

COMPARATIVE ANALYSIS OF CLASSICAL MACHINE LEARNING AND QUANTUM MACHINE LEARNING FOR PREDICTING DENGUE OUTBREAKS IN NEPAL: AN SEIR-VECTOR BASED APPROACH



BSC project Proposal

Submitted to
Department of Physics
Tri-Chandra Multiple Campus

By

Roshan Adhikari

Roll No: 706

Supervisor
Arjun Acharya
Assist. Professor
Department of Physics
Tri-Chandra Multiple Campus

December 19, 2025

Contents

1	Introduction	3
1.1	Background	3
1.2	Problem Statement	3
1.3	Objectives	4
2	Literature Review	5
2.1	Epidemiological Modeling	5
2.2	Machine Learning in Disease Forecasting	5
2.3	Quantum Machine Learning	5
3	Theoretical Framework	6
3.1	The Biophysical SEIR-Vector Model	6
3.1.1	Human Population Dynamics	6
3.1.2	Vector Population Dynamics	6
3.2	Classical ML: Random Forest	6
3.3	Quantum ML: Quantum Support Vector Machine	7
4	Methodology	8
4.1	Tools and Environment	8
4.2	Data Sources	8
4.3	Research Plan	8
5	Expected Results	10
5.1	Simulation	10
5.2	Comparative Analysis	10
6	Project Timeline	11
7	Estimated Budget	11

List of Figures

1	Graph Showing the historical trend of dengue virus cases in Nepal (2004 - 2024)	3
2	SEIR Vector Model Diagram	5
3	Proposed Research Workflow Methodology	9

1 Introduction

1.1 Background

In Nepal, dengue fever, a virus spread by mosquitoes, has quickly become a significant public health concern. Due to urbanisation and climate change, the virus, which was once limited to the tropical lowlands (Terai), has spread to higher altitudes, including the Kathmandu Valley. The main vectors, *Aedes aegypti* and *Aedes albopictus*, are extremely sensitive to thermodynamic factors, especially temperature, which controls the duration of the virus incubation period and their reproduction cycles.

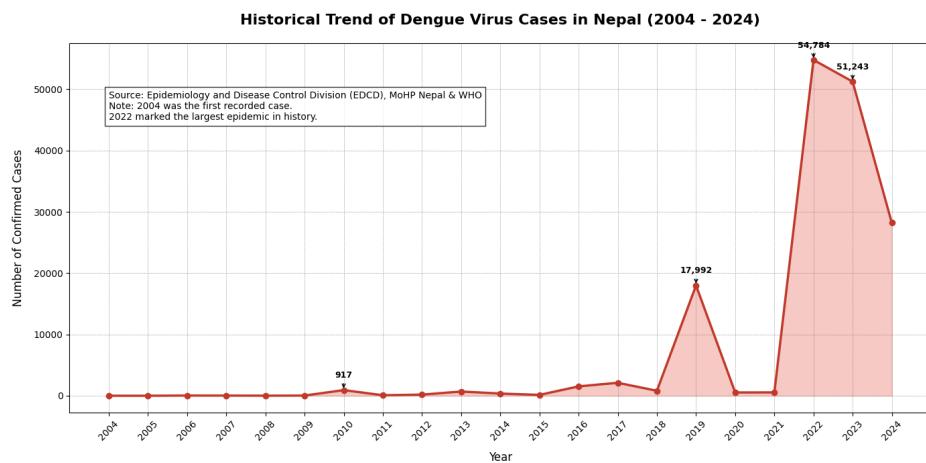


Figure 1: Graph Showing the historical trend of dengue virus cases in Nepal (2004 - 2024)

Classical Machine Learning (CML) methods, such as Random Forest, have been extensively used in predictive modelling to anticipate disease outbreaks using meteorological data. However, traditional approaches may encounter computational constraints as epidemiological datasets become more complex and multidimensional. Using quantum mechanical concepts such as superposition and entanglement to process data in high-dimensional feature spaces, quantum machine learning (QML), and more especially, algorithms such as the Quantum Support Vector Machine (QSVM) that offers a revolutionary method.

1.2 Problem Statement

Nepal's current surveillance systems are primarily reactive, depending on cases reported by hospitals after an outbreak has already started. By neglecting the vector population and the time-lagged impact of temperature on viral incubation, standard mathematical models, like the basic SIR model, frequently oversimplify the dynamics.

Although disease forecasting has been enhanced technologically by Classical Machine Learning (CML) techniques like Random Forest, these techniques may not be able to properly resolve the complex, non-linear connections present in environmental epidemiological data because they operate within limited feature spaces. In theory, a solution is provided by Quantum Machine Learning (QML), which maps data into high-dimensional Hilbert spaces to find subtle correlations that are missed by classical kernels. It is still unclear, nevertheless, if these theoretical benefits result in useful prediction improvements for epidemiology. There are currently few comparison studies that thoroughly assess how well quantum versus classical algorithms predict dengue outbreaks in the particular setting of Nepal.

1.3 Objectives

Based on a biophysical knowledge of the disease, the main goal of this study is to assess the predictive power of Quantum Machine Learning models as compared to Classical Machine Learning models for dengue outbreaks. The particular goals are:

1. To simulate Dengue transmission dynamics using a **Biophysical SEIR-Vector Model** that accounts for temperature-dependent incubation periods.
2. To implement a Classical **Random Forest Regressor** to predict infection trends using historical weather and case data.
3. To implement a **Quantum Support Vector Machine (QSVM)** using a quantum kernel estimator to perform the same prediction task.
4. To perform a comparative analysis of both models based on prediction accuracy (RMSE), computational cost, and ability to generalize from limited data.

2 Literature Review

2.1 Epidemiological Modeling

The mathematical foundation of epidemiology was laid by Kermack and McKendrick (1927) with the SIR model [1]. However, this model is insufficient for vector-borne diseases. Esteva and Vargas (1998) expanded this to a coupled Vector-Host model, which explicitly accounts for the mosquito population [2]. This project builds on their framework by integrating the thermodynamic findings of Watts et al. (1987), who demonstrated that the extrinsic incubation period of the Dengue virus is strictly temperature-dependent [3]. This model was termed as the SIER Vector Model.

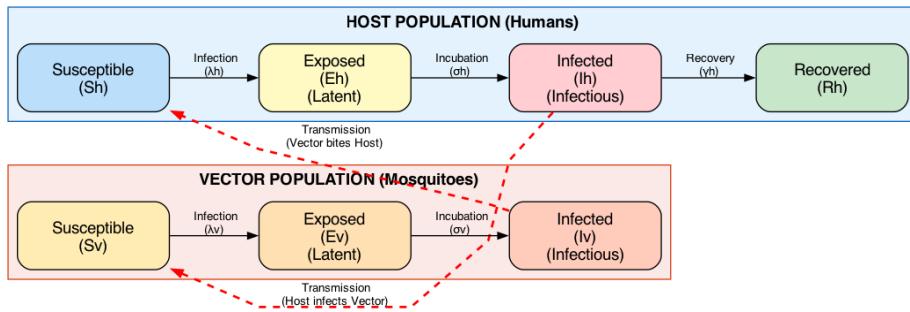


Figure 2: SEIR Vector Model Diagram

2.2 Machine Learning in Disease Forecasting

For outbreak forecasting, Random Forest has become a reliable technique. Benedum et al. (2019) demonstrated that ensemble approaches perform better than basic linear regressions when handling noisy environmental data by successfully using Random Forest to forecast Dengue cases in Peru. [4]. Dhimal et al. (2015) found an association between vector distribution and climate factors in Nepal, however these investigations were primarily statistical rather than predictive machine learning. [5].

2.3 Quantum Machine Learning

A new field called quantum machine learning (QML) uses quantum processors to analyse data. By introducing the idea of the "Quantum Kernel," Havlíček et al. (2019) showed how quantum computers may effectively compute inner products in Hilbert spaces that are computationally impossible for classical computers. [6]. This feature implies that QSVMs may be able to spot minute non-linear trends in epidemiological data that traditional kernels could overlook as 'noise'.

3 Theoretical Framework

3.1 The Biophysical SEIR-Vector Model

The host (human) and vector (mosquito) populations are modelled in this work using a system of coupled Non-Linear Ordinary Differential Equations (ODEs).

3.1.1 Human Population Dynamics

The human population is divided into Susceptible (S_h), Exposed (E_h), Infected (I_h), and Recovered (R_h) compartments:

$$\frac{dS_h}{dt} = -\beta_h S_h I_v \quad (1)$$

$$\frac{dE_h}{dt} = \beta_h S_h I_v - \sigma_h E_h \quad (2)$$

$$\frac{dI_h}{dt} = \sigma_h E_h - \gamma I_h \quad (3)$$

$$\frac{dR_h}{dt} = \gamma I_h \quad (4)$$

3.1.2 Vector Population Dynamics

The mosquito population includes Susceptible (S_v), Exposed (E_v), and Infected (I_v) compartments, with logistic growth:

$$\frac{dS_v}{dt} = \Lambda_v - \beta_v S_v I_h - \mu_v S_v \quad (5)$$

$$\frac{dE_v}{dt} = \beta_v S_v I_h - \sigma_v E_v - \mu_v E_v \quad (6)$$

$$\frac{dI_v}{dt} = \sigma_v E_v - \mu_v I_v \quad (7)$$

3.2 Classical ML: Random Forest

Random Forest is a bagging ensemble method. It aggregates the predictions of B individual decision trees:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B f_b(x) \quad (8)$$

Its capacity to manage non-linear interactions between weather variables (such as temperature and humidity) and its resilience against overfitting are the reasons it was selected.

3.3 Quantum ML: Quantum Support Vector Machine

The QSVM maps classical data \mathbf{x} into a quantum state $|\Phi(\mathbf{x})\rangle$ using a feature map. The kernel function is evaluated as the overlap between two quantum states:

$$K(\mathbf{x}_i, \mathbf{x}_j) = |\langle \Phi(\mathbf{x}_i) | \Phi(\mathbf{x}_j) \rangle|^2 \quad (9)$$

The best separating hyperplane (or regression line) in the high-dimensional feature space is then found using this "Quantum Kernel" in a traditional SVM optimisation procedure.

4 Methodology

4.1 Tools and Environment

The research will be conducted on a MacOS-based system using Python.

- **Simulation:** SciPy (`solve_ivp`) for integrating the SEIR ODEs.
- **Classical ML:** Scikit-Learn for Random Forest implementation.
- **Quantum ML:** PennyLane for creating quantum circuits and kernel estimation.

4.2 Data Sources

1. **Epidemiological Data:** Dengue case counts from the Epidemiology and Disease Control Division (EDCD), Nepal.
 - **Timeframe Selection:** Despite the fact that dengue infections have been documented since 2004, data from 2004 to 2014 shows a low-transmission phase with few, irregular cases. This study will concentrate on the endemic period from 2014 to 2024 in order to guarantee statistical significance and prevent zero-inflation bias in the machine learning models.
 - **Temporal Resolution:** Epidemiological case data will be aggregated on a weekly basis, while raw data is available every day. This frequency is selected to maintain a suitable sample size (≈ 520 weeks) for Quantum Machine Learning training while matching the biological incubation time of the virus (around 10–14 days).
2. **Meteorological Data:** Daily temperature (T_{max}, T_{min}), rainfall, and humidity obtained from the NASA POWER database.

4.3 Research Plan

- **Data Preprocessing:** Cleaning data, handling missing values, and engineering lag features (e.g., Temperature 14 days ago).
- **SEIR Calibration:** Tuning the β (transmission) and σ (incubation) parameters of the mathematical model.
 - **Multi-Scale Simulation Strategy:** The Biophysical SEIR-Vector model will be simulated using a daily time-step ($dt = 1$ day) to accurately capture the temperature-dependent mosquito lifecycle. The output of this daily

simulation will then be aggregated into weekly totals to match the real-world case data and the input format required for the Machine Learning models.

- **Model Training:**

- Train Random Forest on 80% of the weekly-aggregated dataset.
- Train QSVM using a feature map on the same training set.

- **Evaluation:** Compare Root Mean Square Error (RMSE) and R-Squared (R^2) values on the 20% test set.

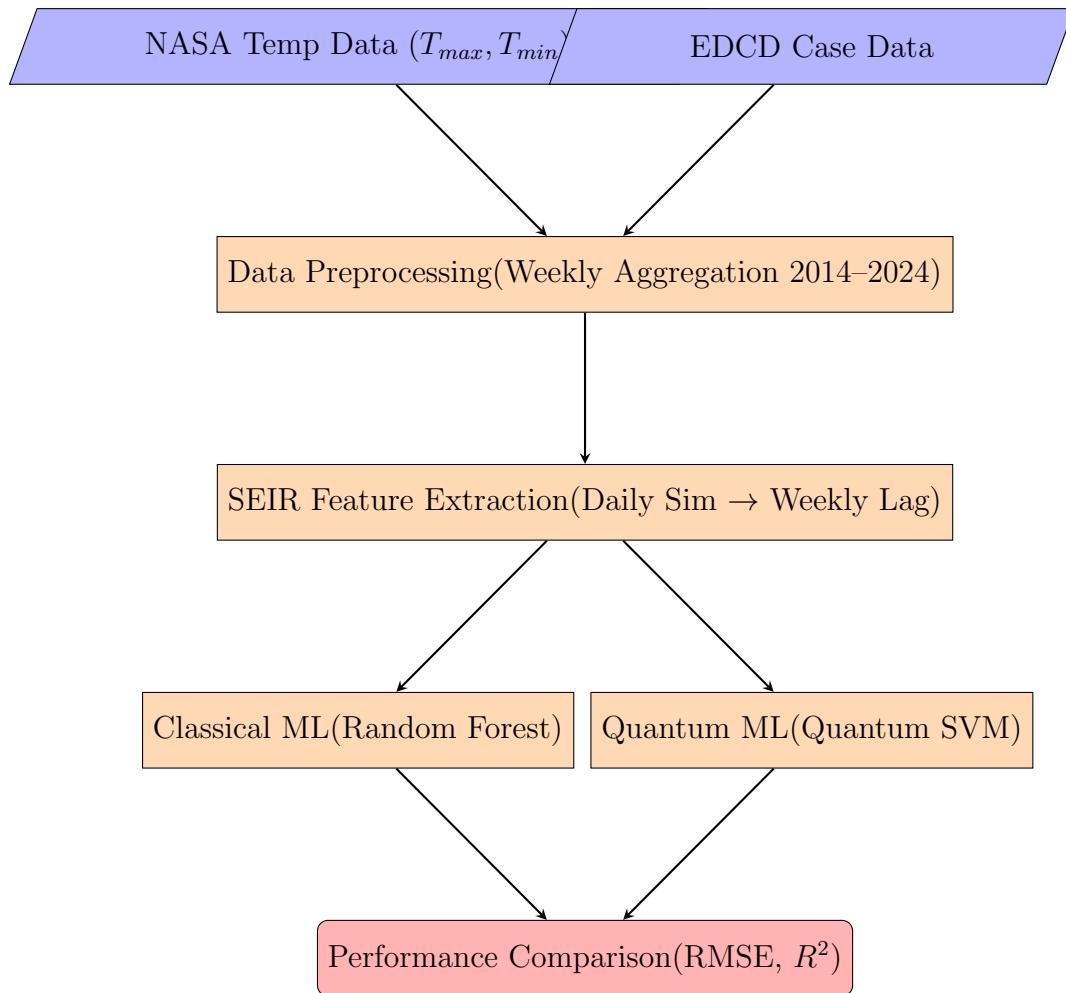


Figure 3: Proposed Research Workflow Methodology

5 Expected Results

5.1 Simulation

It is anticipated that the SEIR-Vector model will show the lag between peak temperature and peak infections, confirming the virus's thermodynamic mechanism.

5.2 Comparative Analysis

Because of its maturity, we expect the Random Forest model to offer a high baseline accuracy. In situations with high feature complexity, the Quantum SVM is expected to meet or somewhat surpass this accuracy.

Metric	Random Forest	Quantum SVM
Training Time	Fast	Slower (Simulated)
Accuracy (R^2)	High (> 0.85)	To be evaluated
Data Efficiency	Moderate	Potentially High

Table 1: Framework for comparing Classical vs Quantum models.

6 Project Timeline

Task / Month	Month 1	Month 2	Month 3	Month 4
Literature Review	X			
Data Collection	X			
SEIR Modeling		X		
ML & QML Coding		X	X	
Thesis Writing			X	X
Final Defense				X

Table 2: Proposed project schedule.

7 Estimated Budget

- **Internet & Data:** NRs. 3,000
- **Cloud Resources:** NRs. 2,000
- **Printing & Binding:** NRs. 4,000
- **Miscellaneous:** NRs. 1,000
- **Total Estimated Cost:** NRs. 10,000

References

- [1] Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A*, 115(772), 700-721.
- [2] Esteva, L., & Vargas, C. (1998). Analysis of a dengue disease transmission model. *Mathematical Biosciences*, 150(2), 131-151.
- [3] Watts, D. M., Burke, D. S., Harrison, B. A., Whitmire, R. E., & Nisalak, A. (1987). Effect of temperature on the vector efficiency of Aedes aegypti for dengue 2 virus. *American Journal of Tropical Medicine and Hygiene*, 36(1), 143-152.
- [4] Benedum, C. M., Shea, K. M., Jenkins, H. E., Kim, L. Y., & Markuzon, N. (2019). Weekly dengue forecasts in Iquitos, Peru; predictive modeling with meteorological and epidemiological data. *PLOS Neglected Tropical Diseases*, 13(10), e0007790.
- [5] Dhimal, M., Ahrens, B., & Kuch, U. (2015). Climate change and spatiotemporal distributions of vector-borne diseases in Nepal. *Parasitology Research*, 114(8), 2817-2828.
- [6] Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., & Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747), 209-212.
- [7] Adhikari, S. R., & Supakankunti, S. (2020). Economic burden of Dengue in Nepal. *Journal of Nepal Health Research Council*, 18(3), 475-480.