

RAPIDMINER-GENERATED ALGORITHM USING MEAN CORPUSCULAR HEMOGLOBIN & HEMATOCRIT & HEMOLYSIS AREA : A NEW PREDICTION TOOL FOR CARRIERS OF α^0 -THALASSEMIA (SOUTHEAST ASIAN TYPE) AMONG MICROCYTIC BLOOD SAMPLES

Thanusak Tatu^{1,2,*}, Pasinee Pornaram¹, Wonsakorn Samran¹, Santipong Jaisutham², and Jindarat Somwong²

Received: January 16, 2025; Revised: May 23, 2025; Accepted: June 26, 2025

Abstract

The SEA- α^0 thalassemia is common in Thailand. Carriers of this type of thalassemia can be found in microcytic blood samples. Although several laboratory tests are performed to screen the carriers of this type of thalassemia, it is questioned if only RBC parameters can effectively screen the carriers of SEA- α^0 -thalassemia. This study was aimed to evaluate the effectiveness of IC strip, OFT, HbH-IB test, Hemolysis Area and RBC parameters in screening for SEA- α^0 thalassemia carriers in microcytic blood samples. Fifty-four (54) non-anemic blood samples having MCV <80 fL were tested for OFT, HbH-IB test, HA, IC strip test and RBC indices. An algorithm of RBC indices for the SEA- α^0 -thalassemia carrier screening was established by the *RapidMiner* software. The Gap-PCR was carried out to detect the SEA- α^0 thalassemia genotype. A variety of RBC indices was observed. In screening for SEA- α^0 thalassemia, the OFT, HA and IC strip tests had 95.47%, 95.4% and 100% sensitivity, and 53.1%, 50%, and 43.7% specificity, respectively, while the conventional and modified HbH inclusion body tests had 45.4%, 69.1% sensitivity and 100%, 93.7% specificity, respectively. A newly established algorithm [MCH(≤ 23.0 pg) $\rightarrow$ Hct(>36.5 %) $\rightarrow$ HA(≤ 70.0 units)] was found to have 80% sensitivity, 100% specificity, 100% PPV and 92.8% NPV, infinite positive LR and 0.002 negative LR in screening for SEA- α^0 thalassemia. Therefore, OFT, HA, IC strip, but not HbH-IB test, were still effective in screening for carriers of SEA- α^0 -thalassemia in the microcytic blood samples. Due to its ease of use, the newly established algorithm was an additional means to predict the carriers of SEA- α^0 thalassemia in the microcytic cohort.

Keywords: Rapidminer algorithm; Screening tests; SEA- α^0 -thalassemia; Thalassemia carriers; α -thalassemia

¹ Division of Clinical Microscopy, Department of Medical Technology, Faculty of Associated Medical Sciences, Chiang Mai University, Chiang Mai, Thailand. E-mail: tthanu@hotmail.com*; pasinee_p@cmu.ac.th; wongsakorn_samran@cmu.ac.th

² Center of Research/Equipment and Department of Hematology, Clinical Microscopy and Parasitology, Faculty of Medical Technology, Nation University, 444 Village No. 2, Vajiravuth Damnern Road, Prabat sub-district, Mueang district, Lampang province 52000, Thailand. E-mail: tthanu@hotmail.com*; 6317010001@nation.ac.th ; 6317010022@nation.ac.th

* Corresponding author

Introduction

Alpha⁰ thalassemia of Southeast Asian (SEA) type (SEA- α^0 -thalassemia) is caused by deletion of 19.301 kb of the α globin gene cluster spanning Genbank coordinates of 26264 to 45564 and thus removing both *cis* α -globin genes (Harteveld and Higgs, 2010). SEA- α^0 thalassemia is common in Thailand with the prevalence of carriers shown to be 4.30-5.29% in the central part, 4.78-5.39% in the northeastern part, 3.06% in the southern part and the highest of 9.4%-10.6% in the northern part (Tienthavorn *et al.*, 2006).

SEA- α^0 -thalassemia is one of the severe forms of thalassemia. Carriers who are heterozygous for the SEA- α^0 -thalassemia gene can generate offspring suffering from the fatal Hb Bart's hydrops fetalis syndrome; the most severe form of thalassemia in humans. Therefore, screening for this carrier is necessary for the nations having high prevalence of this type of thalassemia.

Screening for carriers of SEA- α^0 -thalassemia at present relies on several laboratory techniques including one-tube osmotic fragility test (OFT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), the recently developed hemolysis area (HA) and immunochromatic (IC) strip test for Hb Bart's detection. All carriers of SEA- α^0 thalassemia have positive results for OFT, MCV <80 fL, MCH <27 pg, HA <52 units, and positive IC strip test (Old, 2003; Sirichotiyakul *et al.*, 2004; 2005; Tayapiwatana *et al.*, 2009; Wanapirak *et al.*, 2011; Prayalaw *et al.*, 2014; Sudjaroen, 2015; Winichagoon *et al.*, 2015; Jomoui *et al.*, 2017; Bunkall *et al.*, 2018; Tatu and Sweatman, 2018; Nelson *et al.*, 2019; Tatu 2020). The blood samples positive for these thalassemia screening tests must be sent further for molecular analysis, such as Gap-PCR.

It is clear that those aforementioned protocols are effective in screening for the carriers of SEA- α^0 -thalassemia in the general population. However, it is not clear whether these protocols are still effective in detecting carriers of SEA- α^0 thalassemia in microcytic blood samples. It is also questioned whether there is an alternative approach to be performed prior to the Gap-PCR for identifying the carriers of SEA- α^0 thalassemia in microcytic blood samples. This study was thus aimed to answer these questions.

During recent years, there were several studies that introduced new platforms for predicting thalassemia in both general population and microcytic blood samples. These studies emphasized the possibility of using artificial intelligence for predicting SEA- α^0 thalassemia in

the cohort of microcytic blood samples (Laengsri *et al.*, 2019; Fu *et al.*, 2021; Feng *et al.*, 2022; Phirom *et al.*, 2022; Rustam *et al.*, 2022; Mo *et al.*, 2023; Saleem *et al.*, 2023; Zhang *et al.*, 2023; Lachover-Roth *et al.*, 2024; Schipper *et al.*, 2024; Saha *et al.*, 2025; Wang *et al.*, 2025).

Materials and Methods

Blood Samples

This experimental research study used 54 EDTA blood samples sent for routine complete blood count in the Clinical Laboratory Department of Maharaj Nakorn Chiang Mai Hospital. All 54 blood samples were non anemic (Hb >10 g/dL), having microcytic erythrocytes (MCV <80 fL). Therefore, the MCV values were not included in the algorithm to be established. The protocol of this study was reviewed and granted approval by the Faculty Research Ethics Committee.

Complete Blood Count (CBC)

CBC tests of the 54 blood samples were performed by an automated blood cell analyzer (Sysmex XN-9000) according to manufacturer instructions.

One-tube Osmotic Fragility Test (OFT)

One-tube osmotic fragility test using 0.36% (w/v) NaCl was performed following the procedure described previously (Tatu, 2020). Briefly, 20 μ L of EDTA blood was mixed with 2 mL of 0.36% (w/v) NaCl solution. The mixture was left at room temperature for 5-10 min before naked-eye visualizing the mixture inside the test tube. "Positive OFT" was reported when a red-turbid mixture was



Figure 1. Result of One-tube Osmotic Fragility Test (OFT) using 0.36 (w/v) NaCl. Negative, Suspicious, and Positive OFT results are shown in Tubes 1, 2, 3, respectively

seen. “Negative OFT” was report if a red-clear mixture was observed. Finally, “Suspicious OFT” was reported if the appearance of the mixture was between postive and negative OFT (Figure 1).

Determination of Hemolysis Area (HA)

Hemolysis Area determination was performed according to a procedure described previously (Tatu and Sweatman, 2018; Tatu, 2020). Briefly, 20 μ L of EDTA blood was mixed with 5 mL of 0.45% Glycerine Saline Solution (GSS) in a 12 $\times$ 75 mm-polystyrene tube, followed by end-over-end mixing, before placing the tube immediately (within 10 sec) in the cuvette socket of the *home-made* portable spectrophotometer. At 120 sec reaction time, the hemolysis graph was generated with the hemolysis area (HA) computation. The cutoff point of HA to discriminate thalassemia trait (α -thalassemia 1 and β -thalassemia traits) and

normal was 52.4 arbitrary unit (AU). Blood samples having HA of less than 52.4 was reported as “Positive HA”, and *vice versa* (Figure 2).

Conventional HbH Inclusion Body Test

Conventional HbH inclusion body test was performed according to the procedure described elsewhere (Chan *et al.*, 1996; Pranpanus *et al.*, 2009). Briefly, 100 μ L of EDTA blood was mixed with 100 μ L of 2% (w/v) brilliant cresyl blue (BCB). The mixture was incubated in a 37°C waterbath for 1 h, followed by making a thin film at the end of incubation. This thin film was examined under a 100x objective lens for the red blood cells containing HbH inclusion body which looked like a golf ball as shown in Figure 3. At least 1 red blood cell containing HbH inclusion body was seen among 10,000 red blood cells for α -thalassemia 1 trait including the SEA- α thalassemia 1 trait.

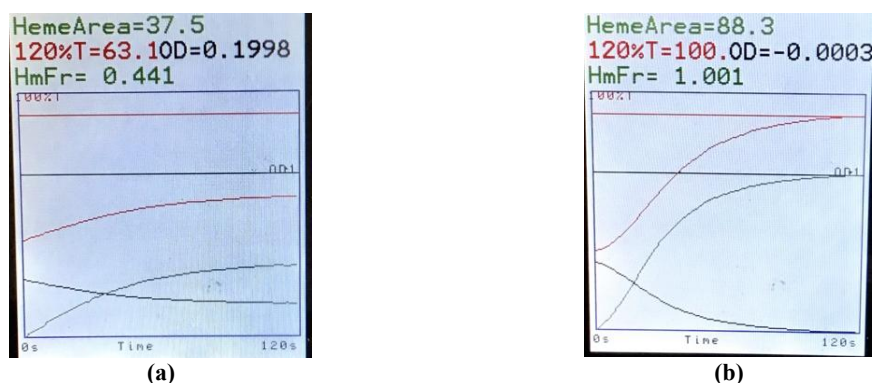


Figure 2. Hemolysis graph demonstrating Hemolysis Area (HA) of SEA- α thalassemia 1 trait with HA 37.5 AU (a) and normal (b) with HA 88.3 AU. “HemeArea” shown in the screen of the machine is the “Hemolysis Area (HA)”

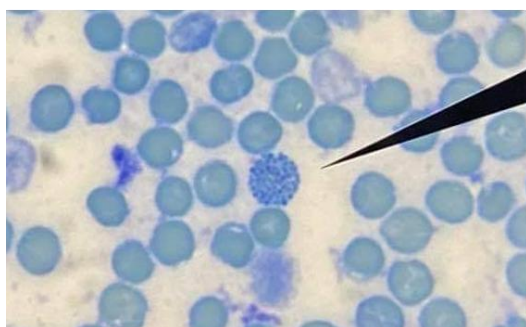


Figure 3. A golf ball-like red blood cell containing HbH inclusion body (pointed) seen in blood samples of the SEA- α^0 thalassemia trait after being incubated with 2% brilliant cresyl blue

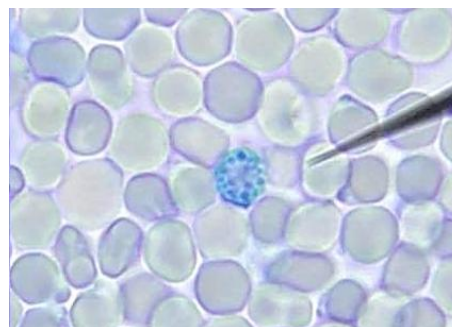


Figure 4. A golf ball-like red blood cell containing HbH inclusion body (pointed) seen in blood samples of the SEA- α^0 thalassemia trait after being incubated with 1% brilliant cresyl blue precoated on a microscopic glass slide

Modified HbH Inclusion Body Test

The modified HbH inclusion body test was performed according to the procedure described previously (Tatu *et al.*, 2003). Briefly, 10 μ L of EDTA blood was mixed with 1% brilliant cresyl blue (BCB) in methanol precoated on a clean microscopy glass slide, followed by pressing a glass coverslip on the blood drop to allow mixing of blood with BCB. The BCB-blood mixture on the precoated glass slide was left at room temperature for 24 h before examining for RBCs containing HbH inclusion body (golf-ball like appearance) under an oil-powered field as shown in Figure 4.

Immunochromatographic Strip Test

Immunochromatographic (IC) strip test for α -thalassemia was carried out following the manufacturer's instructions (*i+Lab α THAL_IC* Strip Test, iMed Laboratory, Bangkok, Thailand) (Tayapiwatana *et al.*, 2009; Wanapirak *et al.*, 2011; Winichagoon *et al.*, 2015; Bunkall *et al.*, 2018). Briefly, EDTA blood was mixed with hemolysis reagent to destroy erythrocyte, resulting in hemoglobin solution. The IC strip was then dipped in the hemoglobin solution to allow hemoglobin solution to move upward to the test and control area. After washing the background out, red bands appeared on the strip. Positive result was determined by the presence of both control and test bands, while only the control band was seen in negative samples (Figure 5).

Identification of SEA- α Thalassemia 1 Genotype

The genotypes of the SEA- α^0 thalassemia were identified through the Gap-PCR established in our laboratory. The PCR was performed in a total volume of 25 μ L containing 500 ng of genomic DNA, 100 μ M of dNTPs, 2.5 units of i-Taq DNA polymerase (Intron), 0.1 μ M of SEA-1 (5'-TGA CTC CAA TAA ATG GAT GAG GA-3') and 0.07 μ M of each primers SEA-2 (5'-GCC TGC GCC GGG GAA CGT AAC CA-3') and SEA-3 (5'-CGC CAA AGA TGG CTA CTC GGA GA-3') in 10 μ M Tris-HCl (pH 8.3), 50 mM KCl, 2% DMSO and 2.5 mM MgCl₂. A total of 40 cycles was carried out with each cycle comprising DNA denaturation at 95°C for 1 min, primer annealing at 60°C for 1 min and primer extension at 72°C for 1 min. The initial denaturation was extended to 5 min while the final extension was prolonged to 7 min. A 757-bp amplified product was produced in the presence of the α -thalassemia 1 of SEA deletion, whereas a 652-bp fragment was obtained with those negative for SEA deletion (Figure 6).

Establishment of Prediction Algorithm

A predictive algorithm was established with the Decision Tree mode of the online-software named *RapidMiner*, following the procedure described elsewhere (Tatu *et al.*, 2025). Red blood cell (RBC) indices including RBC count (M/ μ L), Hemoglobin concentration (Hb; g/dL), Hematocrit (Hct; %), Mean Corpuscular Hemoglobin (MCH; pg), Red Cell Distribution Width (RDW; %) and Hemolysis Area (HA; Unit) of the SEA- α thalassemia 1 and non SEA- α thalassemia 1, genotyped by the Gap-PCR technique, were analyzed to establish the prediction algorithm. After the algorithm was set up, the algorithm was assessed using another cohort of microcytic samples.

Statistical Analysis

Descriptive statistics including mean, standard deviation were calculated for all parameters. Differences of parameters among thalassemia groups were analyzed by the Kruskal-Wallis test with *p*-value of less than 0.05 considered significant. All the statistical analyses were performed using the statistical software (SPSS Statistics 16.0 for Windows, SPSS Inc, Chicago).

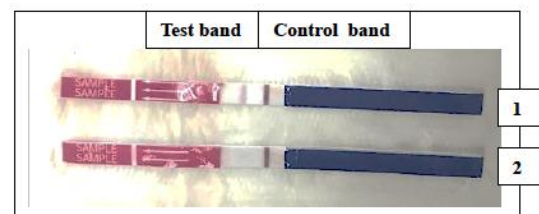


Figure 5. Results of *i+Lab α THAL_IC* Strip Test in which sample 1 had positive result and sample 2 had negative result

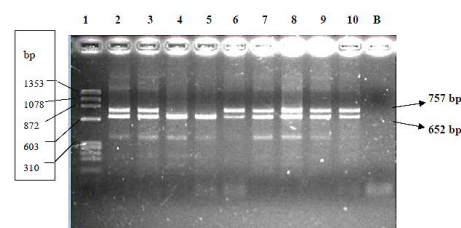


Figure 6. Results of Gap-PCR for identifying SEA- α^0 thalassemia genotype. Lanes 2,3, 6-10 were positive for SEA- α thalassemia 1. Lanes 4, 5 were negative for SEA- α thalassemia 1. Lane B was blank in which DW was used instead of DNA. Lane 1 was Φ X174/ *Hae* III-fragment size standard

Results

Hematologic Parameters and Prevalence of SEA- α Thalassemia 1 among the Studied Microcytic Blood Samples

Among 54 microcytic blood samples analyzed in this study, all were not anemic with Hb concentration and Hct levels of 11.3 -16.0 g/dL (12.8 ± 0.9) and 34.6-50.7%, respectively. Based on phenotypic and genotypic analysis, 11 types of thalassemia were found in these microcytic blood samples. These included SEA- α^0 thalassemia carriers, HbE carriers, homozygous HbE, $\beta^{41/42}$ thalassemia carriers, α -thalassemia carriers (non SEA type), double carriers of α -thalassemia (non SEA type) and $\beta^{41/42}$ thalassemia, combined α -thalassemia carrier (non SEA type) and homozygous HbE, double carriers of HbE and α -thalassemia (non SEA type), double carriers of $\beta^{41/42}$ thalassemia and SEA- α^0 thalassemia, double carriers of β^{17} thalassemia and α -thalassemia (non SEA type), and double carriers of HbE and SEA- α^0 thalassemia. Prevalence of carriers of the SEA- α^0 thalassemia in this cohort was 22 in 54 (40.7%). Those samples positive for HbH inclusion body tests and/or Alpha IC strip test but negative of SEA- α^0 thalassemia genotyping were classified as α -thalassemia (non SEA- α thalassemia type). Heterogeneity of hematological parameters was observed (Table 1). Significant differences (p -values < 0.05) of RBC, MCV, MCH, RDW and

HA values were shown among these 11 types of thalassemia in these microcytic blood samples.

Assessment of Laboratory Test in Screening for SEA- α Thalassemia 1 Carriers

The power of OFT, HA, IC strip test, conventional HbH inclusion body test, and modified HbH inclusion body test in screening for the SEA- α^0 thalassemia carriers were assessed against the SEA- α^0 thalassemia genotype determined by the Gap-PCR. Table 2 shows numbers of samples having true positive (TP), true negative (TN), falsely positive (FP), and falsely negative (FN) results for each test. It was found that OFT, HA and IC strip test had high sensitivity, but quite low specificity in screening for the SEA- α^0 thalassemia carriers in these microcytic blood samples. In contrast, both HbH inclusion body tests had low sensitivity, with high specificity in screening for the SEA- α^0 thalassemia in this cohort.

Establishment of Predictive Algorithm for SEA- α Thalassemia 1 Carriers

RBC parameters including RBC count, Hb, Hct, MCH, RDW, and HA were tested whether they could be used to establish the algorithm for prediction of the SEA- α^0 thalassemia carriers in these microcytic samples. The *RapidMiner* analysis showed that MCH was the strongest predicting parameter. Four algorithms were established for predicting the SEA- α^0 thalassemia carriers (Figure 7). Assessment of these four algorithms

Table 1. Hematological parameters and Hemolysis Area (mean $\pm$ SD) in 11 types of thalassemia existing in microcytic blood samples analyzed in this study. (S: Significant difference, NS: Non-significant difference)

Genotypes	N (%)	RBC (M/ μ L)	Hb (g/dL)	Hct (%)	MCV (fL)	MCH (pg)	MCHC (g/dL)	RDW (%)	HA (U)
--SEA/ $\alpha\alpha$	18 (33.3)	6.23 $\pm$ 0.67	12.9 $\pm$ 1.1	41.1 $\pm$ 3.5	66.2 $\pm$ 3.5	20.8 $\pm$ 1.2	31.5 $\pm$ 0.8	17.5 $\pm$ 2.1	30.5 $\pm$ 14.3
β^A/β^E	10 (18.5)	5.67 $\pm$ 0.51	13.2 $\pm$ 1.2	40.6 $\pm$ 3.0	71.9 $\pm$ 4.5	23.2 $\pm$ 1.7	32.4 $\pm$ 0.6	15.6 $\pm$ 1.8	48.0 $\pm$ 20.9
β^E/β^E	1 (1.8)	5.68	11.9	34.6	60.9	21.0	34.4	15.5	12.4
$\beta^A/\beta^{41/42}$	3 (5.5)	5.55 $\pm$ 0.31	12.3 $\pm$ 0.3	36.5 $\pm$ 0.9	65.8 $\pm$ 2.9	22.2 $\pm$ 1.7	33.7 $\pm$ 1.6	16.0 $\pm$ 2.1	22.2 $\pm$ 6.0
--/ $\alpha\alpha$	13 (24.1)	5.54 $\pm$ 0.32	12.6 $\pm$ 0.9	39.4 $\pm$ 2.5	71.3 $\pm$ 4.1	22.8 $\pm$ 1.8	31.9 $\pm$ 1.2	16.2 $\pm$ 1.9	58.8 $\pm$ 24.7
--/ $\alpha\alpha$	1 (1.8)	6.90	13.6	43.5	63.0	19.7	31.3	19.7	27.3
$\beta^A/\beta^{41/42}$	1 (1.8)	5.78	12.2	36.3	62.8	21.1	33.6	16.8	21.1
β^E/β^E	2 (3.7)	4.93 $\pm$ 0.64	13.1 $\pm$ 0.3	39.1 $\pm$ 2.5	76.5 $\pm$ 0.8	25.7 $\pm$ 1.4	33.5 $\pm$ 1.4	13.5 $\pm$ 0.4	84.8 $\pm$ 4.0
--SEA/ $\alpha\alpha$	1 (1.8)	5.42	11.9	37.1	68.5	22.0	32.1	18.3	43.0
$\beta^A/\beta^{41/42}$	1 (1.8)	6.69	12.9	40.7	60.8	19.3	31.7	18.7	13.6
--SEA/ $\alpha\alpha$	3 (5.5)	6.91 $\pm$ 0.56	13.9 $\pm$ 0.9	44.6 $\pm$ 2.9	64.6 $\pm$ 1.0	20.2 $\pm$ 0.5	31.2 $\pm$ 0.7	18.9 $\pm$ 0.3	18.6 $\pm$ 7.3
β^A/β^E									
p -value		0.004 (S)	0.229 (NS)	0.056 (NS)	0.002 (S)	0.005 (S)	0.071 (NS)	0.026 (S)	0.005 (S)

Table 2. Assessment of screening tests for SEA- α^0 thalassemia carriers in microcytic blood samples. True positive, True negative, False positive and False negative results are shown as well as sensitivity (Sens.) and specificity (Spec.) of the screening tests

		SEA- α thalassemia 1			Sens.	Spec.
		Pos.	Neg.	Total	(%)	(%)
OFT	Pos.	21	15	36	95.4	53.1
	Neg.	1	17	18		
HA	Pos.	21	16	37	95.4	50.0
	Neg.	1	16	17		
IC Strip Test TM	Pos.	22	18	40	100.0	43.7
	Neg.	0	14	14		
Conv. HbH-IB	Pos.	10	0	10	45.4	100.0
	Neg.	12	32	44		
Mod. HbH-IB	Pos.	13	2	15	69.1	93.7
	Neg.	9	30	39		

by the command “*Apply Model*” showed that they were able to predict 19 in 22 (86.4%) SEA- α^0 thalassemia carriers in these microcytic blood samples. The shortest algorithm was modified for easy use and turned out to be $MCH(\leq 23.0) \rightarrow Hct(>36.5) \rightarrow HA(\leq 70.0)$. Assessment of this modified algorithm in another cohort of 18 microcytic blood samples, with $Hb > 10.0$ g/dL and $MCV < 80$ fL, showed that this modified algorithm had 80% sensitivity, 100% specificity, 100% positive predictive value and 92.8% negative predictive value, infinite positive likelihood ratio and 0.002 negative likelihood ratio for predicting the SEA- α^0 thalassemia carriers.

Discussions

Screening for the SEA- α^0 thalassemia carriers is necessary for a thalassemia-rich population to prevent birth of the fatal Hb Bart’s hydrops fetalis babies. Carriers of this type of thalassemia are common in Thailand. All of the carriers of SEA- α^0 thalassemia are characterized by microcytosis (Harteveld and Higgs, 2010). The microcytic blood samples analyzed in this report harbored 40.7% of carriers of the SEA- α^0 thalassemia. This information highlighted the prioritization of screening for SEA- α^0 thalassemia in microcytic blood samples of northern Thailand origin. Although high prevalence of SEA- α^0 thalassemia in Thailand has been documented recently by Hockham C and colleagues (Hockham *et al.*, 2019), that information originated from a population-based survey. This report, however, appears to be the first evidence of the prevalence of SEA- α^0 thalassemia carriers among microcytic blood samples in Thailand. Therefore, identifying carriers of SEA- α^0 thalassemia must be prioritized in microcytic blood

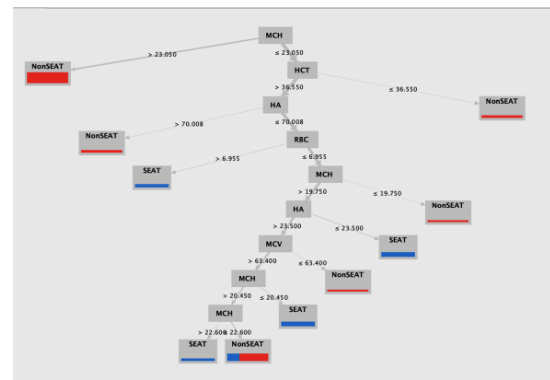


Figure 7. The *RapidMiner* algorithm established for predicting SEA- α^0 thalassemia in the microcytic blood samples. SEAT stands for Southeast Asian Trait which represents SEA- α^0 thalassemia carriers (Blue box). NonSEAT stands for non-Southeast Asian Trait, which represents nonSEA- α^0 thalassemia carriers (Red box). Possible working algorithms are marked by squares at the SEAT boxes. [RBC: RBC count, Hb: Hemoglobin, Hct: Hematocrit, MCV: Mean Corpuscular Volume, MCH: Mean Corpuscular Hemoglobin, HA: Hemolysis Area]

samples in northern Thailand and southeast Asian countries.

Conventional screening laboratory tests performed to screen carriers of SEA- α^0 thalassemia in these microcytic blood samples showed the efficiency resembling that found previously for population-based screening. This finding was not surprising as these tests worked under the same principles. Therefore, they should be identical in the determined efficiencies.

It was clear that RBC parameters in this microcytic cohort varied widely, depending on

types of thalassemia. Therefore, a predicting algorithm was established by using the free software so-called *RapidMiner*. *RapidMiner*-based predictions have been performed in several disciplines in medical sciences (Amirhajlou *et al.*, 2019; Martins *et al.*, 2021; CheshmehSohrabi *et al.*, 2023; Mir *et al.*, 2024). Here, *Decision Tree* option of *RapidMiner* was performed and a suitable working algorithm was established. The newly established algorithm was found to be as effective as the conventional screening tests. The algorithm required Hct, MCH and HA values that were routinely and easily analyzed. This approach should ease the screening for carriers of SEA- α^0 thalassemia in microcytic blood samples. In particular, including

HA in this working algorithm increased specificity in screening for SEA- α^0 thalassemia compared to the HA alone; 50% by HA alone to 100% by the new algorithm. Moreover, compared to the conventional tests; OFT, *IC Strip Test*TM, this new algorithm also had higher specificity. This high specificity would reduce the need of performing confirmatory tests in no-risk cases. More importantly, considering this algorithm in combination with other screening tests would increase the ability to screen for the carrier of SEA- α thalassemia 1. To our knowledge, this is also the first time that this algorithm has been established for predicting SEA- α thalassemia 1 carriers from microcytic blood samples. All methodologies involved in this algorithm are based on numerical data which ensured no human bias. Therefore, this algorithm can be performed accurately in all thalassemia-screening laboratories.

As stated in the previous part, this study compared the efficiency of the *RapidMiner*-based algorithm with that of the conventional screening tests for thalassemia carriers. Comparable efficiency was found which made this algorithm applicable in screening and predicting carriers of SEA- α thalassemia 1. In fact, the data of the SEA- α thalassemia 1 and non SEA- α thalassemia 1 analyzed in this study were not equally distributed. This could lead to unreliability of the decision-tree based algorithm established in this study. Therefore, establishment and assessment of other predicting algorithms using other published and documented machine-learning tools are planned. This would strengthen the effectiveness of using RBC indices in predicting the SEA- α thalassemia 1 carriers.

Conclusions

In conclusion, for microcytic and positive OFT blood samples, the algorithm of $MCH(\leq 23.0)$ -

$>Hct(>36.5)$ - $>HA(\leq 70.0)$ was effective in predicting for carriers of SEA- α^0 thalassemia. This algorithm could be an alternative way to screen for these carriers in laboratories having limitation in performing molecular analysis of SEA- α thalassemia 1. Further analysis with increased number of samples would strengthen this conclusion.

Acknowledgement

The authors thank Faculty of Associated Medical Sciences, Chiang Mai University and Faculty of Medical Technology, Naiton University for supporting materials used in this study. PP, WS, SJ and JS were students of Medical Technology curriculum. The authors also thank Dr. Denis Sweatman for proofreading English usage in this article.

Conflict of Interest

None to declare

References

- Amirhajlou, L., Sohrabi, Z., Alebouyeh, M. R., Tavakoli, N., Haghighi, R. Z., Hashemi, A., and Asodeh, A. (2019). Application of data mining techniques for predicting residents' performance on pre-board examinations: A case study. *Journal of Education and Health Promotion*, 8:108. https://doi.org/10.4103/jehp.jehp_394_18
- Bunkall, C., Ghallyan, N., Elliott, C., Van de Water, N., and Chan, G. (2018). Evaluation of an immunochromatographic strip test for α -thalassaemia screening. *International Journal of Laboratory Hematology*, 40(6):691-696. <https://doi.org/10.1111/ijlh.12905>
- Chan, A.Y., So, C.K., and Chan, L.C. (1996). Comparison of the HbH inclusion test and a PCR test in routine screening for α -thalassaemia in Hong Kong. *Journal of Clinical Pathology*, 49(5):411-412. <https://doi.org/10.1136/jcp.49.5.411>
- CheshmehSohrabi, M. and Mashhadi, A. (2023). Using data mining, text mining, and bibliometric techniques to the research trends and gaps in the field of language and linguistics. *Journal of Psycholinguistic Research*, 52(2):607-630. <https://doi.org/10.1007/s10936-022-09911-6>
- Feng, P., Li, Y., Liao, Z., Yao, Z., Lin, W., Xie, S., Hu, B., Huang, C., Liu, W., Xu, H., Liu, M., and Gan, W. (2022). An online α -thalassemia carrier discrimination model based on random forest and red blood cell parameters for low HbA₂ cases. *Clinica Chimica Acta*, 525:1-5. <https://doi.org/10.1016/j.cca.2021.12.003>
- Fu, Y.K., Liu, H.M., Lee, L.H., Chen, Y.J., Chien, S.H., Lin, J.S., Chen, W.C., Cheng, M.H., Lin, P.H., Lai, J.Y., Chen, C.M., and Liu, C.Y. (2021). The TVGH-NYCU Thal-Classifier: Development of a machine-learning classifier for differentiating thalassemia and non-thalassemia patients. *Diagnostics*, 11(9):1725. <https://doi.org/10.3390/diagnostics11091725>
- Hartevelde, C.L. and Higgs, D.R. (2010). α -thalassaemia. *Orphanet Journal of Rare Diseases*, 5:13. <https://doi.org/10.1186/1750-1172-5-13>

- Hockham, C., Ekwattanakit, S., Bhatt, S., Penman, B.S., Gupta, S., Viprakasit, V., and Piel, F.B. (2019). Estimating the burden of α -thalassaemia in Thailand using a comprehensive prevalence database for Southeast Asia. *eLife*, 8:e40580. <https://doi.org/10.7554/eLife.40580>
- Jomoui, W., Fucharoen, G., Sanchaisuriya, K., and Fucharoen, S. (2017). Screening of ($-\text{SEA}$) α -thalassaemia using an immunochromatographic strip assay for the ζ -globin chain in a population with a high prevalence and heterogeneity of haemoglobinopathies. *Journal of Clinical Pathology*, 70(1):63-68. <https://doi.org/10.1136/jclinpath-2016-203765>
- Lachover-Roth, I., Peretz, S., Zoabi, H., Harel, E., Livshits, L., Filon, D., Levin, C., and Koren, A. (2024). Support vector machine-based formula for detecting suspected α -thalassaemia carriers: A path toward universal screening. *International Journal of Molecular Sciences*, 25(12):6446. <https://doi.org/10.3390/ijms25126446>
- Laengsri, V., Shoombuatong, W., Adirojananon, W., Nantasenamat, C., Prachayasittikul, V., and Nuchnoi, P. (2019). ThalPred: A web-based prediction tool for discriminating thalassemia trait and iron deficiency anemia. *BMC Medical Informatics and Decision Making*, 19(1):212. <https://doi.org/10.1186/s12911-019-0929-2>
- Martins, B., Ferreira, D., Neto, C., Abelha, A., and Machado, J. (2021). Data mining for cardiovascular disease prediction. *Journal of Medical Systems*, 45(1):6. <https://doi.org/10.1007/s10916-020-01682-8>
- Mir, A., Ur Rehman, A., Ali, T.M., Javaid, S., Almufareh, M.F., Humayun, M., and Shaheen, M. (2024). A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques. *ESC Heart Failure*, 11(6):3742-3756. <https://doi.org/10.1002/ehf2.14942>
- Mo, D., Zheng, Q., Xiao, B., and Li, L. (2023). Predicting thalassemia using deep neural network based on red blood cell indices. *Clinica Chimica Acta*, 543:117329. <https://doi.org/10.1016/j.cca.2023.117329>
- Nelson, A.C., Motum, P.I., and Emeto, T.I. (2019). Evaluation of an immunochromatographic test for α -thalassaemia screening in a multi-ethnic population. *International Journal of Laboratory Hematology*, 41(3):397-403. <https://doi.org/10.1111/ijlh.12994>
- Old, J.M. (2003). Screening and genetic diagnosis of haemoglobin disorders. *Blood Reviews*, 17(1):43-53. [https://doi.org/10.1016/S0268-960X\(02\)00061-9](https://doi.org/10.1016/S0268-960X(02)00061-9)
- Phirom, K., Charoenkwan, P., Shoombuatong, W., Sirichotiyakul, S., and Tongsong, T. (2022). DeepThal: A deep learning-based framework for the large-scale prediction of the α -thalassaemia trait using red blood cell parameters. *Journal of Clinical Medicine*, 11(21):6305. <https://doi.org/10.3390/jcm11216305>
- Pranpanus, S., Sirichotiyakul, S., Srisupundit, K., and Tongsong, T. (2009). Sensitivity and specificity of mean corpuscular hemoglobin for screening of α -thalassaemia-1 trait and β -thalassaemia trait. *Journal of the Medical Association of Thailand*, 92(6):739-743.
- Prayalaw, P., Fucharoen, G., and Fucharoen, S. (2014). Routine screening for α -thalassaemia using an immunochromatographic strip assay for haemoglobin Bart's. *Journal of Medical Screening*, 21(3):120-125. <https://doi.org/10.1177/0969141314538611>
- Rustam, F., Ashraf, I., Jabbar, S., Tutusaus, K., Mazas, C., Barrera, A.E.P., and de la Torre Díez, I. (2022). Prediction of β -thalassaemia carriers using complete blood count features. *Scientific Reports*, 12(1):19999. <https://doi.org/10.1038/s41598-022-22011-8>
- Saha, S., Sharma, P., Jain, A.K., Dutta, B., Martínez, L., Saleh, S., Dolai, T.K., Kaviraj, A., Sanyal, T., Nielsen, I., and Das, R. (2025). Detection of β -thalassaemia trait from a heterogeneous population with red cell indices and parameters. *Computers in Biology and Medicine*, 192(PartA):110151. <https://doi.org/10.1016/j.complbiomed.2025.110151>
- Saleem, M., Aslam, W., Lali, M.I.U., Rauf, H.T., and Nasr, E.A. (2023). Predicting thalassemia using feature selection techniques: A comparative analysis. *Diagnostics*, 13(22):3441. <https://doi.org/10.3390/diagnostics13223441>
- Schipper, A., Rutten, M., van Gimmeren, A., Harteveld, C.L., Urrechaga, E., Weerkamp, F., den Besten, G., Krabbe, J., Schoonen, L., Broeren, M., van Wijnen, M., Huijskens, M. J.A.J., Koopmann, T., van Ginneken, B., Kusters, R., and Kurstjens, S. (2024). Machine learning-based prediction of hemoglobinopathies using complete blood count data. *Clinical Chemistry*, 70(8):1064-1075. <https://doi.org/10.1093/clinchem/hvae081>
- Sirichotiyakul, S., Maneerat, J., Sa-nguansermisri, T., Dhananjayanonda, P., and Tongsong, T. (2005). Sensitivity and specificity of mean corpuscular volume testing for screening for α -thalassaemia-1 and β -thalassaemia traits. *Journal of Obstetrics and Gynaecology Research*, 31(3):198-201. <https://doi.org/10.1111/j.1447-0756.2005.00280.x>
- Sirichotiyakul, S., Tantipalakorn, C., Sanguansermisri, T., Wanapirak, C., and Tongsong, T. (2004). Erythrocyte osmotic fragility test for screening of α -thalassaemia-1 and β -thalassaemia trait in pregnancy. *International Journal of Gynecology and Obstetrics*, 86(3):347-350. <https://doi.org/10.1016/j.ijgo.2004.04.037>
- Sudjaroen, Y. (2015). Efficiency assessment of immunochromatographic strip test for the diagnosis of α -thalassaemia-1 carriers. *Journal of Laboratory Physicians*, 7(1):4-10. <https://doi.org/10.4103/0974-2727.157779>
- Tatu, T. (2020). Laboratory Diagnosis of Beta Thalassemia and HbE. In: Zakaria, M. and Hassan, T. (eds.). *Beta Thalassemia*. London: IntechOpen, 174p. <https://doi.org/10.5772/intechopen.90317>
- Tatu, T. and Sweatman, D. (2018). Hemolysis area: A new parameter of erythrocyte osmotic fragility for screening of thalassemia trait. *Journal of Laboratory Physicians*, 10(2):214-220. https://doi.org/10.4103/JLP.JLP_136_17
- Tatu, T., Jannoi, S., Jamwuttipreecha, K., and Sa-nguansermisri, T. (2003). Screening for α -thalassaemia 1 using the dried brilliant cresyl blue method. *Thai Journal of Hematology and Transfusion Medicine*, 13(4):315-320.
- Tatu, T., Klasamut, S., Sorntham, H., Chantanaskulwong, P., Kaewkhampa, N., and Khantarang, P. (2025). Predicting the double heterozygotes of HbE and α -thalassaemia-1 (Southeast Asian Type) using *RapidMiner*-generated hematologic algorithm. *Hemoglobin*, 49(6):390-398. <https://doi.org/10.1080/03630269.2025.2590224>
- Tayapiwatana, C., Kuntaruk, S., Tatu, T., Chiampanichayakul, S., Munkongdee, T., Winichagoon, P., Fucharoen, S., and Kasinrerak, W. (2009). Simple method for screening of α -thalassaemia 1 carriers. *International Journal of Hematology*, 89(5):559-567. <https://doi.org/10.1007/s12185-009-0331-4>
- Tienthavorn, V., Pattanapongsthorn, J., Charoensak, S., Sae-Tung, R., Charoenkwan, P., and Sa-nguansermisri, T. (2006). Prevalence of thalassemia carriers in Thailand (in Thai). *Thai Journal of Hematology and Transfusion Medicine*, 16:307-312.
- Wanapirak, C., Piyamongkol, W., Sirichotiyakul, S., Tayapiwatana, C., Kasinrerak, W., and Tongsong, T. (2011). Accuracy of immunochromatographic strip test in diagnosis of α -thalassaemia-1 carrier. *Journal of the Medical Association of Thailand*, 94(7):761-765.
- Wang, W., Ye, R., Tang, B., and Qi, Y. (2025). MultiThal-classifier: A machine learning-based multi-class model for thalassemia diagnosis and classification. *Clinica Chimica Acta*, 567:120025. <https://doi.org/10.1016/j.cca.2024.120025>

- Winichagoon, P., Kumpan, P., Holmes, P., Finlayson, J., Newbound, C., Kabral, A., Li, B., Nuinoon, M., Fawcett, T., Tayapiwatana, C., Kasinrer, W., and Fucharoen, S. (2015). Validation of the immunochromatographic strip for α -thalassemia screening: A multicenter study. *Translational Research*, 165(6):689-695. <https://doi.org/10.1016/j.trsl.2014.10.013>
- Zhang, F., Yang, J., Wang, Y., Cai, M., Ouyang, J., and Li, J. (2023). TT@MHA: A machine learning-based webpage tool for discriminating thalassemia trait from microcytic hypochromic anemia patients. *Clinica Chimica Acta*, 545:117368. <https://doi.org/10.1016/j.cca.2023.117368>