BAREKENG: Journal of Mathematics and Its Applications
June 2026 Volume 20 Issue 2 Page 1501–1512
P-ISSN: 1978-7227 E-ISSN: 2615-3017
https://doi.org/10.30598/barekengvol20iss2pp1501-1512

APPLICATION OF DISCRETE HIDDEN MARKOV MODELS IN ANALYZING BLOOD TYPE INHERITANCE PATTERNS

Nahrul Hayati [1*], **Eko Sulistyono** [2], **Andini Setyo Anggraeni** [3]

[1,3]*Mathematics Department, Faculty of Information Technology, Institut Teknologi Batam*
Jln. Gajah Mada, Tiban, Batam, 29425, Indonesia

[2]*Mathematics Department, Faculty of Mathematics and Natural Sciences, Universitas Tanjungpura*
Jln. Prof. Dr. H. Hadari Nawawi, Pontianak, 78124, Indonesia

*Corresponding author's e-mail: * **nahrul@iteba.ac.id***

Article Info

Article History:

Received: 12th June 2025
Revised: 12th August 2025
Accepted: 24th September 2025
Available online: 26th January 2026

Keywords:

blood type inheritance;
Discrete Hidden Markov Model.

ABSTRACT

This research investigates the application of a Discrete Hidden Markov Model (DHMM) to analyze inheritance patterns of ABO blood types. Leveraging the DHMM's ability to model systems with hidden states, the study aims to improve the understanding of blood type inheritance dynamics in populations. The model employs six hidden states representing ABO genotypes (I^AI^A, I^Ai, I^BI^B, I^Bi, I^AI^B, and ii) and four observable states corresponding to blood type phenotypes (A, B, AB, and O). The transition and emission matrices followed Mendelian inheritance principles using population allele frequencies, whereas the initial probabilities were computed under Hardy-Weinberg Equilibrium (HWE) assumptions, with parameters calibrated to Indonesian blood type distributions. As a case study, we calculated the likelihood of observing phenotype A across five consecutive generations. Using the forward-backward algorithm, the probability of this sequence was calculated as 19%. The Viterbi algorithm further identified the most probable sequence of hidden genotypes, revealing a transition from the heterozygous I^Ai to the homozygous I^AI^A genotype over the five generations. One iteration of the Baum-Welch algorithm improved model accuracy, increasing log-likelihood from -1.661 to 0. Our results demonstrate the DHMM's efficacy in decoding complex inheritance dynamics and provide a foundation for future population genetics research.

How to cite this article:

N. Hayati, E. Sulistyono and A. S. Anggraeni., "APPLICATION OF DISCRETE HIDDEN MARKOV MODELS IN ANALYZING BLOOD TYPE INHERITANCE PATTERNS", *BAREKENG: J. Math. & App.*, vol. 20, no. 2, pp. 1501-1512, Jun, 2026.

Journal homepage: *https://ojs3.unpatti.ac.id/index.php/barekeng/*
Journal e-mail: *barekeng.math@yahoo.com; barekeng.journal@mail.unpatti.ac.id*

Research Article · **Open Access**

1. INTRODUCTION

The ABO blood group system, first identified by Karl Landsteiner in the early 20th Century, remains a cornerstone of human genetics and medicine [1]. While its Mendelian principles are well established, accurately modeling its inheritance is challenging due to the complexities introduced by multiple alleles, environmental influences, and transitions across generations [2]. These challenges require advanced computational tools, such as Hidden Markov Models (HMMs) to infer hidden genotypic states from observable phenotypic data.

HMMs provide a robust framework for modeling systems with hidden states, making them particularly suitable for genetic inheritance studies [3]. Their capability to handle uncertainty and infer hidden parameters from observed data has proven valuable in various bioinformatics applications, from sequence analysis to gene mapping [4], [5], [6], [7], [8]. Previous research has demonstrated the versatility of HMMs in genetics. Beyond human genetics, Hayati et al. applied Discrete Hidden Markov Models (DHMMs) to analyze crosses in diploid and tetraploid plants [4], [9]. In blood group prediction, Giollo et al. developed BOOGIE, an HMM-based method that uses high-throughput sequencing data to predict blood groups from genomic data [10]. However, while these studies showcase HMMs for static genotype-phenotype prediction, their application to model the temporal dynamics of blood type inheritance across multiple generations remains unexplored. Our study directly addresses this gap by developing a DHMM framework specifically designed to trace and analyze the evolution of inheritance patterns over time, moving beyond single-generation analysis to provide a longitudinal perspective.

The ABO blood group system presents a unique challenge due to its complexity, involving multiple alleles and their interactions [11]. DHMMs are particularly well-suited for this task, as they can accurately model discrete transitions between genotypes and their corresponding categorical phenotypic manifestations. Additionally, HMMs can handle uncertainty and incomplete data, making them ideal for analyzing genetic datasets that often contain ambiguities or missing data [12]. This is especially relevant in generational studies, where ancestral genotypes may not always be known or fully observable.

This research aims to explore and develop the application of DHMMs in analyzing ABO blood type inheritance patterns. By integrating existing genetic knowledge of the ABO system with the computational power of DHMMs, we aim to deepen our understanding of blood type inheritance dynamics across populations. Specifically, our study focuses on modeling the transition of genotypes and phenotypes across generations, providing a framework for predicting genotypes based on phenotypic data and offering new insights into population genetics and evolutionary patterns. Our work builds upon foundational studies in genetic modeling, such as that of Hayati et al. (2016), who applied the Jukes-Cantor model to determine the probability of nitrogen base inheritance in offspring [13]. By extending these approaches, we aim to contribute to the growing body of research that leverages mathematical and computational tools to unravel the complexities of genetic inheritance.

2. RESEARCH METHODS

2.1 The ABO Blood Group System

Karl Landsteiner, an Austrian-American scientist, discovered the ABO blood group system in 1900, a discovery that later earned him the Nobel Prize in Physiology or Medicine in 1930 [14]. This discovery was revolutionary because it explained why some blood transfusions succeeded while others failed, paving the way for safe blood transfusions; it became one of the first genetic characteristics studied in humans, providing valuable insights into genetic inheritance; and it has wide applications in forensic medicine, anthropological studies, and human evolution research.

The ABO gene is located on the long arm of chromosome 9, specifically at position 9q34.1 to 9q34.2 [15]. This specific location is essential because it enables accurate genetic mapping, helps identify mutations or genetic variations that may affect blood type expression, and facilitates studies on the relationship between blood types and certain health conditions linked to nearby genes.

The ABO gene has three main alleles: I^A, I^B, and i. The characteristics of these alleles are:

1. I^A and I^B are codominant, meaning both are expressed when present together.

2. i is recessive to I^A and I^B, meaning its effect is only seen if no I^A or I^B allele is present.

The combination of these alleles results in six possible genotypes. Although there are six genotypes, there are only four phenotypes (A, B, AB, O) due to the codominant nature of I^A and I^B and the recessive nature of i [16].

1. $I^A I^A$: homozygous A
2. $I^A i$: heterozygous A
3. $I^B I^B$: homozygous B
4. $I^B i$: heterozygous B
5. $I^A I^B$: heterozygous AB
6. ii : homozygous O

The inheritance of ABO blood groups follows Mendelian principles, but with additional complexities. The presence of three alleles (rather than two as in classical Mendelian genetics) adds variation in possible offspring genotypes and the codominant nature of I^A and I^B results in a mixed phenotype (AB) not seen in simple dominant-recessive inheritance [17].

2.2 The Hardy-Weinberg Equilibrium (HWE)

The Hardy-Weinberg Equilibrium (HWE) is a fundamental principle in population genetics that describe how allele and genotype frequencies remain constant from generation to generation in an ideal population. This principle was first formulated independently by G.H. Hardy and Wilhelm Weinberg in 1908 [18]. Although the principle is often introduced using simple Mendelian traits with two allele, it is also powerfully applicable to more complex genetic systems.

In the context of the ABO blood group system, which is governed by a multiple allelic system with three main alleles (I^A, I^B, and i), the Hardy-Weinberg equilibrium has complex yet important applications. The ABO blood group system is a classic example of multiple alleles and incomplete dominance in human genetics. The basic assumptions of Hardy-Weinberg equilibrium include large population size, random mating, no mutation, no selection, and no gene flow (migration).

For the three alleles governing the ABO system (I^A, I^B, and i), genotype frequencies under Hardy-Weinberg equilibrium are given by the multinomial expansion $p^2 + q^2 + r^2 + 2pq + 2pr + 2qr = 1$, where p, q, and r represent the frequencies of alleles I^A, I^B, and i, respectively. The corresponding phenotype frequencies for blood groups A, B, AB, and O are derived directly from these genotype probabilities.

Although Hardy-Weinberg equilibrium rarely occurs perfectly in real populations, this principle remains a valuable tool for understanding population genetic dynamics and detecting factors that may influence allele frequencies, such as selection or gene flow [19].

2.3 The Discrete Hidden Markov Model (DHMM)

The DHMM is a statistical framework used to model stochastic processes with unobservable states that can be inferred from observable variables [20]. It consists of an unseen Markov chain $X = \{X_k\}$ paired with an observation process $Y = \{Y_k\}$ with $k \in N$. In this model, X_{k+1} influences the Markov chain X_k, while X_k affects the observation Y_k. Key components of a DHMM include:

1. A transition probability matrix $\boldsymbol{A}$, which represents the likelihood of moving from one hidden state to another. Each element is defined as:

$$a_{ij} = P(X_{k+1} = j | X_k = i) \text{ for } i, j = 1,2, \cdots, N$$

where $a_{ij} \geq 0$ and $\sum_{j=1}^{N} a_{ij} = 1$ for all i. In our context of blood type inheritance, this matrix is constructed based on Mendelian inheritance rules. For example, the probability a_{ij} defines the chance of an offspring having genotype j given the parental genotype i.

2. An emission probability matrix $\boldsymbol{B}$, which indicates the probability of observing a particular output given the current hidden state. Each element is defined as:

$$b_i(j) = P(Y_k = j | X_k = i) \text{ for } i = 1,2, \cdots, N \text{ and } j = 1,2, \cdots, M$$

where $b_i(j) \geq 0$ and $\sum_{j=1}^{M} b_i(j) = 1$ for all i. Crucially, in this specific model, the emission probabilities are deterministic. This means a genotype always produces a single, specific phenotype with a probability of 1 (e.g., $P(\text{Phenotype } A | \text{Genotype } I^A I^A) = 1$). This choice

reflects the well-established biological principle that the ABO allele interactions are complete and predictable for phenotype expression. However, the DHMM framework allows for stochastic emission to model uncertainty, which is a direction for future work.

3. An initial state probability vector $\boldsymbol{\pi}$, which denotes the likelihood of the system starting in each hidden state. Each element is defined as:

$$\pi_i = P(X_1 = i) \text{ for } i = 1,2,\cdots,N$$

where $\sum_{i=1}^{N} \pi_i = 1$. In this study, the initial probabilities were calculated from population allele frequencies using the HWE principle.

These elements are collectively represented as the parameter set $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$.

2.4 Key Algorithms of The Discrete Hidden Markov Model

The DHMM framework addresses three core problems, calculating the probability of an observation sequence, finding the most probable sequence of hidden states, and learning the model parameters that best fit the data. A dedicated algorithm exists for each problem.

1. The Forward-Backward Algorithm

   The Forward-Backward algorithm solves the first problem of evaluating the probability of a given observation sequence $O = (o_1, o_2, \cdots, o_K)$ under a model λ. The forward algorithm proceeds chronologically, while the backward algorithm works in reverse. Forward algorithm steps:

   a. Initialization: $\alpha_1(i) = \pi_i b_i(y_1)$ for $i = 1,2,\cdots,N$.

   b. Induction: $\alpha_{k+1}(j) = \left(\sum_{i=1}^{N} \alpha_k(i) a_{ij}\right) b_j(y_{k+1})$ for $j = 1,2,\cdots,N$ and $k = 1,2,\cdots,K-1$.

   c. Termination: $P(O|\lambda) = \sum_{i=1}^{N} \alpha_K(i)$.

2. The Viterbi Algorithm

   The Viterbi algorithm solves the second problem of finding the single best state sequence $Q = (q, q_2, \cdots, q_K)$ that maximizes $P(Q|O,\lambda)$. Viterbi algorithm steps:

   a. Initialization: $\delta_1(i) = \pi_i b_i(o_1)$ and $\psi_1(i) = \emptyset$ for $i = 1,2,\cdots,N$.

   b. Recursion:

   $$\delta_k(j) = b_j(y_k) \max_{1 \le i \le N} \{a_{ij} \delta_{k-1}(i)\} \text{ and } \psi_k(j) = \arg \max_{1 \le i \le N} \{a_{ij} \delta_{k-1}(i)\} \text{ for } k = 2,3,\cdots,K-1.$$

   c. Termination: $P^* = \max_{1 \le i \le N} \{\delta_K(i)\}$ and ${x_K}^* = \arg \max_{1 \le i \le N} \{\delta_K(i)\}$.

   d. Backtracking: ${x_k}^* = \psi_{k+1}({x_{k+1}}^*)$ for $k = K-1, K-2, \cdots, 1$.

3. The Baum-Welch Algorithm

   The Baum-Welch algorithm is an Expectation-Maximization (EM) procedure that solves the third problem of re-estimating the parameter $\lambda = (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$. Given variables:

   $$\xi_k(i,j) = \frac{\alpha_k(i) a_{ij} b_j(y_{k+1}) \beta_{k+1}(j)}{\sum_{j=1}^{N} \alpha_k(j) \beta_k(j)} \text{ and } \gamma_k(i) = \sum_{j=1}^{N} \xi_k(i,j) \text{ for } i = 1,2,\cdots,N \text{ and } k = 1,2,\cdots,K.$$

   a. A transition probability matrix $\widehat{\boldsymbol{A}}$: $\hat{a}_{ij} = \frac{\sum_{k=1}^{K-1} \xi_k(i,j)}{\sum_{k=1}^{K-1} \gamma_k(i)}$ for $i = 1,2,\cdots,N$ and $j = 1,2,\cdots,K$.

   b. An emission probability matrix $\widehat{\boldsymbol{B}}$: $\hat{b}_i(j) = \frac{\sum_{k=1, \text{s.t} y_k = j}^{K} \gamma_k(i)}{\sum_{k=1}^{K} \gamma_k(i)}$.

   c. An initial state probability vector $\widehat{\boldsymbol{\pi}}$: $\hat{\pi}_i = \gamma_1(i)$ for $i = 1,2,\cdots,N$.

These algorithms aim to find parameters $\hat{\lambda}$ that satisfy $P(O|\hat{\lambda}) \ge P(0|\lambda)$, where O is the observation sequence and λ represents the current model parameters [21].

2.4 Data Source

The study utilized phenotype frequency data for ABO blood types ($A, B, AB,$ and O) from the Indonesian Ministry of Home Affairs' Directorate General of Population and Civil Registration (Dirjen Dukcapil). The dataset, updated as of 31 December 2024 [22], compiles civil registration records from 38 Indonesian provinces. Data were compiled from civil registration records across all provinces in Indonesia. Blood type information was self-reported by individuals during national ID card registration updates [23], [24]. The dataset reflects blood type distributions across diverse ethnic groups and geographic regions in Indonesia. This dataset comprised approximately 38.4 million individuals, representing a nationally aggregated sample.

However, it is important to acknowledge the limitations of this administrative dataset. As the blood type information was self-reported without subsequent serological verification, the potential for misclassification bias cannot be ruled out. Furthermore, the data were used in their raw, aggregated form at the provincial level. No further data cleaning, verification procedures, or specific exclusion criteria were applied, as the study relied on the pre-processed aggregate statistics provided by Dirjen Dukcapil.

The choice of this dataset was based on:

1. Representativeness

   The massive, nationwide sample size provides unparalleled geographic and ethnic coverage, ensuring statistical robustness for initial allele frequency estimation.

2. Availability

   Raw genotype data were unavailable. Therefore, genotype frequencies ($I^AI^A, I^Ai, I^BI^B, I^Bi, I^AI^B,$ and ii) were derived from the observed phenotype frequencies using HWE equations (Section 2.2).

Table 1. Distribution and Proportion of Each Blood Type

Blood Type	Individuals	Proportion
A	8 390 388	0.218
B	8 666 202	0.226
AB	3 341 441	0.087
O	18 021 283	0.469
Total	**38 419 314**	**1**

Table 1 presents the absolute counts of individuals in Indonesia across the four ABO blood type phenotypes ($A, B, AB,$ and O), as recorded by the Directorate General of Population and Civil Registration (Dirjen Dukcapil) as of December 31, 2024. Among the total sample 38 419 314 individuals, representing a subset of Indonesia's population. Blood type O was the most prevalent with 18.021.283 people (46.9%), followed by type B (22.6%), type A (21.8%), and type AB (8.7%). This distribution suggests that the recessive allele i (which determines blood type O) is more common than the I^A and I^B alleles, consistent with global trends where type O is frequently the most widespread. The relatively high frequency of type AB may reflect genetic diversity or a specific selection factor in Indonesia. This data served as the foundation for calculating allele frequencies using the HWE for initializing the DHMM parameters.

2.5 Data Analysis Procedure

The analysis followed a structured pipeline to model blood type inheritance using a DHMM:

1. DHMM Initialization

   Transition matrix $\boldsymbol{A}$ was constructed based on Mendelian inheritance rules, while emission matrix $\boldsymbol{B}$ was defined deterministically. Initial probabilities $\boldsymbol{\pi}$ were calculated using HWE with Indonesian allele frequencies.

2. Model Training and Validation

   First, the forward-backward algorithm was applied to estimate the probability of observed phenotype sequences (e.g., five generations of type A), while the Viterbi algorithm decoded hidden genotypes. Model accuracy was improved via one iteration of the Baum-Welch algorithm, with convergence monitored through log-likelihood values.

3. RESULTS AND DISCUSSION

3.1 Discrete Hidden Markov Model Initialization

In the context of ABO blood type, hidden states represent possible genotypes. There are six hidden states: I^AI^A, I^Ai, I^BI^B, I^Bi, I^AI^B, and ii. These states are not directly observable and form the basis of the hidden Markov model. Observations in this model are the blood type phenotypes that can be directly observed through laboratory test. There are four possible observations: A, B, AB, and O (Table 2).

Table 2. ABO Blood Type Genotype-Phenotype Frequencies

Genotypic Frequency	Genotype	Phenotype
p^2	I^AI^A	A
$2pr$	I^Ai	A
q^2	I^BI^B	B
$2qr$	I^Bi	B
$2pq$	I^AI^B	AB
r^2	ii	O

The transition matrix $(\boldsymbol{A})$ is a 6×6 matrix that defines the probability of an offspring having a specific genotype based on parental genotype, governed by Mendelian inheritance laws. The entries of matrix $\boldsymbol{A}$ are derived from the probabilities of allele segregation, as illustrated by the Punnett squares in Figure 1.

This figure illustrates how the probabilities in the transition matrix $\boldsymbol{A}$ are calculated based on Mendelian inheritance. Each square shows the potential offspring genotypes resulting from a specific parental genotype cross. The standard genotype abbreviations are used: I^AI^A (homozygous A), I^Ai (heterozygous A), I^BI^B (homozygous B), I^Bi (heterozygous B), I^AI^B (heterozygous AB), and ii (homozygous O). The probabilities of each offspring genotype are calculated from the allele combinations.

I^AI^A	×	I^AI^A
I^AI^A 100%		

I^Ai	×	I^Ai
(I^A, i)		(I^A, i)

	I^A	i
I^A	I^AI^A	I^Ai
i	I^Ai	ii

$I^AI^A\left(\frac{1}{4}\right), I^Ai\left(\frac{1}{2}\right), I^BI^B(0), I^Bi(0), I^AI^B(0), ii\left(\frac{1}{4}\right)$

I^BI^B	×	I^BI^B
I^BI^B 100%		

ii	×	ii
ii 100%		

I^Bi	×	I^Bi
(I^B, i)		(I^B, i)

	I^B	i
I^B	I^BI^B	I^Bi
i	I^Bi	ii

$I^AI^A(0), I^Ai(0), I^BI^B\left(\frac{1}{4}\right), I^Bi\left(\frac{1}{2}\right), I^AI^B(0), ii\left(\frac{1}{4}\right)$

I^AI^B	×	I^AI^B
(I^A, I^B)		(I^A, I^B)

	I^A	I^B
I^A	I^AI^A	I^AI^B
I^B	I^AI^B	I^BI^B

$I^AI^A\left(\frac{1}{4}\right), I^Ai(0), I^BI^B\left(\frac{1}{4}\right), I^Bi(0), I^AI^B\left(\frac{1}{2}\right), ii(0)$

Figure 1. Derivation of Genotype Transition Probabilities using Punnett Square

Figure 1 provides a visual derivation of these probabilities. For example, a homozygous parent (I^AI^A) can only pass the I^A allele to its offspring and a heterozygous parent (e.g., I^Ai) has a 50% probability of passing either the I^A allele or the i allele to its offspring. Based on the principles shown in Figure 1, the probabilities for offspring genotype can be calculated. For instance:

1. The probability of two I^AI^A parents producing a homozygous I^AI^A offspring is 1.
2. The probability of two I^Ai (heterozygous A) parents producing a homozygous I^AI^A offspring is 0.25.
3. The probability of them producing another heterozygous I^Ai offspring is 0.5.
4. The probability of them producing a homozygous ii (O) offspring is 0.25.

Therefore, the complete transition matrix $\boldsymbol{A}$ is defined as follows:

$$\boldsymbol{A} = \begin{bmatrix} a_{I^AI^A,I^AI^A} & a_{I^AI^A,I^Ai} & a_{I^AI^A,I^BI^B} & a_{I^AI^A,I^Bi} & a_{I^AI^A,I^AI^B} & a_{I^AI^A,ii} \\ a_{I^Ai,I^AI^A} & a_{I^Ai,I^Ai} & a_{I^Ai,I^BI^B} & a_{I^Ai,I^Bi} & a_{I^Ai,I^AI^B} & a_{I^Ai,ii} \\ a_{I^BI^B,I^AI^A} & a_{I^BI^B,I^Ai} & a_{I^BI^B,I^BI^B} & a_{I^BI^B,I^Bi} & a_{I^BI^B,I^AI^B} & a_{I^BI^B,ii} \\ a_{I^Bi,I^AI^A} & a_{I^Bi,I^Ai} & a_{I^Bi,I^BI^B} & a_{I^Bi,I^Bi} & a_{I^Bi,I^AI^B} & a_{I^Bi,ii} \\ a_{I^AI^B,I^AI^A} & a_{I^AI^B,I^Ai} & a_{I^AI^B,I^BI^B} & a_{I^AI^B,I^Bi} & a_{I^AI^B,I^AI^B} & a_{I^AI^B,ii} \\ a_{ii,I^AI^A} & a_{ii,I^Ai} & a_{ii,I^BI^B} & a_{ii,I^Bi} & a_{ii,I^AI^B} & a_{ii,ii} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ \frac{1}{4} & \frac{1}{2} & 0 & 0 & 0 & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ \frac{1}{4} & 0 & \frac{1}{4} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

The emission matrix ($\boldsymbol{B}$) describes the probabilities of a genotype producing a certain phenotype. It is a 6×4 matrix that links hidden states to observations. The matrix reflects the deterministic relationship between genotypes and phenotypes in the ABO blood type system. The examples of some matrix elements are genotype I^AI^A always produces phenotype A, genotype I^Ai always produce phenotype A, genotype I^AI^B always produces phenotype AB and genotype ii always produces phenotype O.

$$\boldsymbol{B} = \begin{bmatrix} b_{I^AI^A}(A) & b_{I^AI^A}(B) & b_{I^AI^A}(AB) & b_{I^AI^A}(O) \\ b_{I^Ai}(A) & b_{I^Ai}(B) & b_{I^Ai}(AB) & b_{I^Ai}(O) \\ b_{I^BI^B}(A) & b_{I^BI^B}(B) & b_{I^BI^B}(AB) & b_{I^BI^B}(O) \\ b_{I^Bi}(A) & b_{I^Bi}(B) & b_{I^Bi}(AB) & b_{I^Bi}(O) \\ b_{I^AI^B}(A) & b_{I^AI^B}(B) & b_{I^AI^B}(AB) & b_{I^AI^B}(O) \\ b_{ii}(A) & b_{ii}(B) & b_{ii}(AB) & b_{ii}(O) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The initial probability matrix ($\boldsymbol{\pi}$) describes the initial probability distribution of each genotype in the populations. The matrix π reflects the relative frequencies of each genotype in the studied population and are typically based on epidemiological or population genetic data.

$$\boldsymbol{\pi} = [\pi_{I^AI^A} \quad \pi_{I^Ai} \quad \pi_{I^AI^B} \quad \pi_{I^Bi} \quad \pi_{I^AI^B} \quad \pi_{ii}]$$

To apply a DHMM in analyzing blood type inheritance patterns, the value of the initial probability matrix ($\boldsymbol{\pi}$) were determined using the HWE based on an example dataset from Table 1. The step-by-step procedure is as follows:

1. Step 1: Determine The HWE Equation.

$$P(A) = p^2 + 2pr = 0.218$$
$$P(B) = q^2 + 2qr = 0.226$$
$$P(AB) = 2pq = 0.087$$
$$P(O) = r^2 = 0.469$$

2. Step 2: Calculate Allele Frequency.

   Solve the Step 1 using a Non-Linear Equation System:

$$p = 0.2057, q = 0.2114, r = 0.4279$$
$$p + q + r = 0.8450$$

3. Step 3: Normalize to Ensure $p + q + r = 1$.

$$\hat{p} = \frac{0.2057}{0.8450} = 0.2434, \hat{q} = \frac{0.2114}{0.8450} = 0.2502, \hat{r} = \frac{0.4279}{0.8450} = 0.5064.$$

Based on the calculation results, the allele frequencies are as follows:

1. Frequency of I^A allele $(p) = 0.2434$;
2. Frequency of I^B allele $(q) = 0.2502$;
3. Frequency of i allele $(r) = 0.5064$.

These three alleles appear to form a multiple allele system (a system with more than two variants) since the total sum of the three allele frequencies equals 1. This frequency distribution indicates significant genetic

diversity in the population, with one dominant allele and two minority alleles. With this data, the initial probability matrix can be calculated as follows:

1. $P(I^AI^A) = p^2 = (0.2434)^2 = 0.059$;
2. $P(I^Ai) = 2pr = 2(0.2434)(0.5064) = 0.247$;
3. $P(I^BI^B) = q^2 = (0.2502)^2 = 0.063$;
4. $P(I^Bi) = 2qr = 2(0.2502)(0.5064) = 0.253$;
5. $P(I^AI^B) = 2pq = 2(0.2434)(0.2502) = 0.122$;
6. $P(ii) = r^2 = (0.5064)^2 = 0.256$.

Thus, the initial probability matrix becomes:

$$\boldsymbol{\pi} = [0.059 \quad 0.247 \quad 0.063 \quad 0.253 \quad 0.122 \quad 0.256].$$

Based on these results, it can be determined that genotype ii (blood type O) has the highest probability with 25.6%. Followed by I^Bi (heterozygous blood type B) with 25.3%. I^Ai (heterozygous blood type A) has a probability of 24.7%. Genotype I^AI^A (homozygous blood type B) has the lowest probability (5.9%).

These three matrices ($\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{\pi}$) form the main parameters of the DHMM used to analyze and predict ABO blood type inheritance pattern. This model allows for prediction of genotypes based on observed phenotype sequences, analysis of inheritance patterns in family pedigrees, estimation of allele frequencies in populations, and better understanding of the genetic dynamics of ABO blood types.

3.2 Model Training and Validation

The first problem in a DHMM is to calculate the probability of an observation sequence using the forward and backward algorithm. For example, let's calculate the probability of five generations having phenotype A consecutively using the forward algorithm.

Table 3. Calculation of the Observation Sequence Probability using the Forward Algorithm

t	1	2	3	4	5
$\alpha_t(1)$	0.059	0.121	0.152	0.167	0.175
$\alpha_t(2)$	0.247	0.124	0.062	0.031	0.015
$\alpha_t(3)$	0	0	0	0	0
$\alpha_t(4)$	0	0	0	0	0
$\alpha_t(5)$	0	0	0	0	0
$\alpha_t(6)$	0	0	0	0	0
		$\boldsymbol{P(O\|\lambda)}$			**0.190**

Table 3 presents the step-by-step results of the forward algorithm for calculating the probability of observing phenotype A across five consecutive generations, i.e., $P(O = (A, A, A, A, A)|\lambda)$. The forward probability $\alpha_t(j)$ represents the probability of the partial observation sequence up to time t and being in state j at time t, given the model λ. The final probability, obtained by summing the forward probabilities at $t = 5$ for all states emitting phenotype A (I^AI^A and I^Ai), is 0.190 or 19%.

$$P(A, A, A, A, A) = \alpha_5(I^AI^A) + \alpha_5(I^Ai) = \alpha_5(1) + \alpha_5(2) = 0.175 + 0.015 = 0.190$$

This probability is relatively high, considering that phenotype A can be produced by two genotypes (I^AI^A and I^Ai). The increasing value of $\alpha_t(I^AI^A)$ and the decreasing value of $\alpha_t(I^Ai)$ from generation to generation indicate a cumulative effect, where the probability becomes increasingly concentrated in homozygous I^AI^A state, which always produces offspring with phenotype A.

The next problem is to find the most likely sequence of hidden states (genotypes) based on the given sequence of observations (phenotypes) using the Viterbi algorithm. For example, five generations have phenotype A consecutively.

Table 4. Traceback of the Most Probable Genotype Sequence using Viterbi Algorithm's Path Pointers

t	1	2	3	4	5
$\psi_t(1)$	0	I^Ai	I^AI^A	I^AI^A	I^AI^A
$\psi_t(2)$	0	I^Ai	I^Ai	I^Ai	I^Ai

t	1	2	3	4	5
$\psi_t(3)$	0	0	0	0	0
$\psi_t(4)$	0	0	0	0	0
$\psi_t(5)$	0	0	0	0	0
$\psi_t(6)$	0	0	0	0	0

The Viterbi variable $\psi_t(j)$ stores the most likely previous state that leads to state j at time t, used to reconstruct the optimal hidden path. Table 4 shows the backtracking values for the observation sequence of five consecutive phenotype A generations. The most probable genotype sequence, determined by tracing the ψ pointers from the final step, is

$$x^* = \{I^A i, I^A I^A, I^A I^A, I^A I^A, I^A I^A\}.$$

The most probable sequence of hidden genotypes, identified by applying the traceback function to the values in Table 4, is $I^A i, I^A I^A, I^A I^A, I^A I^A, I^A I^A$. This result reveals a critical hidden dynamic, although the observed phenotype (A) remained constant for five generations, the underlying genotype likely shifted from heterozygous ($I^A i$) in the first generation to homozygous ($I^A I^A$) for the subsequent four. This transition is genetically logical, as a cross between two $I^A i$ parents have a probability of 0.25 of producing a homozygous $I^A I^A$ offspring. Once the $I^A I^A$ genotype is established, it will stably produce offspring with the same genotype and phenotype indefinitely. This finding powerfully illustrates the utility of the DHMM and the Viterbi algorithm in inferring hidden genetic states that are not directly observable from phenotypic data alone.

The final problem is to re-estimate DHMM parameters to maximize the probability of the observation sequence. To update the model parameters using the Baum-Welch algorithm, we need to perform several iterations [4]. However, due to the observation data for five generations with phenotype A, only one iteration is performed as an example. The updated model:

1. Initial state matrix $\hat{\boldsymbol{\pi}}$: $\hat{\pi}_{I^A I^A} = 0.31$ and $\hat{\pi}_{I^A i} = 0.69$ (the values for other states are 0).
2. Transition matrix $\hat{\boldsymbol{A}}$: $\hat{a}_{I^A i, I^A I^A} = 0.441$, and $\hat{a}_{I^A i, I^A i} = 0.559$ (other transitions remain the same as before).
3. Emission matrix $\hat{\boldsymbol{B}}$: unchanged because all observations are A.

Based on the result, the initial probability for genotype $I^A I^A$ increased from 0.061 to 0.314 and the initial probability for genotype $I^A i$ increased from 0.249 to 0.686. The transition probability from $I^A i$ to $I^A I^A$ increased from 0.25 to 0.466, and the transition probability from $I^A i$ to $I^A i$ increased from 0.5 to 0.534. This shows an increased tendency to stay in or transition to genotypes that produce phenotype A. The emission matrix remains unchanged because all observations are phenotype A and the genotype-phenotype relationship remains deterministic. The new parameters are as follows:

$$\hat{\boldsymbol{A}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0.441 & 0.559 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.25 & 0.5 & 0 & 0.25 \\ 0.25 & 0 & 0.25 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \hat{\boldsymbol{B}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \hat{\boldsymbol{\pi}} = \begin{bmatrix} 0.31 \\ 0.69 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

The model generated through one iteration of the Baum-Welch algorithm provides an initial picture of how DHMM parameters can be optimized to analyze blood type inheritance patterns. However, it is essential to recognize that these results are just the initial step in a more comprehensive process. In actual implementation, the Baum-Welch algorithm performs many iterations, continuously updating the model parameters (initial probabilities, transition matrix, and emission matrix) until convergence is reached or a predetermined maximum number of iterations is met.

The observation of five consecutive generations with phenotype A used in this study represents an effective case study to illustrate the application of DHMMs in genetic inheritance analysis. This approach can be further developed by expanding the dataset to include longer and more diverse sequences of generations, involving all blood type phenotypes (A, B, AB, O). This would result in a more robust model with more accurate parameter estimates, thereby improving the quality of predictions and understanding regarding the dynamics of blood type inheritance in populations.

The re-estimated model can also serve as a starting point for subsequent iterations, where the forward, backward, Viterbi, and Baum-Welch algorithms are executed again to further refine the model parameters. This iterative process will result in a model that increasingly converges toward an optimal value. In each iteration, the parameters obtained from the previous iteration are used as the initial value for the next re-estimation process.

Table 5. Forward Algorithm Computation for the Observation Sequence after Baum-Welch Re-Estimation

t	1	2	3	4	5
$\alpha_t(1)$	0.310	0.614	0.784	0.880	0.933
$\alpha_t(2)$	0.690	0.386	0.216	0.120	0.067
$\alpha_t(3)$	0	0	0	0	0
$\alpha_t(4)$	0	0	0	0	0
$\alpha_t(5)$	0	0	0	0	0
$\alpha_t(6)$	0	0	0	0	0
		$\boldsymbol{P(O\|\lambda)}$			**1**

The forward variable $\alpha_t(j)$ represent the probability of the partial observation sequence until time t and being in state j at time t, given the re-estimated model $\hat{\lambda}$. The final probability $P(O|\hat{\lambda}) = \sum \alpha_5(j)$ for all states j that emit phenotype A (i.e., $I^A I^A$ and $I^A i$) yields the result of 1.

$$P(A, A, A, A, A) = \alpha_5(I^A I^A) + \alpha_5(I^A i) = \alpha_5(1) + \alpha_5(2) = 0.933 + 0.067 = 1.$$

As detailed in Table 5, the forward algorithm was executed using the re-estimated parameters ($\hat{\lambda}$) obtained from the Baum-Welch algorithm. The result shows a perfect probability score, $P(O|\hat{\lambda}) = 1$, for the observation sequence of five consecutive phenotype A generations. This signifies a 100% probability under the optimized model, demonstrating a dramatic increase from the initial model's probability and unequivocally confirming the effectiveness of the Baum-Welch re-estimation process. Furthermore, the concentration of probability mass in the $\alpha_5(I^A I^A)$ value (0.933) suggests the optimized model now strongly favors the homozygous $I^A I^A$ genotype as the most probable hidden state underlying the sustained phenotypic observation.

One effective method for evaluating the performance of the DHMM model and monitoring the convergence process is to use the log likelihood. Comparing the log likelihood results between the initial model and the re-estimated model provides important insights into the effectiveness of the re-estimation process. In the case of this blood type inheritance model, the initial model's log likelihood value was -1.661, while the model after Baum-Welch re-estimation achieved a value of 0. This increase of 1.640 indicates that the re-estimated model is significantly better at explaining the observed data.

$$\mathcal{L}(\lambda) = \ln P(O|\lambda) = \ln P(A, A, A, A, A)$$

1. The initial model's log likelihood: $\mathcal{L}(\lambda) = \ln(0.194) = -1.661$;
2. The re-estimated model's log likelihood: $\mathcal{L}(\hat{\lambda}) = \ln(1) = 0$.

A log likelihood value of 0 (likelihood = 1) in the re-estimated model suggests that the model provides maximum probability or absolute certainty in generating a sequence of phenotype A consecutively over five generations. This perfect log likelihood is direct consequence of the highly specific and deterministic nature of our experimental setup. It occurs because:

1. The emission probabilities in our model are defined deterministically (e.g., $P(A|I^A I^A = 1)$). A genotype always produces its corresponding phenotype.
2. The Baum-Welch algorithm re-estimated the transition matrix to perfectly explain the single, noiseless observation sequence of five 'A's. In essence, it learned a specialized model where the most probable path perfectly generates the observations.
3. The model was trained on a single, short, and perfectly observed sequence. In real-world applications, models are trained on large, noisy datasets with diverse sequences, making a perfect likelihood statistically impossible and indicative of overfitting.

Therefore, while this result demonstrates the algorithm's ability to achieve perfect convergence on a constrained problem, it also highlights that our initial model parameters were not optimal for this specific sequence. For practical applications, techniques such as regularization or using larger dataset would be necessary to prevent overfitting and ensure generalizability.

4. CONCLUSION

This research applied a DHMM to analyze the inheritance patterns of ABO blood types. Through the use of forward-backward, Viterbi, and Baum-Welch algorithms, the model successfully estimated the probability of phenotype sequences, identified the most likely hidden genotype sequence, and optimized model parameters. The results demonstrate the model's capability to adapt to observational data. Specifically, the probability of observing phenotype A consecutively across five generations was calculated to be 0.190 (19%) under the initial model. A key insight from the Viterbi algorithm was that although phenotype A was observed consistently, the underlying genotype likely shifted from heterozygous ($I^A i$) in the first generation to homozygous ($I^A I^A$) in the subsequent generations. This highlights the complexity of genotype-phenotype relationships and underscores the importance of inferring hidden genetic states. The Baum-Welch re-estimation process significantly improved the model's fit to the specific observed sequence. The log likelihood increased from -1.661 to 0, indicating that the re-estimated parameters yielded a perfect likelihood for this training sequence. While this demonstrates the algorithm's effectiveness in parameter optimization, this specific result of $P(O|\hat{\lambda}) = 1$ is a characteristic of its convergence on a short, noiseless sequence with deterministic emissions and may indicate overfitting. To build upon this work, future research should focus on applying the DHMM framework to larger and more diverse datasets, incorporating genomic sequencing data for validation of predicted genotypes, and extending the analysis to other complex multiallelic traits. Investigating methods to prevent overfitting, such as regularization or the use of prior distributions, would also be valuable for enhancing the model's generalizability to real-world genetic data.

Author Contributions

Nahrul Hayati: Conceptualization, Funding Acquisition, Methodology, Project Administration, Supervision, Writing-Original Draft. Eko Sulistyono: Data Curation, Formal Analysis, Investigation, Software, Validation, Writing-Review and Editing. Andini Setyo Anggraeni: Data curation, Resources, Visualization, Writing-Review and Editing. All authors discussed the results and approved the final version of the manuscript.

Funding Statement

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Acknowledgment

The authors are deeply thankful for the thorough and insightful comments provided by the anonymous reviewers, which have significantly improved the quality of this manuscript.

Declarations

The authors declare no conflicts of interest to report study.

Declaration of Generative AI and AI-assisted technologies

ChatGPT was utilized only to improve the readability and grammatical structure of the manuscript. No AI tool was used to generate or alter the research data, methodology, results, or interpretations. All content was verified by the authors for accuracy and consistency with the study.

REFERENCES

[1] C. Kim, "HUMAN RED BLOOD CELL (RBC) BLOOD GROUPS SYSTEM," *In: Glycoimmunology in Xenotransplantation*, Singapore: Springer, pp. 35–45, 2024, doi: https://doi.org/10.1007/978-981-99-7691-1_6.

[2] S. Hazari, S. Bhavusaheb, P. Santanu, and N. M. Paul, *FUNDAMENTALS OF PLANT GENETICS EDITORS*. National Press Associates, 2024. [Online]. Available: www.npapublishing.in

[3] M. Franzese and A. Iuliano, "HIDDEN MARKOV MODELS," *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, vol. 1–3, pp. 753–762, Jan. 2019, doi: https://doi.org/10.1016/B978-0-12-809633-8.20488-3.

[4] N. Hayati, B. Setiawaty, and I. Purnaba, "THE APPLICATION OF DISCRETE HIDDEN MARKOV MODEL ON CROSSES OF DIPLOID PLANT," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 17, no. 3, pp. 1449–1462, 2023, doi: https://doi.org/10.30598/barekengvol17iss3pp1449-1462.

[5] M. Mollinari and A. A. F. Garcia, "LINKAGE ANALYSIS AND HAPLOTYPE PHASING IN EXPERIMENTAL AUTOPOLYPLOID POPULATIONS WITH HIGH PLOIDY LEVEL USING HIDDEN MARKOV MODELS," *G3: Genes, Genomes, Genetics*, vol. 9, no. 10, pp. 3297–3314, Oct. 2019, doi: https://doi.org/10.1534/g3.119.400378.
[6] M. Fang, X. Wang, Y. Chen, P. Wang, L. Lu, J. Lu, F. Yao and Y. Zhang., "GENOME SEQUENCE ANALYSIS OF AURICULARIA HEIMUER COMBINED WITH GENETIC LINKAGE MAP," *Journal of Fungi*, vol. 6, no. 1, Mar. 2020, doi: https://doi.org/10.3390/jof6010037.
[7] M. Feldgarden, V. Brover, N. G. Escalona, J. G. Frye, J. Haendiges, D. H. Haft, M. Hoffmann, J. B. Pettengill, A. B. Prasad, G. E. Tillman, G. H. Tyson and W,. Klimke., "AMRFINDERPLUS AND THE REFERENCE GENE CATALOG FACILITATE EXAMINATION OF THE GENOMIC LINKS AMONG ANTIMICROBIAL RESISTANCE, STRESS RESPONSE, AND VIRULENCE," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: https://doi.org/10.1038/s41598-021-91456-0.
[8] M. Sesia, E. Katsevich, S. Bates, E. Candès, and C. Sabatti, "MULTI-RESOLUTION LOCALIZATION OF CAUSAL VARIANTS ACROSS THE GENOME," *Nat Commun*, vol. 11, no. 1, Dec. 2020, doi: https://doi.org/10.1038/s41467-020-14791-2.
[9] N. Hayati, E. Sulistyono, and V. Aprilla Handayani, "UTILIZING DISCRETE HIDDEN MARKOV MODELS TO ANALYZE TETRAPLOID PLANT BREEDING," *Jurnal Matematika UNAND*, vol. 13, no. 4, pp. 244–256, 2024, doi: https://doi.org/10/25077/jmua.13.4.244-256.2024.
[10] M. Giollo, G. Minervini, M. Scalzotto, E. Leonardi, C. Ferrari, and S. Tosatto, "BOOGIE: PREDICTING BLOOD GROUPS FROM HIGH THROUGHPUT SEQUENCING DATA," *PLoS One*, vol. 10, no. 4, pp. 1–15, 2015.
[11] N. Girma, "FREQUENCY OF ABO, RH BLOOD GROUP ALLELES AMONG OROMO, AMHARA AND WOLAYITA ETHNIC GROUP STUDENTS IN ROBE SECONDARY, PREPARATORY AND ZEYBELA PRIMARY SCHOOL, BALE, ETHIOPIA," *International Journal of Genetics and Genomics*, vol. 5, no. 2, p. 19, 2017, doi: https://doi.org/10.11648/j.ijgg.20170502.11.
[12] K. I. Shahin, C. Simon, P. Weber, A. Johansen, and M. B. Kjærgaard, "PROGNOSTIC CONSIDERING MISSING DATA: AN INPUT OUTPUT HIDDEN MARKOV MODEL-BASED SOLUTION," *Proc Inst Mech Eng O J Risk Reliab*, vol. 237, no. 5, pp. 980–993, Oct. 2023, doi: https://doi.org/10.1177/1748006X221119853.
[13] N. Hayati, M. Kiftiah, and B. Prihandono, "APPLICATION OF THE JUKES CANTOR MODEL IN DETERMINING THE PROBABILITY OF NITROGEN BASES IN THE OFFSPRING OF AN INDIVIDUAL (APLIKASI MODEL JUKES CANTOR DALAM MENENTUKAN PELUANG BASA NITROGEN KETURUNAN SUATU INDIVIDU)," *Bimaster: Buletin Ilmiah Matematika, Statistika dan Terapannya*, vol. 5, no. 2, pp. 119–128, 2016, doi: https://doi.org/10.26418/bbimst.v5i02.15843.
[14] N. H. Mohd Noor and M. J. Siti Asmaa, "KARL LANDSTEINER (1868–1943): A VERSATILE BLOOD SCIENTIST," *Cureus*, Sep. 2024, doi: https://doi.org/10.7759/cureus.68903.
[15] H. F. Jelinek *et al.*, "ALLELIC VARIANTS WITHIN THE ABO BLOOD GROUP PHENOTYPE CONFER PROTECTION AGAINST CRITICAL COVID-19 HOSPITAL PRESENTATION," *Front Med (Lausanne)*, vol. 8, Jan. 2022, doi: https://doi.org/10.3389/fmed.2021.759648.
[16] P. Vasisth, O. M. Limbalkar, and M. Sharma, "GENETICS OF MULTIPLE ALLELES: CONCEPT AND FUNCTION," *Advances in Genetic Polymorphisms*, Jul. 2023, doi: https://doi.org/10.5772/intechopen.1001464.
[17] S. Strome, N. Bhalla, R. Kamakaka, U. Sharma, and W. Sullivan, "CLARIFYING MENDELIAN VS NON-MENDELIAN INHERITANCE," *Genetics*, vol. 227, no. 3, pp. 1–7, Jul. 2024, doi: https://doi.org/10.1093/genetics/iyae078.
[18] T. A. Khan, G. K. Nanjundan, D. M. Basvarajaih, and M. Azharuddin, "STATISTICAL MODEL DERIVATION AND EXTENSION OF HARDY – WEINBERG EQUILIBRIUM," *Int J Curr Microbiol Appl Sci*, vol. 7, no. 10, pp. 2402–2409, Oct. 2018, doi: https://doi.org/10.20546/ijcmas.2018.710.279.
[19] R. Waples, "TESTING FOR HARDY–WEINBERG PROPORTIONS: HAVE WE LOST THE PLOT?," *Journal of Heredity*, vol. 106, no. 1, pp. 1–19, 2015, doi: https://doi.org/10.1093/jhered/esu062.
[20] V. Kulkarni, *MODELING AND ANALYSIS OF STOCHASTIC SYSTEMS*. Chapman and Hall/CRC, 2016.
[21] M. Awad and R. Khanna, "HIDDEN MARKOV MODEL," in *Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers*, Berkeley: CA: Apress, 2015, ch. 5, pp. 81–104. doi: https://doi.org/10.1007/978-1-4302-5990-9_5.
[22] R. K. Perdana, "INDONESIAN POPULATION DATA BASED ON BLOOD TYPE IN 2024 (DATA JUMLAH PENDUDUK INDONESIA BERDASARKAN GOLONGAN DARAH PADA 2024)," Dinas Kependudukan dan Pencatatan Sipil (Dukcapil) Kementerian Dalam Negeri (Kemendagri). Accessed: Jun. 11, 2025. [Online]. Available: https://dataindonesia.id/varia/detail/data-jumlah-penduduk-indonesia-berdasarkan-golongan-darah-pada-2024
[23] Kementerian Dalam Negeri Republik Indonesia, "DIRECTOR GENERAL OF CIVIL REGISTRATION: INDONESIA HAS A DATABASE OF 37.9 MILLION BLOOD TYPES (DIRJEN DUKCAPIL: INDONESIA MILIKI BANK DATA 37,9 JUTA GOLONGAN DARAH)," Direktorat Jenderal Kependudukan dan Pencatatan Sipil. Accessed: Feb. 20, 2025. [Online]. Available: https://dukcapil.kemendagri.go.id/blog/read/dirjen-dukcapil-indonesia-miliki-bank-data-379-juta-golongan-darah
[24] Kementerian Dalam Negeri Republik Indonesia, "THE MINISTRY OF HOME AFFAIRS' CIVIL REGISTRATION OFFICE URGES THE PUBLIC TO UPDATE THEIR BLOOD TYPE TO HELP OTHERS (DUKCAPIL KEMENDAGRI IMBAU MASYARAKAT UPDATE GOLONGAN DARAH AGAR BISA BANTU SESAMA)," Direktorat Jenderal Kependudukan dan Pencatatan Sipil. Accessed: Feb. 24, 2025. [Online]. Available: https://dukcapil.kemendagri.go.id/blog/read/dukcapil-kemendagri-imbau-masyarakat-update-golongan-darah-agar-bisa-bantu-sesama