

CHAPTER I

INTRODUCTION

1.1 Research Background

Dengue Fever (DF) is a disease caused by the dengue virus (DENV), primarily transmitted through the bites of infected mosquito. *Aedes aegypti* is one of the mosquitoes that can carry the DENV. This dengue virus has four unique strains, ranging from DENV-1 to DENV-4, which typically carried by mosquitoes that are part of the Flaviviridae family, such as *Aedes polynesiensis*, *Aedes scutellaris*, and *Aedes albopictus* (Lusno et al., 2023). This disease is known as the most common arthropod-borne viral disease in humans (Pourzangiabadi et al., 2025). The virus will incubate inside the infected person for several days until the symptoms appear.

Arthropod-borne viruses or Arbovirus, spreads to people through the bites of mosquitoes, ticks, sand flies, and other arthropods, where they are predominantly seen in tropical and sub-tropical regions (State of Hawaii, 2024). The infections will likely to occur during warm and humid periods of the year, specifically during the rainy season. This is when mosquitoes are most active.

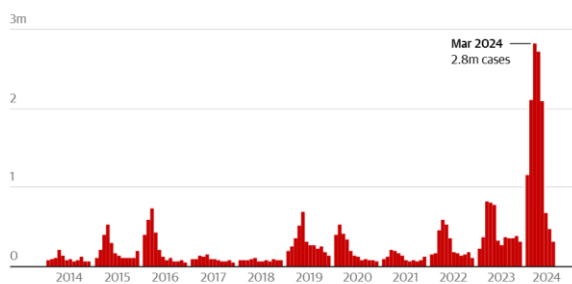
In general, *Flavivirus* infections spread by mosquitoes can cause a wide range of symptoms from a low-grade fever and joint pain (arthralgia) to acute and severe complications such as liver, kidney, and brain damage. On the other hand, DENV infections range from a life-threatening hemorrhagic and capillary leak condition known as dengue hemorrhagic fever (DHF) or dengue shock syndrome (DSS) to an acute, self-limiting, disabling illness (Pourzangiabadi et al., 2025).

Dengue fever typically begins three to fourteen days after the bite of an infected mosquito, with symptoms such as fever, muscle pain, frontal headache, joint pain, and rash. Many non-medical individuals struggle to distinguish dengue fever from common fevers, leading to delayed treatment and increased transmission risks. Without an early and proper treatment, the disease can progress to dengue hemorrhagic fever, a severe and potentially fatal form of dengue (Divya Jacob, 2023).

As of 30 April 2024, over 7.6 million dengue cases have been reported by the World Health Organization (WHO), around 44.74% of those were confirmed, 0.21% of severe cases, including more than 3,000 deaths, where half of the world's population is at risk of contracting DENV, which is mostly prevalent in tropical and subtropical urban. Indonesia, being the most populous country in dengue-endemic Southeast Asia, has consistently ranked at the top with the most dengue cases. The Indonesian Ministry of Health noted that throughout 2024, Indonesia experienced 244,409 cases of dengue fever with 0.585% of those resulting in deaths. Until January 23, 2025, the top six regions with the highest number of cases in Indonesia being: NTT, East Java, Bali, Lampung, West Java, and DKI Jakarta (CNN Indonesia, 2025). This puts Indonesia as a dangerous breeding ground for the disease.

Global cases of dengue fever rose steeply in 2024

Monthly global cases, millions



Guardian graphic. Source: WHO. Note: case reporting requirements vary by country

Figure 1.1 Global Dengue Virus Spread (2014-2024)

Dengue fever cases in 2024

0 cases 1 500 5,000 50,000+ No data



Guardian graphic. Source: WHO. Note: case reporting requirements vary by country

Figure 1.2 Dengue Fever Cases in 2024

Source: *The Guardian Graphic* (2024)

Source: *The Guardian Graphic* (2024)

As of February 16, 2025, a total of 10,752 cases had been reported nationwide. The Incidence Rate (IR) was approximately 3.79 per 100,000 population, with 48 deaths resulting in a Case Fatality Rate (CFR) of 0.48 (M Iqbal Al Machmudi, 2025). Bali has one of the highest historical dengue fever IR and is considered one of the top provinces with the highest number of dengue fever cases in Indonesia as of 2024. Within Bali, Buleleng Regency is ranked third in cases with 1,947 incidences of dengue. However, the actual number is likely greater as more mild infections, approximately thousand cases went unreported (The Bali Sun, 2025).

One of the challenges in solving the root cause of dengue fever in society largely comes from the lack of awareness. Those with subprime awareness are likely to ignore and not report symptoms, nor any intention to treat them, which speeds up the infection and reduces the already-thin chances of survival. Dengue fever shares overlapping symptoms with other diseases such as influenza, chikungunya, and other common fever conditions, including febrile conditions, most notably high body temperature. This makes it difficult to have an accurate diagnosis without any prior laboratory testing. Therefore, education and awareness are crucial for identifying symptoms and providing early treatment. To support this, machine learning techniques can be utilized to develop a tool that can help early detect the dengue fever symptoms, specifically by classifying whether a patient is likely to have dengue fever or not based on the symptoms occurring. This model is built upon the common dengue fever symptoms which it analyzes to make inferences regarding the binary status of the dengue fever in patient's body. This supports faster preliminary assessments before laboratory test confirmation is available.

In recent years, with the rapid advancement of technology and the implementation of machine learning techniques, several studies have applied these techniques to predict or classify infectious diseases, yielding promising results. However, the conditions of real-world data are not always good. In order to achieve promising results, high-quality and comprehensive data are needed. Unfortunately, in the real world, the data is often incomplete, biased, noisy, and imbalanced, which might hinder the performance of the models used. Therefore, to address this challenge, it is necessary to evaluate more than one model within the same case focus to identify which model delivers the best results and works well with this dataset's characteristics.

This study will compare two ensemble learning models to identify which yields better performance. Those two models, namely Random Forest and Gradient Boosting, which are known to be used for classification and regression tasks. Random Forest is an ensemble method that is known to provide more stable results and minimize overfitting because it consists of multiple decision trees that uses a

bagging (bootstrap aggregating) method where each tree is trained on a random subset of the data and feature (Win et al., 2024). This algorithm works well for incomplete data or handles missing values within the datasets. Gradient Boosting is also an ensemble method that combines several weak learners to form a strong learner and builds trees sequentially, each new tree attempting to correct the errors of its predecessor by minimizing a loss function (Win et al., 2024).

This study aims to support early detection of dengue fever by helping non-medical communities recognize its symptoms through an intuitive, symptom-based interface. The dataset used in this study is independent, where each patient's record represents a one-time condition only. Therefore, it is not considered time-series data because it does not consist of periodic data that is recorded continuously over a certain period. The system performs a binary classification to determine whether a patient is likely to have dengue, based on inputted symptoms observed at the triage stage and before the laboratory test confirmation is available.

Various prior studies demonstrated the implementation of Random Forest and Gradient Boosting. Study by Sarma et al. (2020), compared the Random Forest (RF) and Decision Tree (DT) models on dengue patient cases obtained 79% accuracy for DT and 74% for RF. Similarly, the study by Veena Kumari et al. (2022) achieved 98.79% for the proposed model, which is RF with GridSearchCV. Pohan et al. (2022) achieved 88.41% accuracy for RF as the highest model results amongst the model used in the research. While, the study by Sarwar & Al Mamun (2022) showed XGBoost performed better (92%) than Logistic Regression (89%) but Support Vector Machine still achieved a slightly better result (97%). On the other hand, the study by Nsanzabandi & Nizeyimana (2024) with a different focus, which is Malaria disease, showed the comparison of RF and GB. In this study, GB (98.20%) slightly outperforms RF (97.81%). These prior studies have demonstrated the effectiveness of both RF and GB in disease prediction and classification. RF showed better performance by outperforming the other models. However, GB also showed great results when compared to RF in the malaria case, which has quite similar characteristics to dengue disease. Therefore, this study will do a comparison analysis for both RF and GB in the early detection of dengue.

1.2 Problem Identification

According to the background that has been explained in the previous section, there are several problems raised and have been identified in this study, as follows:

1. As of February 2025, the incidence rate remained high, which was 3.79 per 100,000 population, with 48 deaths resulting in a case fatality rate of 0.48. This number highlights the urgent need for improvement in early detection the dengue to reduce the risk of severe cases and fatalities.
2. There is a lack of public awareness regarding the early symptoms of dengue fever. The symptoms typically appear from the third to the fourteenth day after the first bite of the mosquito, resulting in people often mistakenly thinking they are having a common fever. This misperception contributes to delays in seeking medical help and increases the likelihood of more severe complications if not properly and promptly treated.
3. Several prior studies have been applying machine learning to medical diagnostics. However, few studies have conducted a comparative analysis between Random Forest and Gradient Boosting for early detection using non-laboratory data. Most of the prior studies also focus on patients with severe or confirmed dengue, leaving a gap in classified models that use the early-stage symptoms, which is crucial to prevent the case leads to more severe conditions.

1.3 Research Question

With those problem identifications, there are some problems obtained and will be discussed further in this study, including:

1. How to implement RF and GB for the early detection of dengue fever?
2. How do the RF and GB algorithms perform in modeling early detection of dengue fever, and what are the resulting outcomes?
3. How can the resulting model be accessible and interpretable for non-medical users?

1.4 Research Limitation

In order to overcome the problem stated previously, this study needs to have a clear scope and focus, so that the research can be carried out more specifically. The limitations for this study are as follows:

1. The clinical records obtained from the triage form at *Rumah Sakit Umum Daerah* (RSUD) Buleleng will be used as input features, while lab test results will determine the dataset labels. Lab parameters themselves are excluded from the feature set to ensure the model supports early-stage classification before laboratory test confirmation is available.
2. The datasets focused on dengue fever cases from January 2025 in Buleleng regency.
3. This study will compare two supervised machine learning algorithms, namely Random Forest and Gradient Boosting.
4. The study does not include real-time vector surveillance or non-clinical predictive factors like weather or mosquito population trends.
5. The temperature of the patient's body will be used in degrees Celsius.
6. The strategy of feature selection evaluation metrics will be used accuracy, precision, recall, and F1-score to assess model performance. With recall will be the main prioritized.
7. The data preprocessing, modeling, and evaluation of this study will be conducted using Python Environment and use Google Colab as a tool.
8. The results of this study will be limited to only the results of the performance of Random Forest and Gradient Boosting to compare and find which models give the best outcomes for this case study, as well as an intuitive and simple interface that integrates with the model.

1.5 Research Objective

This study aims to evaluate the performance of ensemble machine learning algorithms in the early detection of dengue fever, based on clinical datasets obtained from the triage form medical records at *Rumah Sakit Umum Daerah* Buleleng. The models will classify whether the patient is likely to have dengue fever or not based on the inputted symptoms that are observable for the patients prior to the laboratory test. The objective is to support faster preliminary assessments for the patients and prevent the symptoms from developing into more severe conditions. This study will

help in evaluating the effectiveness of machine learning by utilizing several evaluation metrics to evaluate the model, including accuracy, precision, recall, and F1-score. However, this study does not prioritize results over accuracy because the goal here is to catch as many cases of most likely dengue as possible so that these patients immediately receive the appropriate and proper treatment. Furthermore, the model will be deployed in a website-based system with a simple interface that can be used as a tool for early detection of dengue fever by inputting values for each symptom feature. The system will then process and provide classification results based on the user's input.

1.6 Research Significance

1.6.1 Theoretical Significance

1. For Readers and Future Researchers
 - a. Provides a comparative analysis of two machine learning algorithms, Random Forest and Gradient Boosting, about the results in the early detection of dengue disease.
 - b. Offers insights into feature importance for dengue classification, aiding future model improvements.
 - c. Serves as a reference for future research and studies on machine learning applications in the medical and healthcare sectors, particularly in classifying and detecting diseases.
2. For Universitas Pendidikan Ganesha (Institution)
 - a. Enhance the university's reputation in applied machine learning research and its role in addressing public health challenges.
 - b. Supports institution improve the quality of graduates and develop an academic culture towards a superior and competitive campus, it is necessary to improve the quality and quantity of scientific work for lecturers and students.

1.6.2 Practical Significance

1. For *Rumah Sakit Umum Daerah* Buleleng (Hospital and Healthcare Sector)
 - a. Supports the hospital in classifying dengue trends, allowing proactive response to outbreaks.

- b. Helps the hospital integrate machine learning-based diagnostic support tools into its system, improving decision-making efficiency.
- c. Assists medical professionals in diagnosing dengue fever earlier and more accurately, potentially reducing misdiagnosis.

2. For Society

- a. Increases public awareness by helping non-medical communities recognize the early signs and symptoms of dengue fever, encouraging earlier medical consultation.
- b. Empowers communities with accessible information, minimizing confusion between dengue and common fevers, which often leads to late or incorrect treatment.
- c. Providing an early detection tool that estimates their likelihood of having dengue fever, which enables faster response and medical attention, reducing severe cases and preventing outbreaks.

