

# Factors associated with advanced HIV stage among Thai people with HIV by using Random Forest Algorithm

Chanya Srithawatpong<sup>a,\*</sup>, Sirinya Teeraananchai<sup>a,b</sup>

<sup>a</sup>Master of Biomedical Data Science program, Faculty of Science, Kasetsart University

<sup>b</sup>Department of Statistics, Faculty of Science, Kasetsart University

\*Corresponding author. E-mail address: chanya.srit@ku.th

## INTRODUCTION

Despite better access to antiretroviral therapy (ART), late treatment initiation remains a major hurdle, with many patients starting therapy only after their immune systems are severely compromised. Even with significant resources dedicated to HIV screening and ART provision, Thailand still faces a persistently high incidence of delayed treatment.

## OBJECTIVES

This study applied Random Forest (RF) algorithm to determine factors associated with advanced HIV disease at ART initiation using database from Thai Universal health Coverage program.

## MATERIALS & METHODS

Data Selection: Extracted 302,420 HIV medical records (2014–2024) of people living with HIV (PLHIV) aged 15 and above from the NHSO.

- Data Cleaning: Excluded 23,392 cases missing CD4 counts, leaving 279,028 cases for analysis.
- Train–Test split (70:30)
- RF Model Building using 5–fold cross –validation on training data.
- Optimization using GridsearchCV
- Model Evaluation

## RESULTS

- A total 279,028 PLHIV were included. Among PLHIV with advanced HIV, the majority belonged to male (69.8%), aged 35–49 (43.23%), northeastern (25.9%), asymptomatic HIV (46.7%), 8–30 days (31.12%), 2017–2020 (38.38%), Present of opportunistic infection (15.21%), and NNRTI+NRTI drug group (80.06%) as shown in Table 1.

Table 1 Characteristics of population

Baseline CD4 (cells/mm <sup>3</sup> )	<200 (123,294, 44.19%)	≥200(155,734, 55.81%)	Total (279,028, 100%)
<b>Sex</b>			
Female	37232 (42.31%)	50766 (57.69%)	87998 (31.54%)
Male	86062 (45.05%)	104968 (54.95%)	191030 (68.46%)
<b>Age</b>			
15–19	2539 (19.42%)	10537 (80.58%)	13076 (4.69%)
20–24	11216 (28.91%)	27584 (71.09%)	38800 (13.91%)
25–34	38068 (43.48%)	49477 (56.52%)	87545 (31.37%)
35–49	53298 (51.16%)	50886 (48.84%)	104184 (37.34%)
≥50	18173 (51.3%)	17250 (48.7%)	35423 (12.70%)
<b>Region</b>			
BKK	17038 (32.34%)	35640 (67.66%)	52678 (18.88%)
Central	21014 (44.18%)	26554 (55.82%)	47568 (17.05%)
Eastern	14587 (45.07%)	17778 (54.93%)	32365 (11.60%)
Northeastern	31931 (49.76%)	32233 (50.24%)	64164 (23.00%)
Northern	19133 (44.52%)	23848 (55.48%)	42981 (15.40%)
Southern	14212 (49.3%)	14617 (50.7%)	28829 (10.33%)
Western	5379 (51.51%)	5064 (48.49%)	10443 (3.74%)
<b>HIV Stages</b>			
AIDS	41937 (54.4%)	35147 (45.6%)	77084 (27.63%)
Asymptomatic HIV	57576 (36.37%)	100732 (63.63%)	158308 (56.74%)
Symptomatic HIV	23781 (54.5%)	19855 (45.5%)	43636 (15.64%)
<b>Duration before start ART</b>			
sdART	9803 (31.87%)	20960 (68.13%)	30763 (11.03%)
2–7days	9619 (37.49%)	16039 (62.51%)	25658 (9.20%)
8–30days	38371 (50.92%)	36978 (49.08%)	75349 (27.00%)
>1–3months	35293 (59.68%)	23840 (40.32%)	59133 (21.19%)
>3months	30208 (34.28%)	57917 (65.72%)	88125 (31.58%)
<b>Year starting ART</b>			
2014–2016	39832 (47.38%)	44232 (52.62%)	84064 (30.13%)
2017–2020	47321 (41.41%)	66958 (58.59%)	114279 (40.96%)
2021–2024	36141 (44.79%)	44544 (55.21%)	80685 (28.92%)
<b>Opportunistic infection</b>			
Yes	18758 (78.41%)	5164 (21.59%)	23922 (8.57%)
No	104536 (40.98%)	150570 (59.02%)	255106 (91.43%)
<b>First regimen</b>			
DTG–based	22027 (44.6%)	27359 (55.4%)	49386 (17.70%)
NNRTI+NRTI	98704 (44.98%)	120739 (55.02%)	219443 (78.65%)
PIs+NRTI	2563 (25.13%)	7636 (74.87%)	10199 (3.66%)

## REFERENCES

- Teeraananchai, S., Boettiger, D. C., Lertpiriyasawat, C., Triamwichanon, R., Benjaratnapan, P., & Phanuphak, N. 2025. The impact of same-day and rapid ART initiation under the Universal Health Coverage programme on HIV outcomes in Thailand: a retrospective real-life cohort study. *Journal of the International AIDS Society*, 28(1): e28406. doi.org/10.1002/jia2.28406
- Eansakulrat, P., & Kiertiburanakul, S. 2022. The Impact of Timing of Antiretroviral Therapy Initiation on Retention in Care, Viral Load Suppression and Mortality in People Living with HIV: A Study in a University Hospital in Thailand. *Journal of the International Association of Providers of AIDS Care (JIAPAC)*, 21, 23259582221082607. doi.org/10.1177/23259582221082607
- Ooka, T., Jhono, H., Nakamoto, K., Yoda, Y., Yokomichi, H., & Yamagata, Z. 2021. Random forest approach for determining risk prediction and predictive factors of type 2 diabetes: large-scale health check-up data in Japan. *BMJ Nutrition Prevention & Health*, 4(1), 140–148. doi.org/10.1136/bmjnp-2020-000200

## ACKNOWLEDGEMENT

The author would like to express my deepest gratitude to my principal advisor, Asst. Prof. Dr. Sirinya Teeraananchai for her invaluable guidance, intellectual support, and consistent encouragement throughout the entire process of this research project. This study was approved by the Institutional Review Board of Kasetsart University Research Ethics Committee (KUREC–HSR67/043), Thailand. A waiver of informed consent was granted for this secondary data analysis. All data were de-identified by the National Health Security Office prior to analysis.

- The model achieved an AUC of 0.72 (CI: [0.7207 – 0.7273]) as shown in figure 1. The overall accuracy was 67% and F1–score of 0.62. The model achieved a sensitivity (recall) of 61%, indicating the model’s ability to identify people with advanced HIV stage at ART initiation as shown in Table 2.

Fig. 1 ROC Curve of RF Model

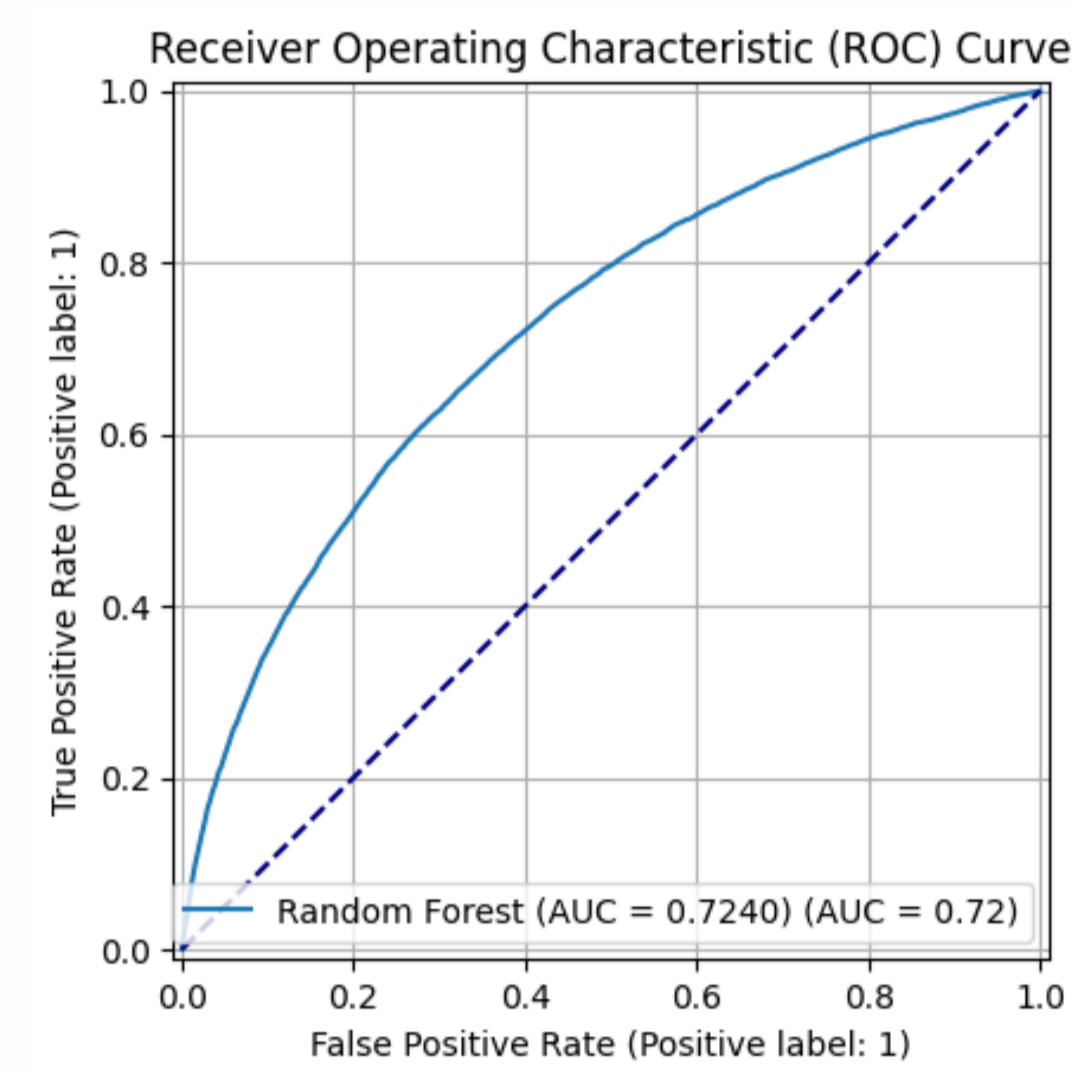
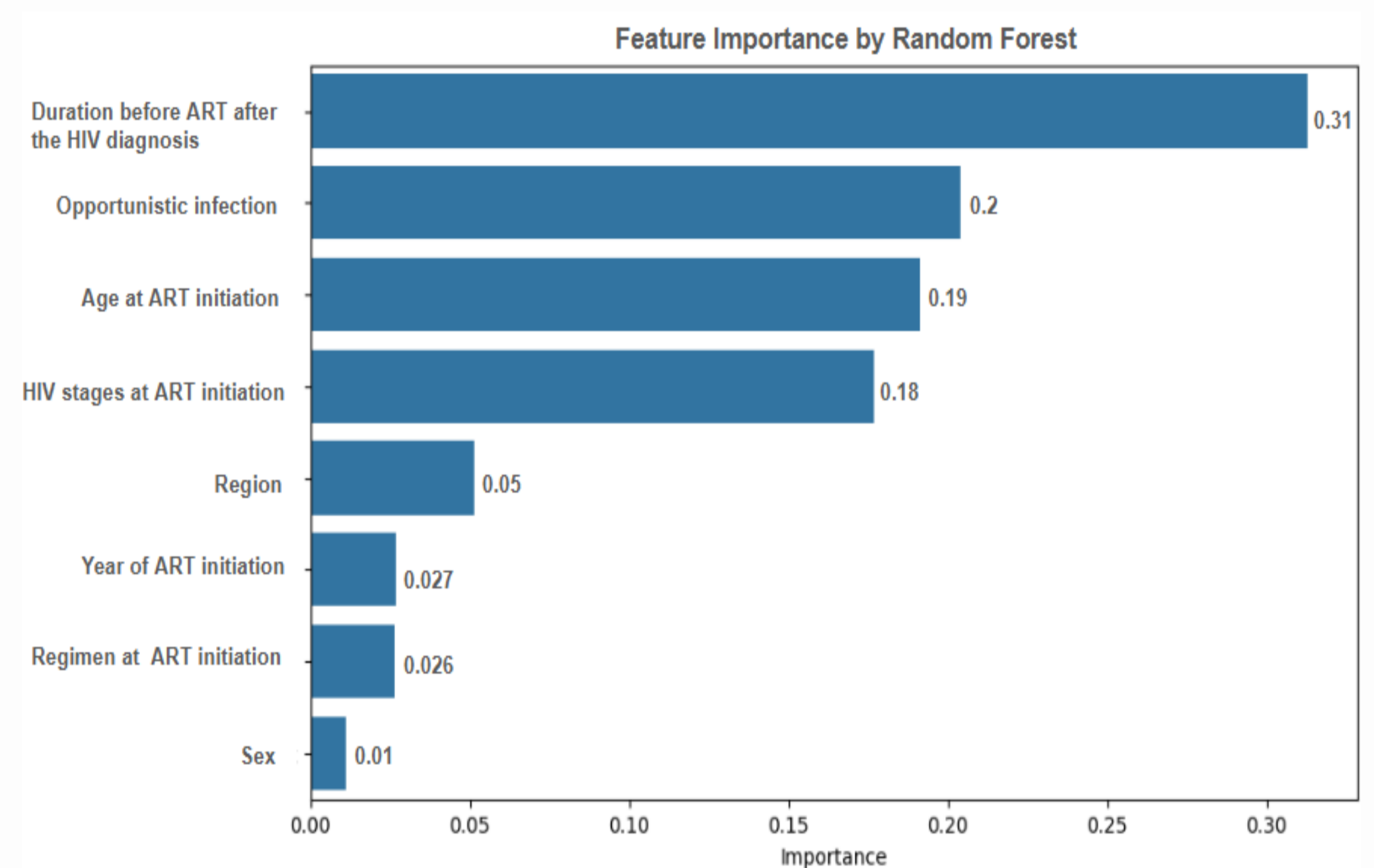


Table 2 Model Performance Metrics

Accuracy	0.6709
Sensitivity	0.6101
Specificity	0.719
Precision	0.63
F1–score	0.621
AUC	0.724

- The duration before starting ART was the most significant factor, followed by a history of opportunistic infections and age at ART initiation as shown in figure 2.

Fig. 2 Feature Importance by Random Forest



- Compared to the reference group (starting ART within 2–7 days), patients who waited 8–30 days had 1.73 times higher odds of advanced HIV, while those waiting 1–3 months faced the highest risk with 2.47 times higher odds as shown in Table 3.
- Initiating ART on the same day as diagnosis showed a lower likelihood of advanced HIV (OR 0.78) compared to the 2–7 days group, supporting the benefit of rapid treatment for better CD4 recovery.
- The ">3 Months" Paradox: Patients starting ART after more than 3 months showed a paradoxical reduction in odds (OR 0.87); however, this may be due to missing socio–behavioral or clinical variables and requires further investigation.

Table 3 Association between ART Initiation Timing and Advanced HIV (CD4 < 200)

Timing of ART Initiation (vs. 2–7 days)	Odds Ratio (OR)	95% Confidence Interval (CI)
8–30 days	1.73	(1.681, 1.781)
>1–3 months	2.468	(2.395, 2.544)
>3 months	0.87	(0.845, 0.895)
Same date of HIV diagnosis	0.78	(0.753, 0.808)

## CONCLUSION

The analysis conclusively identifies the duration before ART initiation and the presence of opportunistic infection as the most critical determinants of low baseline CD4 on PLHIV. The predictive model can serve as a valuable risk stratification tool in clinical practice, allowing healthcare providers, particularly in resource–limited settings, to prioritize high–risk patients for immediate care and aggressive follow–up, thereby optimizing treatment outcomes.