represents a feature, and each side effect means a label. RF multi-label algorithms with problem transformation methods are suitable potential models for predicting the side effects of ATD for DR-TB compared with DT and XGBoost, with outperformed performance metrics and high AUC. Based on feature importance, Cfz, Km, Bdq, and Lfx had a higher risk for multiple side effects such as gastrointestinal, neuropsychiatric, cardiovascular, musculoskeletal, and others. Predicting multiple side effects using a multi-label RF algorithm model is essential when designing a treatment regimen for DR-TB for better patient management.

FUNDING INFORMATION

This research was funded by the Directorate of Higher Education, Research, and Technology, Ministry of Education, Culture, Research, and Technology by the contract for the Implementation of the Research Program Year 2023 (No: 18935/IT3.D10/PT.01.03/P/B/2023).

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Siti Syahidatul Helma	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Wisnu Ananta Kusuma	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓
Mushthofa	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	
Diah Handayani	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest with respect to the research, authorship, and publication of this article.

ETHICAL APPROVAL

This study has carefully reviewed and approved by the Ethics Committee of the Faculty of Medicine, University of Indonesia (Acta No: KET-1326/UN2/F1/ETIK/PPM.00.02/2022).

DATA AVAILABILITY

The data supporting this study's findings were obtained from the National Tuberculosis Registry under the authority of Central Friendship General Hospital and were made available exclusively for this research. Any further use of the data requires prior approval from the National Tuberculosis Program, Ministry of Health, Republic of Indonesia.

REFERENCES

[1] WHO, "Global tuberculosis report 2023," *World Health Organization*, 2023. [Online]. Available: https://iris.who.int/bitstream/handle/10665/373828/9789240083851-eng.pdf?sequence=1

[2] Z. Zhuang *et al.*, "Trends and challenges of multi-drug resistance in childhood tuberculosis," *Frontiers in Cellular and Infection Microbiology*, vol. 13, 2023, doi: 10.3389/fcimb.2023.1183590.

[3] R. Singh, S. P. Dwivedi, U. S. Gaharwar, R. Meena, P. Rajamani, and T. Prasad, "Recent updates on drug resistance in Mycobacterium tuberculosis," *Journal of Applied Microbiology*, vol. 128, no. 6, pp. 1547–1567, 2020, doi: 10.1111/jam.14478.

[4] N. Dookie, S. Rambaran, N. Padayatchi, S. Mahomed, and K. Naidoo, "Evolution of drug resistance in Mycobacterium tuberculosis: a review on the molecular determinants of resistance and implications for personalized care," *Journal of Antimicrobial Chemotherapy*, vol. 73, no. 5, pp. 1138–1151, 2018, doi: 10.1093/jac/dkx506.

[5] T. W. Yang *et al.*, "Side effects associated with the treatment of multidrug-resistant tuberculosis at a tuberculosis referral hospital in South Korea," *Medicine*, vol. 96, no. 28, 2017, doi: 10.1097/MD.0000000000007482.

[6] T. Törün *et al.*, "Side effects associated with the treatment of multidrug-resistant tuberculosis," *International Journal of Tuberculosis and Lung Disease*, vol. 9, no. 12, pp. 1373–1377, 2005.

[7] X. Zhao, L. Chen, and J. Lu, "A similarity-based method for prediction of drug side effects with heterogeneous information," *Mathematical Biosciences*, vol. 306, pp. 136–144, 2018, doi: 10.1016/j.mbs.2018.09.010.

[8] W. Zhang, F. Liu, L. Luo, and J. Zhang, "Predicting drug side effects by multi-label learning and ensemble learning," *BMC Bioinformatics*, vol. 16, no. 1, 2015, doi: 10.1186/s12859-015-0774-y.

[9] O. C. Uner, H. I. Kuru, R. G. Cinbis, O. Tastan, and A. E. Cicek, "DeepSide: a deep learning approach for drug side effect prediction," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 20, no. 1, pp. 330–339, 2023, doi: 10.1109/TCBB.2022.3141103.

[10] S. Kouchaki *et al.*, "Multi-label random forest model for tuberculosis drug resistance classification and mutation ranking," *Frontiers in Microbiology*, vol. 11, 2020, doi: 10.3389/fmicb.2020.00667.

[11] M. C. E. Simsekler, N. H. Alhashmi, E. Azar, N. King, R. A. M. A. Luqman, and A. Al Mulla, "Exploring drivers of patient satisfaction using a random forest algorithm," *BMC Medical Informatics and Decision Making*, vol. 21, no. 1, 2021, doi: 10.1186/s12911-021-01519-5.

[12] T. Bartz-Beielstein and S. Markon, "Tuning search algorithms for real-world applications: a regression tree based approach," in *2004 Congress on Evolutionary Computation, CEC2004*, 2004, vol. 1, pp. 1111–1118, doi: 10.1109/cec.2004.1330986.

[13] S. Redkar, S. Mondal, A. Joseph, and K. S. Hareesha, "A machine learning approach for drug-target interaction prediction using wrapper feature selection and class balancing," *Molecular Informatics*, vol. 39, no. 5, 2020, doi: 10.1002/minf.201900062.

[14] X. Wu, Y. Gao, and D. Jiao, "Multi-label classification based on random forest algorithm for non-intrusive load monitoring system," *Processes*, vol. 7, no. 6, 2019, doi: 10.3390/pr7060337.

[15] M. Arslan and C. Cruz, "Imbalanced multi-label classification for business-related text with moderately large label spaces," *arXiv-Computer Science*, pp. 1–9, 2023.

[16] Q. Wu, H. Wang, X. Yan, and X. Liu, "MapReduce-based adaptive random forest algorithm for multi-label classification," *Neural Computing and Applications*, vol. 31, no. 12, pp. 8239–8252, 2019, doi: 10.1007/s00521-018-3900-8.

[17] H. Wang, Y. Ding, J. Tang, Q. Zou, and F. Guo, "Identify RNA-associated subcellular localizations based on multi-label learning using Chou's 5-steps rule," *BMC Genomics*, vol. 22, no. 1, 2021, doi: 10.1186/s12864-020-07347-7.

[18] N. Saini, "Multi-label image classification to detect air traffic controllers' drowsiness using facial features," *M.Sc. Research Project*, School of Computing, National College of Ireland, Dublin, Ireland, 2019.

[19] X. Zhang, H. Zhao, S. Zhang, and R. Li, "A novel deep neural network model for multi-label chronic disease prediction," *Frontiers in Genetics*, vol. 10, 2019, doi: 10.3389/fgene.2019.00351.

[20] M. S. Sorower, "A literature survey on algorithms for multi-label learning," *Research Gate*, pp. 1–25, 2010.

[21] K. B. Lin, W. Weng, R. K. Lai, and P. Lu, "Imbalance data classification algorithm based on SVM and clustering function," in *9th International Conference on Computer Science and Education, ICCCSE 2014*, 2014, pp. 544–548, doi: 10.1109/ICCSE.2014.6926521.

[22] H. Alaiz-Moreton, J. Aveleira-Mata, J. Ondicol-Garcia, A. L. Muñoz-Castañeda, I. García, and C. Benavides, "Multiclass classification procedure for detecting attacks on MQTT-IoT protocol," *Complexity*, vol. 2019, no. 1, 2019, doi: 10.1155/2019/6516253.

[23] T. Üstünkök and M. Karakaya, "SS-MLA: a semisupervised method for multi-label annotation of remotely sensed images," *Journal of Applied Remote Sensing*, vol. 15, no. 3, 2021, doi: 10.1117/1.jrs.15.036509.

[24] J. Duan, X. Li, and D. Mu, "Learning multi labels from single label-an extreme weak label learning algorithm," *Wuhan University Journal of Natural Sciences*, vol. 24, no. 2, pp. 161–168, 2019, doi: 10.1007/s11859-019-1381-y.

[25] H. Alsagri and M. Ykhlef, "Quantifying feature importance for detecting depression using random forest," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 5, pp. 628–635, 2020, doi: 10.14569/IJACSA.2020.0110577.

[26] D. De Clercq, N. F. Diop, D. Jain, B. Tan, and Z. Wen, "Multi-label classification and interactive NLP-based visualization of electric vehicle patent data," *World Patent Information*, vol. 58, 2019, doi: 10.1016/j.wpi.2019.101903.

[27] M. R. Choirulfikri, K. M. Lhaksamana, and S. Al Faraby, "A multi-label classification of Al-Quran verses using ensemble method and naïve Bayes," *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 4, pp. 473–479, 2022, doi: 10.47065/bits.v3i4.1287.

[28] M. A. Wani, N. Agarwal, and P. Bours, "Sexual-predator detection system based on social behavior biometric (SSB) features," *Procedia CIRP*, vol. 189, pp. 116–127, 2021, doi: 10.1016/j.procs.2021.05.075.

[29] A. Istomin, A. Masarwah, R. Vanninen, H. Okuma, and M. Sudah, "Diagnostic performance of the Kaiser score for characterizing lesions on breast MRI with comparison to a multiparametric classification system," *European Journal of Radiology*, vol. 138, 2021, doi: 10.1016/j.ejrad.2021.109659.

[30] J. A. M. Stadler, G. Maartens, G. Meintjes, and S. Wasserman, "Clofazimine for the treatment of tuberculosis," *Frontiers in Pharmacology*, vol. 14, 2023, doi: 10.3389/fphar.2023.1100488.

[31] M. Dalcolmo *et al.*, "Effectiveness and safety of clofazimine in multidrug-resistant tuberculosis: a nationwide report from Brazil," *European Respiratory Journal*, vol. 49, no. 3, 2017, doi: 10.1183/13993003.02445-2016.
[32] S. Esposito, S. Bianchini, and F. Blasi, "Bedaquiline and delamanid in tuberculosis," *Expert Opinion on Pharmacotherapy*, vol. 16, no. 15, pp. 2319–2330, 2015, doi: 10.1517/14656566.2015.1080240.
[33] A. T. Deshkar and P. A. Shirure, "Bedaquiline: a novel diarylquinoline for multidrug-resistant pulmonary tuberculosis," *Cureus*, vol. 14, no. 8, 2022, doi: 10.7759/cureus.28519.
[34] S. K. Heysell *et al.*, "Hearing loss with kanamycin treatment for multidrug-resistant tuberculosis in Bangladesh," *European Respiratory Journal*, vol. 51, no. 3, 2018, doi: 10.1183/13993003.01778-2017.
[35] A. J. Mehlhorn and D. A. Brown, "Safety concerns with fluoroquinolones," *Annals of Pharmacotherapy*, vol. 41, no. 11, pp. 1859–1866, 2007, doi: 10.1345/aph.1K347.
[36] A. Rusu, A.-C. Munteanu, E.-M. Arbănas, and V. Uivarosi, "Overview of side-effects of antibacterial fluoroquinolones: new drugs versus old drugs, a step forward in the safety profile?," *Pharmaceutics*, vol. 15, no. 3, 2023, doi: 10.3390/pharmaceutics15030804.
[37] B. Eadie, M. Etminan, and F. S. Mikelberg, "Risk for uveitis with oral moxifloxacin: a comparative safety study," *JAMA Ophthalmology*, vol. 133, no. 1, pp. 81–84, 2015, doi: 10.1001/jamaophthalmol.2014.3598.
[38] Z. E. Humma and P. Patel, *Moxifloxacin*, Treasure Island, Florida: StatPearls Publishing, 2024.
[39] W. X. Wang, F. Liu, J. Y. Li, J. Xue, Y. T. Li, and R. M. Liu, "A cocrystal for effectively reducing the hepatotoxicity of ethionamide," *Journal of Molecular Structure*, vol. 1243, 2021, doi: 10.1016/j.molstruc.2021.130729.

BIOGRAPHIES OF AUTHORS

Siti Syahidatul Helma holds a Master's degree in Computer Science from IPB University. She also received an undergraduate degree in information systems from the State Islamic University of Sultan Syarif Kasim Riau in 2020. She is currently part of editorial teams at the Indonesian Journal of Artificial Intelligence and Data Mining (IJAIDM). Her research includes machine learning, data mining, and decision support systems. She can be contacted at email: syahidah313@gmail.com or syahidatulhelma@apps.ipb.ac.id.

Wisnu Ananta Kusuma earned his bachelor's and master's degrees from the Bandung Institute of Technology and obtained his Ph.D. from the Tokyo Institute of Technology in 2012. He is an Associate Professor in the Department of Computer Science at IPB University. He serves as the Executive Secretary of the Institute for International Research on Advanced Technology at IPB University. Additionally, he coordinates the Bioinformatics Working Group at the Faculty of Mathematics and Natural Science, IPB University. He leads the Bioinformatics and High-Performance Computing Research Group at the Advanced Research Laboratory, IPB University. His research focuses on machine learning, high-performance computing, and bioinformatics, with over 60 published articles and extensive experience reviewing international journals. He also serves as the Chairperson of the Indonesian Society of Bioinformatics and Biodiversity and holds a position as an ExCo Member of the Asia Pacific Bioinformatics Network (APBioNet). He can be contacted at email: ananta@apps.ipb.ac.id.

Mushthofa holds a Bachelor in Computer Science from IPB University, Indonesia, a Master of Science from Technical University of Vienna, Austria and a Ph.D. in Computer Science from Ghent University, Belgium. He is currently a lecturer at the Department of Computer Science, IPB University, Bogor, where his research interest spans artificial intelligence, machine learning, and bioinformatics. He can be contacted at email: mush@apps.ipb.ac.id.

Diah Handayani received a medical doctorate from the Faculty of Medicine Universitas Indonesia (FMUI). She earned a Pulmonology specialist at FMUI in 2007, a Pulmonary Infection Consultant Collegium Pulmonology in 2015, and a Doctor (Ph.D.) at FMUI in 2021. She is currently a lecturer in the Department of Pulmonology and Respirology Medicine FMUI, Persahabatan Hospital, and Universitas Indonesia Hospital. She is also Head of the education coordination committee of Universitas Indonesia Hospital and secretary of Tuberculosis Research Linked (JejaringRiset TB, JetSet TB). She has authored many articles and copyrights focused on tuberculosis and COVID-19. She can be contacted at email: diahzulfitri@yahoo.com.