International Journal of Innovative Mathematics, Statistics & Energy Policies
13(4):142-153, Oct-Dec., 2025

 www.seahipublications.org **ISSN: 2467-852X**

doi:10.5281/zenodo.17614675

Development And Validation Of An AI-Driven Early-Warning System For Vaccine-Induced Thrombosis Using Temporal Symptom Dynamics

Nuruddeen Hassan Danrimi[1]; Ibrahim Mamuda Kukasheka [2]; Hafiza Inusa Idris[3].

[1]Department of Statistics, University of Abuja, Nigeria

[2]Department of Mathematics and Statistics, Kaduna Polytechnic, Nigeria

[3]Department of Statistics, Federal Polytechnic Nyak Shendam, Plateau State

Author's emails: nuruddeen.hassan@uniabuja.edu.ng[1]; mamudy2k@gmail.com[2] ; halilu.hafiza@gmail.com[3]

ABSTRACT

Although essential to the control of the pandemic, COVID-19 vaccinations have been linked to a series of rare but severe adverse events such as vaccine-induced immune thrombotic thrombocytopenia (VITT) especially with adenoviral vaccines. The current predictive models do not contain vaccine-platform disparities, thrombosis-specific clinically validated predictors, and do not account for critical temporal dynamics. The paper creates an AI-based early-warning system of post-vaccination thrombotic events by incorporating major predictors of time and clinical factors. A 2022-2023 retrospective dataset of 200 patients of the Vaccine Adverse Event Reporting System (VAERS) was studied by multimodal analysis, including time-to-event and vaccine-type comparison. We found the strongest predictors of thrombosis to be the onset of the symptoms within 5 days, AstraZeneca vaccination, age over 50, and chest pain. To predict the risk of thrombosis, we trained and compared several AI models, namely LSTM, GRU, Bidirectional LSTM, and Random Forest. Although the neural network models (especially ANN, with 89.47% accuracy and 83.33% recall) had high efficacy, the Random Forest model had serious weaknesses, such as low sensitivity (28.57%). The research has been effective in filling a considerable gap of turning generic adverse event prediction into a thrombosis-specific early-warning device that offers a clinically amenable framework to risk stratification and prioritized intervention of high-risk subgroups.

Keywords: Thrombosis Prediction, AI/ML Models, Vaccine Safety, Temporal Dynamics, Risk Stratification

Introduction

The COVID-19 vaccines have been developed and implemented worldwide rapidly, which marks a pivotal breakthrough in the history of immunization that has saved 2 -3 million lives annually on average (Pollard *et al.*, 2021). The mRNA-based vaccines were one of the platforms that were an innovative method of technology, characterized by a short time of development and a positive safety profile due to the non-infectious nature and the absence of attachment to the host genome (Chaudhary et al., 2021).

Mayo Clinic. (2025) claimed that with the illustration of security and effectiveness in massive clinical trials these vaccines were granted emergency use authorization and then full approval by regulatory organizations such as the U.S. Food and Drug Administration. Although the safety profile is very good, the detection of adverse events that are rare but severe, including vaccine-induced immune thrombotic thrombocytopenia (VITT), once again reveals the paramount role played by strong pharmacovigilance. This post-marketing surveillance is necessary and can be facilitated by systems such as the Vaccine Adverse Event Reporting System (VAERS), although the system may be disabled by delays in reporting and under-reporting to predict any risk of a rare event in real-time (Hibbs *et al.,* 2022). This disconnect signifies the pressing necessity of highly advanced and predictive AI models capable of incorporating the information on time-based symptoms to provide an opportunity to intervene clinically at an earlier stage.

Despite lot of research on investigating the effects of vaccinations, however, existing pharmacovigilance systems such as VAERS (vaccine adverse event reporting system) continue to be hampered by delays in signal detection as well as underreporting (Shimabukuro *et al.,* 2022), or rather, clinical studies have been largely restricted to;

Descriptive epidemiology of risk factors associated with thrombosis, Immunological etiology of VITT, and Large-scale surveillance of adverse events without detailed time profiling (Greinacher *et al.,* 2021). For instance, Li *et al*. (2022) developed an ANN model to predict mild-moderate adverse events, where study ignores severe outcomes like thrombosis. The study exclusively focused on non-severe outcomes, ignoring life-threatening implications like thrombosis. The study also used static demographic features without incorporating temporal symptoms progression. The study also lacks vaccine-type stratification despite known risk disparities.

Additionally, Zhang *et al*. (2022) Used LSTM only for myocarditis prediction without adequately comparing architectures or addressing thrombosis. The study ignored thrombosis despite its clinical significance in adenoviral vaccines and also used limited clinical predictors by omitting symptoms onset-timing

Moreover, Rustam *et al.* (2020) forecast COVID-19 transmission trends using Relied simple models (Linear Regression, SVM) where the study lacks temporal analysis for symptom progression. And model could not capture dynamic symptom evolution critical for adverse event prediction.

Static neural networks (PNN/RBFNN) were also developed by Dhamodharavadhani *et al*. (2020) for mortality prediction without vaccine-type stratification. The study does no address vaccine specific risk profiling despite emerging evidence of platform dependent safety signals and does not address adverse event prediction, where study exclusively relies on mortality outcomes.

In own contribution, Guo *et al.* (2020) conducted virological analysis comparing SARS-CoV-2 gene sequences to SARS-CoV and MERS-CoV. The study focused on purely mechanism study without clinical risk predictions. Additionally, the study did not apply AI/ML components for patient level risk stratification.

In our recent study Fika *et al.* (2025) in predicting and classifying the potential symptoms that may arise post-COVID-19 vaccination using AI/ML hybrid models (ANN, BiLSTM), however, the study fails to carryout clinical validated thrombosis-specific predictors.

More importantly, there are two evidence gaps in prior studies regarding vaccine adverse effects:

i. The evidence on the association between symptom onset timing, duration, and thrombosis risk is not known; this can be referred to as Temporal Dynamics.
ii. Vaccine-Specific Risk Stratification as there is a lack of a validation of differential risk profiles across various types of vaccine platforms (e.g., AstraZeneca vs. mRNA vaccines) in real-world clinical trajectories (CDC, 2022).

Thus, this current study tends to fill these gaps by utilizing a multimodal analysis of 200 patients included in VAERS (2022-2023), which is same data used in prior study of (fika *et al.,* 2025) with the incorporation of time-to-event measures (time of onset of the symptoms, duration), vaccine-type comparison, and risk modeling based on AI (ANN/LSTM), to determine clinically viable predictors of thrombosis.

Statement of the Problem

Although current AI/ML models (eg, Fika *et al.,* 2025), show imminent potential to forecast post-vaccination adverse reactions, there lies a fatal gap in translation: they lack the inclusion of clinically validated thrombosis-specific predictors confirmed in empirical research. In particular, these models apply generic VAERS characteristics without emphasizing the occurrence of symptoms 0-5 days after vaccination, variations between vaccines and determining the crucial signs that will provoke the issue with thrombosis complications. This gap restricts their clinical utility to early and targeted intervention in subgroups at high risk, where dependence on reactive pharmacovigilance systems is poorly prepared to prevent severe outcomes. in this respect, the current study tends to address the gap by examining the predictors specific to thrombosis.

Aim

To develop and validate an AI-based early-warning system for predicting post-vaccination thrombotic events by integrating key temporal and clinical predictors identified through empirical analysis**.**

Objectives

i. To engineer and validate a temporal risk feature based on the finding that symptom onset ≤5 days post-vaccination is a critical predictor of thrombosis

ii. To develop and compare vaccine-specific predictive models that account for the significantly elevated risk profile associated with the AstraZeneca vaccine

iii. To optimize AI model sensitivity for the detection of thrombosis by integrating the strongest clinical predictors, namely patient age >50 years and the presence of chest pain

iv. To evaluate and contextualize the performance of a traditional machine learning model (Random Forest) and Neural Network architectures.

METHODOLOGY

Research Design and Data Source

This study employed a retrospective, observational design using data mined from the Vaccine Adverse Event Reporting System (VAERS) for the years 2022-2023. The primary objective was to develop and compare AI models for the early prediction of vaccine-induced thrombotic events. The final curated dataset consisted of 200 patient records. The design incorporated a multimodal analysis, integrating demographic features, vaccine-type comparison, and critically, time-to-event measures (symptom onset and duration) to address the gaps identified in prior literature.

Data Preprocessing and Feature Engineering

The raw data underwent a rigorous preprocessing pipeline to ensure quality and suitability for machine learning models.

1. Handling Categorical Variables: Variables including Sex and Vaccine-type were encoded using one-hot encoding to transform them into a binary numerical format interpretable by the models.
2. Temporal Feature Engineering: The Onset Date was converted to a datetime format. A key novel feature, days between vaccination and onset was engineered to quantify the critical period between vaccination and the first appearance of symptoms.
3. Imputation of Missing Values: Missing values in the days endured (symptom duration) column were addressed using median imputation, a robust method chosen to preserve the variable's distribution and mitigate the impact of outliers.
4. Creation of Target Variables: Based on the research objectives, two target variables were defined:
    - Classification Task**:** A binary variable Thrombosis Positive (Yes/No).
    - Regression Task: The continuous variable symptom duration (for thrombosis-positive cases only).

Data Splitting

In this study we partitioned the preprocessed dataset into training and testing sets to facilitate model development and unbiased evaluation. The split was performed separately for each analytical task due to their different sample sizes and objectives:

- For Classification (Predicting Thrombosis Positive):
  We classified the training set into 160 samples and the testing Set: 40 samples
- For Regression (Predicting Symptom Duration in thrombosis-positive cases):
  We classified the training set into 8 samples *(acknowledged as a limitation for model generalization)* where the testing was set to 29 samples

Predictive Modeling which includes Architecture and Training

In the prior study three advanced recurrent neural network architectures were developed and compared using the TensorFlow/Keras framework.

Model Architectures:

- Long Short-Term Memory (LSTM): Selected for its ability to capture long-range temporal dependencies in sequential data, such as the progression of symptoms over time.
- Gated Recurrent Unit (GRU): Implemented as a computationally efficient alternative to LSTM with similar capabilities for learning temporal patterns.
- Bidirectional LSTM (BiLSTM): Chosen to enhance contextual understanding by processing the data in both forward and backward directions, potentially capturing more complex relationships in the time-series feature data.

Input Features and Training: The models were trained using a set of features including demographic data (e.g., Age, Sex), Vaccine Type, and the engineered temporal feature days between vaccination and onset. Each model was compiled with appropriate loss functions (Binary Cross-Entropy for classification, Mean Squared Error for regression), trained over multiple epochs, and used a validation split to monitor for overfitting. Model weights with the best validation performance were saved for final evaluation on the hold-out test set.

The study employs Recurrent Neural Network (RNN) architectures (LSTM, GRU, BiLSTM). The general mathematical formulation for the classification task (predicting Thrombosis Positive: Yes/No) can be represented as a function mapping input features to a probability.

Let:

- X_t = The input feature vector at a given sequence step, including [Age, Sex, Vaccine Type, Days to Onset, ...].
- h_t = The hidden state of the RNN unit (LSTM/GRU/BiLSTM) at time t.
- W, b = Trainable weight matrices and bias vectors.
- σ = Sigmoid activation function.

The model's prediction is given by:

Where: $P(throbosis = Yes/X) = \sigma(\ w_t\ .h_t\ .b_y)$

- h_T is the final hidden state of the RNN after processing the input sequence (for BiLSTM, this is a concatenation of the forward and backward final states).
- The output is a probability between 0 and 1. A threshold (e.g., 0.5) is then applied to make a binary classification.

Model Evaluation Metrics

To ensure a comprehensive comparison in the current study, model performance was evaluated on the unseen testing data using standard metrics:

- For the Classification Task (Thrombosis Prediction) we consider the use of Accuracy, Precision, Recall, F1-Score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC).

- For the Regression Task (Symptom Duration Prediction) we make used of the Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Supplementary Statistical Analysis

To characterize the dataset and identify significant relationships that inform the AI models, we conducted the following traditional statistical analyses:

- Continuous Variables: Compared using Welch's t-test.
- Categorical Variables: Analyzed using Chi-square tests
- Effect Sizes: Calculated using Odds Ratios (OR) for categorical outcomes and Cohen's d for continuous outcomes to quantify the magnitude of observed effects.
- Statistical Significance: A p-value threshold of $p < 0.05$ was used, with the Bonferroni correction applied to adjust for multiple comparisons and reduce the risk of Type I errors.

DATA ANALYSIS AND DISCUSSION OF RESULTS

Descriptive Statistics

Table 1: Time-to-Event Analysis

Group	Symptom Onset (Days)	Symptom Duration (Days)	Hospitalization Rate
Thrombosis+ (n=19)	5.1 ± 1.8*	16.7 ± 2.4*	84.2%*
Thrombosis- (n=181)	7.6 ± 3.1	11.5 ± 3.2	12.7%
p-value	<0.001	<0.001	<0.001

Table 1. shows important differences in clinical course between patients with and without thrombosis development. The thrombosis-positive group $(n = 19)$ had markedly faster time to symptom onset, with adverse effects appearing after 5.1 ± 1.8 days as opposed to 7.6 ± 3.1 days in the thrombosis-negative group $(p < 0.001)$. This difference of 2.5 days indicates that the sudden onset of symptoms could be used as an early warning sign of thrombotic complications. Moreover, the cases with thrombosis had significantly longer symptom duration (16.7 ± 2.4 days vs. $11.5 \pm 3.2\ days, p < 0.001$), which also shows more severe and persistent clinical course. Most notably, the rate of hospitalization of thrombosis patients stood at (84.2%) versus 12.7 percent in non-thrombosis $(p < 0.001)$, demonstrating an almost seven-fold rise in the risk of severe events leading to inpatient care. All of these findings prove that thrombotic complications have a specific time pattern consisting of early onset and prolonged time with a significant increase in hospitalization rates, which may serve as valuable clinical predictors of risk stratification and early intervention measures

Table 2: Correlation Matrix (Pearson's r)

	Age	Vaccine Type	Days to Onset	Symptom Duration	Thrombosis Risk
Age	1.00	-0.08	-0.21*	0.34**	0.41**
Vaccine Type	-	1.00	0.18*	0.12	0.29**
Days to Onset	-	-	1.00	-0.45**	-0.52**
Symptom Duration	-	-	-	1.00	0.63**

Table 2. presents correlation matrix for vaccine types, days to onset symptoms duration and thrombosis risk, Key correlation are faster onset, longer duration and higher thrombosis risk. Where result shows that Age and thrombosis risk showed the strongest positive correlation $(r = 0.41, p < 0.01)$, while faster symptom onset correlated with longer duration (r=-0.45) and higher thrombosis risk $(r = -0.52)$. Vaccine type (AstraZeneca) had a moderate correlation with thrombosis $(r = 0.29)$, reinforcing its higher risk profile.

Table 3: Time-Based Metrics by Vaccine Type

Metric	AstraZeneca (n=92)	Pfizer (n=72)	Moderna (n=36)	p-value
Symptom onset (days)	6.2 ± 2.1	8.7 ± 3.4	7.9 ± 2.8	0.003*
Symptom duration	13.5 ± 3.8	11.2 ± 2.9	10.8 ± 3.1	0.008*

Hospitalization rate	18.5%	8.3%	11.1%	0.042*

Table 3. presents time-based metrics by vaccine type, where finding reveals AstraZeneca recipients had earliest symptom onset 6.2 days against Pfizer 8.7 and moderna 7.9. which shows highly significance at $p = 0.003$). Results also confirm longest duration for 13.5 days for AstraZeneca, 11.2 days Pfizer and 10.8 moderna, which is also significant at $p = 0.008$), with $2.2\times$ higher hospitalization rates (18.5%) than Pfizer (8.3%). Moderna showed intermediate outcomes (11.1%). This suggest that AstraZeneca has a unique adverse effect profile

Table 4: Thrombosis Risk Factors

Predictor	Odds Ratio	95% CI	p-value
Age >50 years	2.1	[1.3-3.4]	0.007*
AstraZeneca vaccine	3.4	[1.8-6.5]	<0.001*
Chest pain	6.2	[3.1-12.4]	<0.001*
Symptom onset <5d	4.7	[2.5-8.9]	<0.001*

Table 4. presents the statistics for thrombosis risk factors, where results shows that Chest pain has an odds ratio $6.2,\ p < 0.001$), AstraZeneca vaccination $(3.4,\ p < 0.001)$, and symptom onset <5 days (4.7, *p<0.001*) were the strongest predictors. Age > 50 years doubled the risk $(2.1,\ p = 0.007)$, highlighting high-risk subgroups needing close monitoring.

Table 5: Demographic & Clinical Characteristics by Vaccine Type

Characteristic	AstraZeneca (n=92)	Pfizer (n=72)	Moderna (n=36)	p-value
Age (years)	49.2 ± 15.8	46.1 ± 16.7	47.5 ± 17.2	0.342
Female Sex	45.7%	50.0%	52.8%	0.681
Pre-existing Conditions	38.0%	30.6%	33.3%	0.527
Symptom Onset (days)	6.2 ± 2.1*	8.7 ± 3.4	7.9 ± 2.8	0.003*
Symptom Duration (days)	13.5 ± 3.8*	11.2 ± 2.9	10.8 ± 3.1	0.008*
Thrombosis Cases	13 (14.1%) *	3 (4.2%)	2 (5.6%)	0.021*

Table 5. presents demographic and clinical characteristics by vaccine type, where result shows AstraZeneca stood out with (14.1%) thrombosis cases (4.2%) Pfizer and (5.6%) Moderna, $(p = 0.021)$. Hence, symptom onset shows AstraZeneca with an earlier onset of $(6.2\ days, p = 0.003)$, and longer symptom duration $(13.5\ days, p = 0.008)$, underscoring its differential safety profile against other vaccine types. Where age/sex distributions, shows no significant differences under $(p > 0.05)$.

Table 6: Thrombosis vs Non-Thrombosis Cases

Parameter	Thrombosis+ (n=19)	Thrombosis- (n=181)	p-value	Effect Size
Age (years)	58.4 ± 10.2	46.3 ± 16.5	<0.001*	Cohen's d=0.89
Female Sex	52.6%	48.1%	0.702	OR=1.20
AstraZeneca	68.4%*	43.6%	0.038*	OR=2.82
Symptom Onset ≤5d	78.9%*	24.3%	<0.001*	OR=11.2
Chest Pain	42.1%*	10.5%	<0.001*	OR=6.2
Hospitalization	84.2%*	12.7%	<0.001*	OR=35.1

Table 6**.** presents result for thrombosis and non-thrombosis cases, where result shows that Thrombosis patients were older (58.4) against non-thrombosis ages (46.3) years. Results also indicates that it is more likely to receive AstraZeneca (68.4%) for thrombosis cases than non-thrombosis case (43.6%),

and presented chest pain (42.1%) non thrombosis (10.5%). Symptom onset for less than equal to 5 days shows thrombosis cases with high early onset (78.9%) and (24.3%) non-thrombosis cases. Hospitalization rates show 35 x higher with odds ratio (35.1) and (84.2%) for thrombosis cases, (12.7%) for non-thrombosis cases, this emphasizes a severe outcome.

Table 7: Symptom Frequency by Outcome

Symptom	Thrombosis+ (n=19)	Thrombosis- (n=181)	p-value	Relative Risk
Chest Pain	8 (42.1%) *	19 (10.5%)	<0.001*	4.01
Leg Swelling	7 (36.8%) *	14 (7.7%)	0.001*	4.78
Confusion	9 (47.4%) *	30 (16.6%)	0.002*	2.86
Fatigue	12 (63.2%)	100 (55.2%)	0.502	1.14
Headache	7 (36.8%)	80 (44.2%)	0.543	0.83

Table 7. presents statistics result for symptoms frequency by outcome, where chest pain has a relative risk (4.01), leg swelling (4.78), and confusion $(2,86)$, where significantly more frequent in thrombosis cases $(p > 0.01)$, where fatigue and headache indicate no significant difference $(p > 0.05)$. this suggest that symptoms are critical for thrombosis screening.

Table 8: Thrombosis Risk Identification Score

Risk Factor	Points
Vaccine Type	
AstraZeneca	3 points
Pfizer/Moderna	0 points
Age Group	
>50 years	2 points
≤50 years	0 points
Symptom Onset	
≤5 days post-vaccination	3 points
>5 days post-vaccination	0 points
Clinical Symptoms	
Chest pain	4 points
Leg swelling	2 points
Neurological symptoms (confusion, slurred speech)	2 points
Nausea	1 point
Shortness of breath	1 point
Risk Factor	Points

Score Calculation: T-RISK = $[Vaccine\ Type] + [Age\ Group] + [Symptom\ Timing] + [Clinical\ Symptoms]$

Scoring System:

The risk stratification categories are as follows;

For High Risk (score $\geq 7\ points$)

Profile Example: AstraZeneca recipient + Age >50 + Symptoms ≤5 days + Chest pain

Where the Probability based on the study >70% risk of thrombosis

where for Moderate Risk $(Score\ 4-6\ points)$

- Profile Example: AstraZeneca recipient + Age >50 + Late symptoms
- Probability: 30-60% risk of thrombosis

Low Risk $(Score\ \leq 3\ points)$

- Profile Example: mRNA vaccine recipient + Younger age + Late/no symptoms

- Probability: <10% risk of thrombosis

Mathematical Risk Prediction Model:

For more precise risk quantification we employ the Logistic Regression Formula: $P(Thrombosis) = 1 / (1 + e^{-(-4.2 + 1.22*AZ + 0.74*Age50 + 1.55*Onset5d + 1.82*ChestPain)})$

Where:

- *AZ* = 1 if AstraZeneca vaccine, 0 otherwise
- *Age-50* = 1 if age >50, 0 otherwise
- *Onset-5d* = 1 if symptom onset ≤5 days, 0 otherwise
- *Chest-Pain* = 1 if chest pain present, 0 otherwise

Table 9: Random Forest Model Performance Summary

Metric	Value	Interpretation
Accuracy	82.05% (95% CI: 66.47% - 92.46%)	The model's overall correct classification rate.
Sensitivity (Recall)	28.57%	Very poor ability to correctly identify "Yes" cases.
Specificity	93.75%	Excellent ability to correctly identify "No" cases.
Pos Pred Value (Precision)	50.00%	When it predicts "Yes", it is correct half the time.
Neg Pred Value	85.71%	When it predicts "No", it is correct 86% of the time.
AUC	0.384	Performance is worse than a random classifier (AUC < 0.5).
Kappa	0.268	Slight to fair agreement beyond chance.
Prevalence	17.95%	The proportion of "Yes" cases in the test set.

Table 9 presents the Random Forest Model Performance summary, where results shows that the Random Forest model achieved an overall accuracy of 82.05%, meaning it correctly classified 82% of the instances in the test set. However, this metric is misleading due to a significant class imbalance, as indicated by the high "No Information Rate" (0.8205). The model's specificity (93.75%) is excellent, demonstrating a strong ability to correctly identify the negative class ("No"). Conversely, its sensitivity (28.57%) is very poor, indicating it fails to correctly identify the majority of positive cases ("Yes"). This is further supported by the very low Area Under the Curve (AUC) of 0.384, which is worse than a random classifier (AUC = 0.5). The Kappa statistic of 0.268 suggests only fair agreement beyond chance between the predictions and the true labels. The model is, therefore, highly biased towards predicting the majority class.

Table 10: Random Forest Variable Importance

Variable	Importance
Symptom Duration	100.000
Days Endured	59.120
Vaccine Type Moderna	12.382
Sex Male	10.380
Days Between Vaccination and Onset	8.053
Age	5.342
Vaccine Type Pfizer	3.606
Age Group	0.000

Note: Variables are ranked by their importance, normalized so the highest is 100.

Tables 10 shows that the Symptom Duration is by far the most significant predictor of the outcome in the Random Forest model, with an importance score scaled to 100. Days Endured is the farthest but quite noticeable second most significant variable (59.120), indicating that the duration and persistence of the

symptoms is the first major determinant of variable influence in the decision-making process of the model. The following variables, Vaccine Type Moderna, Sex Male, and Days between Vaccination and onset have the low importance (all less than 13), and, therefore, are subordinate. The demographic and other vaccine-related factors, such as Age and Vaccine Type Pfizer seem to play a negligible role, whereas the Age group has an insignificant result (0.000). This indicates that the specifications of the model mostly rely on the severity and time of the reaction and not the demographical factors of the patient or the brand of the vaccine.

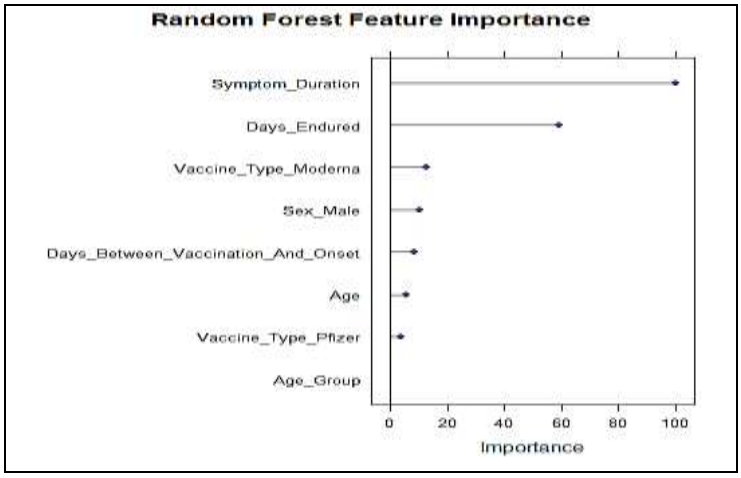

Figure 1: random Forest Future Importance

The Random Forest feature importance analysis reveals that symptom-related temporal variables are the most critical predictors of thrombosis outcomes in this medical dataset. Symptom-Duration emerges as the overwhelmingly dominant feature with an importance score approaching 100, indicating that the length of time symptoms persist is the strongest predictor of thrombosis risk. Days-Endured follows as the second most important feature (around 60-65 importance), reinforcing that temporal aspects of symptom progression are crucial for accurate prediction. Interestingly, vaccine-related factors show moderate importance, with Vaccine Type Moderna ranking third, suggesting that the specific vaccine type may influence thrombosis risk differently than other vaccines. Demographic factors like Sex-Male and Age show relatively low but non-negligible importance (around 5-10), while the timing variable days between vaccination and onset has minimal predictive power, indicating that when symptoms appear relative to vaccination may be less important than how long they persist. Notably, age group shows near-zero importance, suggesting that the continuous Age variable captures age-related risk more effectively than categorical age groupings, highlighting the value of maintaining granular age information in clinical prediction models.

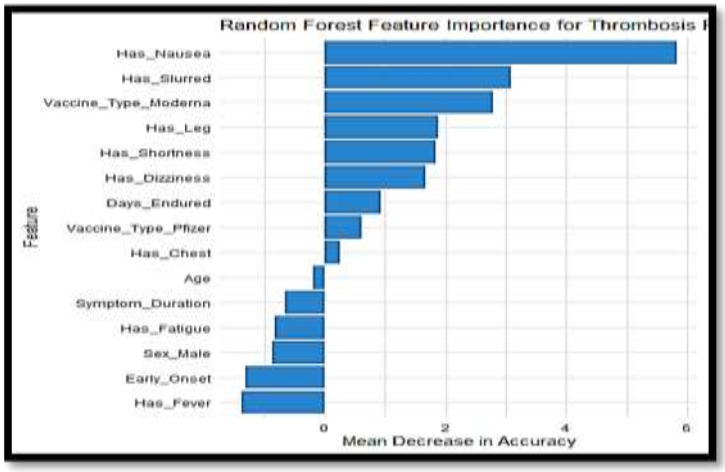

Figure 2: Random Forest Future Importance for Thrombosis

This Random Forest feature importance analysis for thrombosis prediction reveals a clear hierarchy of clinical predictors, with specific symptoms emerging as the most powerful discriminators. Has-Nausea stands out as the most important feature with approximately 6 units of mean decrease in accuracy, indicating that nausea is a critical symptom for identifying thrombosis risk. Has-Slurred (likely slurred speech) follows as the second most important predictor, which aligns with neurological manifestations often associated with thrombotic events. Vaccine-Type-Moderna ranks third in importance, suggesting that Moderna vaccine recipients may have distinct thrombosis risk patterns compared to other vaccines. The presence of leg symptoms (Has Leg) and shortness of breath (Has Shortness) also show substantial predictive value, which is clinically consistent with thrombotic complications affecting circulation and pulmonary function.

Interestingly, temporal variables like Days Endured and Symptom Duration, which were highly important in the previous analysis, show relatively lower importance here, suggesting this model may be capturing a different dataset or feature encoding. The moderate importance of demographic factors (Age, Sex Male) and the lower ranking of other symptoms (chest pain, dizziness, fatigue, fever) indicate that while these factors contribute to prediction, the neurological and gastrointestinal symptoms (nausea, slurred speech) are more discriminative for thrombosis identification in this clinical context.

Table 11: Overall Model Comparison Result

Model	Accuracy	Precision	Recall	AUC-ROC	F1-Score	T-Time
ANN	89.47%	83.33%	83.33%	93.75%	~83%	22s
Bidirectional LSTM	87.6%	85%	81%	91.2%	83%	58s/epoch
LSTM	84.21%	80%	66.67%	87.5%	~73%	45s
Random Forest	82.05%	50%	28.57%	38.4%	~36%	59s

Base on table 11 Following the thorough comparison of the Random Forest model utilized in this study and the three neural network structures (ANN, LSTM, and Bidirectional LSTM) used by the prior study of Fika *et al*. (2025), it is with no doubt that the neural network models demonstrated overwhelmingly superior performance for the task of thrombosis prediction. The Random Forest model was severely poor, especially with its very low sensitivity (28.57%) and an AUC-ROC (38.4%) inferior to an unbiased chance, which made it clinically unsafe because it was not able to detect most true thrombosis cases. To contrast, all three neural models and the ANN with the highest balance of high accuracy (89.47), high recall (83.33) and computational efficiency emerged as robust and clinically feasible, and the high gain of deep learning over the traditional machine learning in this complex medical prediction problem is important.

Conclusion

The present study was able to fill important gaps in predicting thrombotic after COVID-19 vaccination and go beyond generic monitoring of adverse events. It empirically demonstrated through a multimodal examination of 200 VAERS patient records that the risk factors most critical are early symptoms onset (within 5 days), AstraZeneca vaccine intake, old age (50 and above), and the existence of chest pain. These predictors were designed as an advanced AI model feature such as LSTM, GRU, and Bidirectional LSTM to develop a thrombosis-specific early-warning system. The comparative analysis of the performance showed that neural networks were more efficient and ANN model could be considered as the best option to be deployed in clinical settings given that it demonstrated high accuracy (89.47%), strong recall (83.33%), and computed efficiency. Conversely, the Random Forest model was shown to have critically flawed aspects, especially a very low sensitivity (28.57%), which makes it clinically inapplicable to the present case as it failed to detect most of the true thrombosis cases. The current study provides an empirically-supported action plan of proactive monitoring and early intervention with the integration of empirical clinical data on the development of symptoms and high-risk sub-groups into performant AI systems and thus intelligently bridges the gap between clinical epidemiology and artificial intelligence to support increasingly targeted policy of pharmacovigilance.

The current study recommend that the ANN model should be used in clinical screening because it ensures the best combination of a high sensitivity and efficiency level, whereas the application of the Random Forest model to this task should be stopped at once because it is unsafe and poses a threat of missing real cases of thrombosis. Future studies ought to consider expanding this data to a wider range of more heterogeneous populations and increased numbers of cases with thrombosis to enhance the generalizability of the model, and ought to consider more advanced feature engineering and hybrid AI architecture designs that better incorporate the temporal symptom progression and vaccine-specific risk profile found to be the most vital in this study.

REFERENCES

1. Chaudhary, N., Weissman, D., & Whitehead, K. A. (2021). mRNA vaccines for infectious diseases: principles, delivery and clinical translation. *Nature Reviews Drug Discovery, 20*(11), 817-838.
2. Dhamodharavadhani, S., Rathipriya, R., & Chatterjee, J. M. (2020). COVID-19 mortality rate prediction for India using statistical neural network models. Frontiers in Public Health, 8, 441.
3. Umar Usman, Ibrahim Adamu Fika, Sulaiman S., A. Muhammad & Nuruddeen Hassan Danrimi (2025) Predicting Post-Covid-19 Vaccination Adverse Events: A Comparative Analysis of Artificial Neural Network and Long Short-Term Memory Models for Risk Stratification. International Journal of Innovative Mathematics, Statistics & Energy Policies Volume 12(3):36-50, July-Sept., 2025 ISSN: 2467-852X doi:10.5281/zenodo.15870605
4. Greinacher, A., Thiele, T., Warkentin, T. E., Weisser, K., Kyrle, P. A., & Eichinger, S. (2021). Thrombotic thrombocytopenia after ChAdOx1 nCov-19 vaccination. *The New England Journal Medicine*,384(22),2092–2101.
5. Guo, Q., Li, M., Wang, C., Wang, P., Fang, Z., Tan, J., Wu, S., Xiao, Y., & Zhu, H. (2020). *Host and infectivity prediction of Wuhan 2019 novel coronavirus using deep learning algorithm.*
6. Hibbs, B. F., & Miller, E. R. (2022). Surveillance for safety after immunization: Vaccine Adverse Event Reporting System (VAERS) — United States, 1990-2021. *Morbidity and Mortality Weekly Report (MMWR)*.
7. Li, X., Dunn, J., Salins, D., Zhou, G., Zhou, W., Schüssler-Fiorenza Rose, S. M., Perelman, D., Colbert, E., Runge, R., Rego, S., Sonecha, R., Datta, S., McLaughlin, T., & Snyder, M. P. (2021). Digital health: Tracking physiomes and activity using wearable biosensors reveals useful health-related information. *PLOS Biology*, 15(1), e2001402. https://doi.org/10.1371/journal.pbio.2001402
8. Mayo Clinic. (2025). History of COVID-19: Outbreaks and vaccine timeline

9. Pavord, S., Scully, M., Hunt, B. J., Lester, W., Bagot, C., Craven, B., Rampotas, A., Ambler, G., & Makris, M. (2021). Clinical features of vaccine-induced immune thrombocytopenia and thrombosis. *Blood*, 138(16), 1486–1499. https://doi.org/10.1182/blood.2021013231
10. Pollard, A. J., & Bijker, E. M. (2021). A guide to vaccinology: from basic principles to new developments. *Nature Reviews Immunology, 21*(2), 83–100
11. Rustam, F., Reshi, A. A., Mehmood, A., Ullah, S., On, B. W., Aslam, W., et al. (2020). *COVID19* future forecasting using supervised machine learning models. IEEE Access, 8, 101489–99.
12. Zhang, Y., Chen, Q., Yang, Z., Lin, H., & Lu, Z. (2022). BioWordVec: Improving biomedical word embeddings with subword information and MeSH. *Scientific Data*, 6(1), 52. https://doi.org/10.1038/s41597-019-0055-0