International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.139, August 2026

Effects of COVID-19 Vaccines among University Students in Bangladesh: A Survey-based Machine Learning Analysis

Shahriar Arefin Zummon[1,2], Somapika Das[2], Aushtmi Deb[2], Marjana Akter Juti[2]

[1]Shahjalal University of Science and Technology, Sylhet, Bangladesh
[2]Leading University, Sylhet, Bangladesh

ABSTRACT

While young adults are generally considered resilient, the long-term side effects of COVID-19 vaccines in this demographic remain largely unexplored. This study investigates post-vaccination experiences among university students and faculty collected from three universities in Sylhet, Bangladesh through online and offline surveys between January and March 2025. After rigorous data cleaning, a finalized dataset of 591 respondents was obtained covering vaccine type, dosage, demographics, pre-existing health conditions, overseas travel history, and post-vaccination impacts (short-term, medium-term, and long-term). Descriptive, correlation, and predictive analyses were performed. Five supervised machine learning (ML) classifiers - Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), XGBoost, and K-Nearest Neighbors (KNN) - were evaluated alongside K-Means clustering ($k = 4$) with Principal Component Analysis (PCA) for patient risk profiling. Logistic Regression achieved the highest accuracy (70.79%). Random Forest feature importance identified side-effect severity (52.8%) and vaccine type (34.7%) as the dominant predictors. Respiratory conditions and allergies were the leading comorbidities associated with post-vaccination impacts. Findings provide actionable insights for policymakers and healthcare professionals refining vaccination protocols for young, active populations.

General Terms

Experimentation, Measurement, Human Factors

Keywords

COVID-19 vaccine, long-term side effects, vaccine safety, machine learning, university students, Bangladesh, pharmacovigilance

1. INTRODUCTION

The COVID-19 pandemic reshaped global healthcare, accelerating vaccine development at an unprecedented pace. Mass vaccination programs were launched worldwide, with millions receiving different vaccine types in record time [1, 2]. While these vaccines have been instrumental in reducing severe illness and death, questions remain about their long-term effects - particularly among young adults, a group that was not the primary focus during early clinical trials.

University students, typically aged 20–26, represent a unique demographic in this context. They are generally healthy, highly mobile, and actively engaged in social and academic environments that may increase their exposure to both the virus and vaccine-related factors [3]. Unlike older populations, they rarely suffer from chronic conditions, making them an ideal group to study potential vaccine-related side effects that cannot be attributed to pre-existing health conditions. While short-term effects such as fever, fatigue, and muscle pain are well-documented [4, 5], reports of lingering symptoms - ranging from neurological issues to cardiovascular concerns - have raised new questions regarding the prolonged impact of vaccination on this young, active population [6].

Bangladesh has implemented a robust vaccination program, administering vaccines such as Oxford-AstraZeneca (Covishield), Pfizer-BioNTech, Moderna, Sinopharm, and Sinovac [7]. While these vaccines have been largely successful in containing the virus, many individuals have reported extended post-vaccination symptoms [8]. The long-term effects of these vaccines, particularly in young adults, remain insufficiently studied [9], creating a need for more focused research.

Considering this gap, the present study explores post-vaccination experiences among university-affiliated individuals. By analyzing factors such as vaccine type, dosage, and demographic variables, we aim to correlate various factors with post-vaccination outcomes, thereby aiding policymakers and healthcare professionals in refining vaccination policies [10].

2. LITERATURE REVIEW

This section synthesizes existing research on COVID-19 vaccine side effects, with a focus on long-term outcomes, demographic variations, and the unique context of university students in Bangladesh. The review identifies key gaps that motivate the present study.

2.1 Vaccine Efficacy and Short-Term Side Effects

COVID-19 vaccines, including Pfizer-BioNTech, Moderna, AstraZeneca, and Sinopharm, have demonstrated high efficacy in preventing severe illness and hospitalization [1, 2]. Short-term side

effects such as fever, fatigue, and injection-site pain are commonly reported [4] and generally resolve within a few days. Their prevalence varies by vaccine type, age, and gender [11]: mRNA vaccines are associated with higher short-term reactogenicity compared to viral-vector vaccines like AstraZeneca [12].

2.2 Long-Term Side Effects and Demographic Factors

Research on long-term side effects - defined as symptoms persisting for weeks or months after vaccination - remains limited. Reported manifestations include fatigue, muscle pain, and neurological symptoms [5]. Menni et al. found that long-term effects are rare but more prevalent among younger individuals and women [6], suggesting that demographic factors shape the post-vaccination experience significantly. Younger individuals under 30 are more likely to experience side effects due to robust immune responses [11]. Gee et al. reported higher side-effect rates in women compared to men, possibly due to hormonal differences [5], while individuals with pre-existing conditions such as diabetes or hypertension may experience more severe effects [12].

2.3 University Students as a Distinct Population

University students represent a unique demographic due to high levels of social interaction, mobility, and stress, all of which may influence their vaccination experiences [13]. Jones et al. reported that lifestyle factors - such as irregular sleep patterns, high stress, and frequent travel - may exacerbate or mask vaccine-related side effects [3]. Despite these unique characteristics, targeted research on vaccine side effects in this population remains scarce, and existing studies have largely not examined how the interplay of lifestyle and immune factors shapes long-term outcomes among this group.

2.4 Vaccine Side Effects in Bangladesh

Bangladesh has participated extensively in the global vaccination effort [7], and vaccine acceptance has been shown to be influenced by education, income, and trust in the healthcare system [14]. Islam et al. found that short-term side effects such as fever and fatigue were common among Bangladeshi vaccine recipients, but long-term effects remain understudied [8]. Research in low- and middle-income countries (LMICs) has highlighted unique challenges such as limited healthcare infrastructure and vaccine hesitancy [15]. Alam et al. found that individuals with pre-existing conditions were more likely to report side effects [9], while Khan et al. found that university students were more likely to accept vaccines when provided with clear safety information [10].
The present study addresses key literature gaps by: (1) focusing on long-term rather than short-term outcomes; (2) examining demographic interactions specifically within a university student population; and (3) incorporating lifestyle factors such as travel history in shaping vaccine side-effect profiles.

3. DATASET AND COLLECTION PROCEDURE

This section describes the data source, collection procedure, and the key variables captured, as well as the quality-control steps taken to ensure the integrity of the final dataset used for analysis.
The dataset was collected from university-affiliated individuals (students and faculty) from three universities in Sylhet, Bangladesh through a combination of online (Google Forms distributed via Google Classroom and WhatsApp groups) and offline (paper-based survey) methods between January and March 2025, with trained enumerators assisting in offline data collection.

Table 1. Demographic Characteristics of Study Participants ($n = 591$)

Characteristic	Count	%
Gender – Male	297	50.25
Gender – Female	294	49.75
Multiple vaccine takers – Yes	85	14.38
Multiple vaccine takers – No	506	85.62
Overseas travel – Yes	77	13.03
Overseas travel – No	514	86.97
COVID-19 (confirmed by test)	50	8.46
COVID-19 (confirmed by symptoms)	184	31.13
After-vaccination COVID-19	8	1.35
Pre-vaccine health issues – Yes	51	8.63
Pre-vaccine health issues – No	540	91.37

The structured questionnaire, refined through three iterations based on pilot feedback, included questions on: participants' demographics; COVID-19 vaccination history (vaccine type, number of doses, combinations); pre-existing underlying health conditions; history of COVID-19 infection (before and after vaccination); overseas travel history; and post-vaccine impacts experienced (both short-term and long-term). Ethical compliance was ensured with explicit informed consent and secure data storage.
Data quality was maintained through pre-testing, cross-validation, and rigorous cleaning (exclusion of 23 contradictory or invalid responses), resulting in a finalized dataset of 591 respondents. The dataset comprehensively documented reactogenicity patterns, medium-term (1–3 months) symptoms, and long-term (>6 months) health outcomes, while prioritizing participant privacy and methodological robustness. Table 1 summarizes the key demographic and clinical characteristics.

4. METHODOLOGY

This section presents the complete analytical workflow employed in the study. The pipeline proceeds through eight sequential phases - from study design and tool selection to final cluster interpretation - ensuring reproducibility and methodological rigour at every stage. Figure 1 provides a visual overview of the full pipeline.

4.1 Study Design and Tools

The study employs a cross-sectional, survey-based design applied to secondary data collected from three universities in Sylhet, Bangladesh (January–March 2025, $n = 591$). This design is well-suited for characterising prevalence patterns and associations among post-vaccination variables within a defined population window, while being practically achievable without longitudinal follow-up infrastructure.
All analyses were implemented in Python 3.10 using the following libraries: pandas and NumPy for data manipulation; scikit-learn for preprocessing, classification, and clustering; XGBoost for gradient-boosted tree modelling; and matplotlib and seaborn for visualisation. Random seeds were fixed (`random_state = 42`) throughout to ensure reproducibility of all stochastic operations including train–test splitting, classifier training, and cluster initialisation.

4.2 Data Preparation

Before any analysis, the raw survey responses underwent a multi-step preprocessing pipeline to correct inconsistencies, standardise formats, and prepare the data for multi-label analysis.

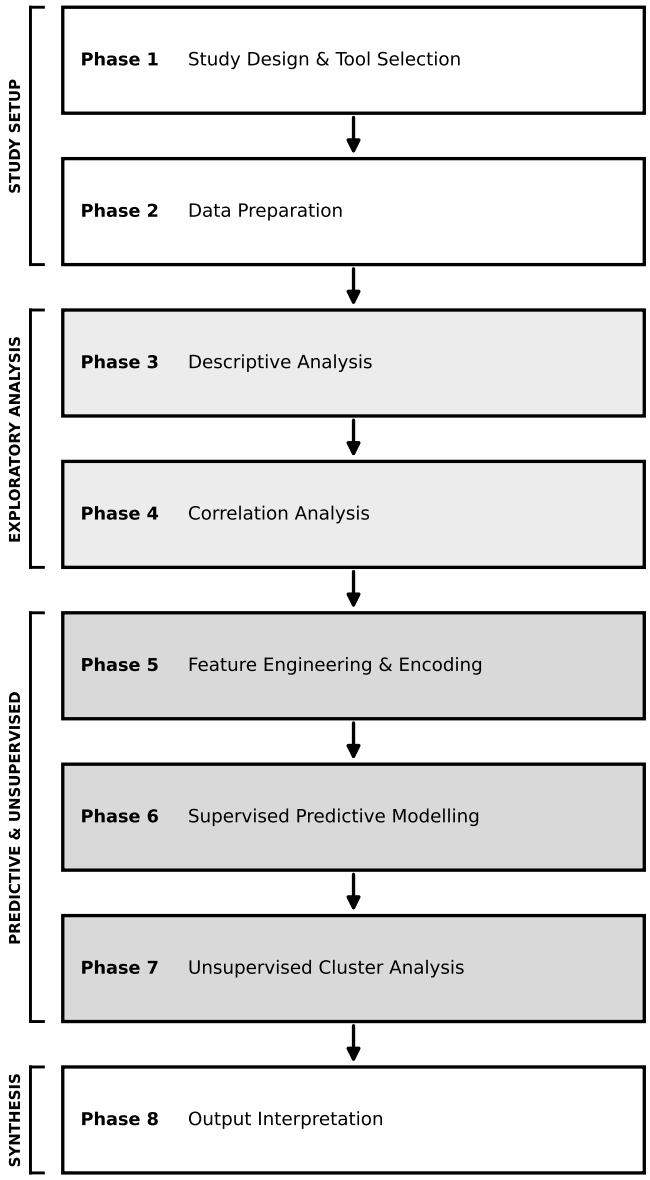

Fig. 1. Eight-phase analytical pipeline: study design, data preparation, descriptive analysis, correlation analysis, feature engineering, supervised predictive modelling, unsupervised cluster analysis, and output interpretation.

All 22 variables were systematically renamed via a predefined mapping dictionary applied with `df.rename()`, replacing verbose questionnaire headings with concise, machine-readable identifiers (e.g., Highest educational qualification → Education).
Null entries were filled with empty strings via `df.fillna('')` rather than dropped, preserving all 591 records for downstream analyses. Rows with contradictory or logically inconsistent responses were excluded during quality control, accounting for the 23 removals described in Section 3.

Six fields - vaccine names, side effects, underlying conditions, long-term effects, post-vaccination impacts, and visited countries - were stored as comma-delimited strings in the raw data. These were converted to Python list objects via `str.split(',')` with per-element `str.strip()`, enabling `df.explode()` operations for per-label frequency and cross-tabulation analyses.
All string-valued fields were lowercased to eliminate case-sensitivity mismatches (e.g., 'Pfizer' vs. 'pfizer') that would otherwise generate spurious distinct categories.

4.3 Descriptive Analysis

Descriptive analysis provided the population baseline against which subsequent correlational and predictive findings are contextualised. Univariate frequency distributions were computed for all demographic and clinical variables using `value_counts()` on exploded multi-label columns. Three outcome dimensions were characterised:

Vaccination profile - distribution of vaccine types, dose counts, and heterologous (mixed-brand) combinations, using `groupby` and `value_counts` on the exploded `Vaccine_Name` column.

Comorbidity burden - frequency of each underlying condition via `explode('Underlying_Conditions').value_counts()`.

Longitudinal outcomes - proportion of respondents reporting persistent effects beyond 3 months, and immediate post-vaccination impacts such as fever and skin reactions.

4.4 Correlation Analysis

Building on the descriptive baseline, correlation analysis examined six structured relationships between vaccination factors and health outcomes, progressing from broad population-level patterns to finer cross-variable interactions:

Dose-dependent side effects - side-effect frequency and the affection ratio (proportion of dose recipients experiencing any post-vaccine impact) were computed per dose count using `groupby('Dose_Received')` and plotted as grouped bar charts.

Vaccine-specific reactogenicity - severity was categorised numerically (1-2 = Mild; 3 = Moderate; 4-5 = Severe) and persistence was extracted from the `Persistent_Effects` field; both were cross-tabulated against vaccine type and visualised as stacked bar charts and a radar plot.

Temporal persistence - duration categories (no effect, $<$24 h, 1-2 days, 3-5 days, $>$1 week, $>$1 month) were compared across vaccine platforms using the radar plot.

Condition-impact associations - a double-explode on `Underlying_Conditions` and `Post_Vaccine_Impacts` followed by `groupby().size().unstack()` produced a full condition $\times$ impact contingency matrix; the top-5 $\times$ top-5 submatrix was visualised as a heatmap.

Vaccine liability - impact counts were cross-tabulated against vaccine type and displayed as a heatmap to identify which platforms were disproportionately associated with specific impact types.

Travel-impact associations - post-vaccine impact totals were aggregated by visited country using `groupby(['Visited_Country', 'Post_Vaccine_Impacts']).size()` to assess whether overseas travel correlated with impact diversity.

4.5 Feature Engineering and Encoding

Prior to predictive modelling, the target variable and input feature set were formally defined. The binary classification target was constructed as:

$$y_i = \begin{cases} 1 & \text{if } |\text{Post\_Vaccine\_Impacts}_i| > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

A preliminary Random Forest fit on a broader candidate feature pool identified the four most predictive variables through feature importance scores, yielding the final feature vector:

$$X = [x_{\text{Severity}}, x_{\text{Vaccine\_Name}}, x_{\text{Dose\_Received}}, x_{\text{Persistent\_Effects}}] \quad (2)$$

with importances of 52.8%, 34.7%, 12.5%, and 0.0% respectively. Categorical columns (`Vaccine_Name`, `Persistent_Effects`) were integer-encoded via `LabelEncoder`; all four features were subsequently scaled with `StandardScaler` (zero mean, unit variance) to ensure magnitude-insensitive classifiers (SVM, KNN) operated on a comparable scale.

4.6 Predictive Modelling

Five supervised classifiers were trained and evaluated on a stratified 85:15 train–test split (`random_state = 42`), ensuring class balance was preserved across both partitions. The classifiers and their key configurations were:

Logistic Regression (LR) - ℓ_2-regularised linear model; serves as an interpretable linear baseline.

Random Forest (RF) - ensemble of 100 decision trees (`n_estimators = 100`); also used for feature importance ranking.

Support Vector Machine (SVM) - linear kernel (`kernel = 'linear'`); maximises the margin between the two classes in the scaled feature space.

XGBoost - gradient-boosted trees with default regularisation; handles non-linear feature interactions efficiently.

K-Nearest Neighbors (KNN) - distance-based classifier using Euclidean distance on the standardised feature space.

All models were evaluated using accuracy, macro-averaged precision, recall, and F1-score. Confusion matrices were computed for each classifier to assess class-specific sensitivity and specificity, with particular attention to recall for the positive class (presence of post-vaccination impact).

4.7 Cluster Analysis

Unsupervised clustering was applied to reveal latent patient groupings without reference to the binary target label. The feature matrix for clustering was constructed through three steps: (1) dose count was mapped to an integer (1, 2, 3, 4+) via a custom `convert_dose()` function; (2) vaccine type and severity were category-encoded via `.cat.codes`; and (3) both `Underlying_Conditions` and `Post_Vaccine_Impacts` were one-hot expanded into binary indicator columns (`Condition_*`, `Impact_*`). The resulting high-dimensional matrix was standardised with `StandardScaler` before applying K-Means ($k = 4$, `random_state = 42`). The value $k = 4$ was selected via the elbow method on within-cluster inertia. Cluster profiles were interpreted by computing per-cluster feature means (`groupby('Cluster').mean()`), identifying the top conditions, top impacts, and average severity in each group. PCA (2 components) was applied for 2D scatter visualisation of cluster separation.

5. RESULTS

The results are organized in parallel with the methodology: descriptive findings are reported first to establish the sample profile, followed by correlation analysis, predictive modelling performance, and cluster characterization. All figures referenced here were generated from the study dataset.

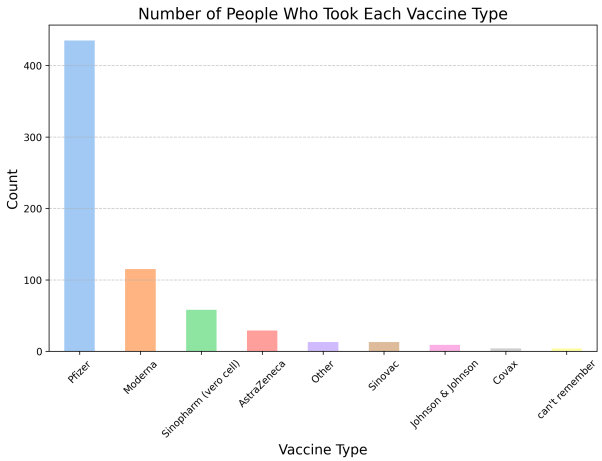

Fig. 2. Vaccine frequency distribution among study participants.

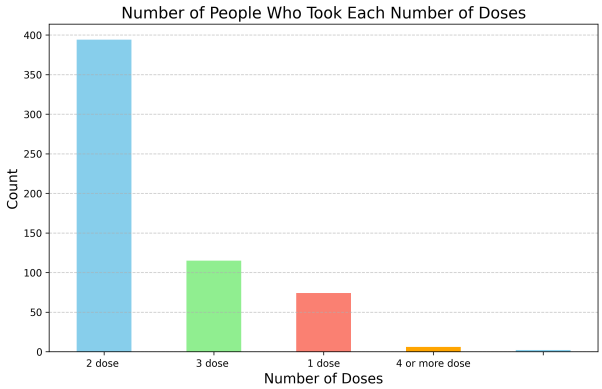

Fig. 3. Number of doses received.

Table 2. Multiple Vaccine Takers

Multiple Vaccines	Count	%
Yes	85	14.38
No	506	85.62

5.1 Descriptive Analysis

Gender distribution was balanced ($\chi^2 = 0.015$, $p = 0.902$), with 297 males (50.25%) and 294 females (49.75%), ensuring representative sampling for subsequent inferential analyses (Table 1).

Vaccination patterns. As shown in Fig. 2, Pfizer-BioNTech was the dominant vaccine ($n = 435$, 73.6% of recipients), followed by Moderna ($n = 115$) and Sinopharm ($n = 58$). Dose stratification (Fig. 3) revealed that 394 participants (66.7%) completed the standard 2-dose regimen. Heterologous vaccination was observed in 85 cases (14.4%; Table 2), with the Pfizer–Moderna combination being most prevalent ($n = 40$, 47.1% of mixed regimens), as illustrated in Fig. 4.

Comorbidities. Baseline health assessments (Fig. 5) identified respiratory conditions ($n = 54$, 9.1%) and allergies ($n = 33$, 5.6%) as the most common comorbidities among the study population.

Longitudinal outcomes. A total of 93 cases (15.7%) reported persistent long-term effects. Immediate post-vaccination impacts were

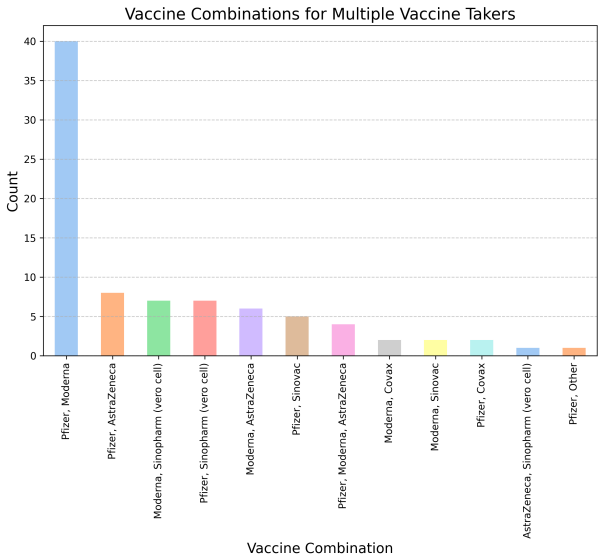

Fig. 4. Vaccine combinations among heterologous vaccination recipients.

Table 3. Prior COVID-19 and Travel Information

Condition / History	Count	%
Overseas Travel – Yes	77	13.03
Overseas Travel – No	514	86.97
COVID-19 – Confirmed by test	50	8.46
COVID-19 – Confirmed by symptoms	184	31.13
After-vaccination COVID-19	8	1.35
Pre-Vaccine Health Issues – Yes	51	8.63
Pre-Vaccine Health Issues – No	540	91.37

frequent: fever ($n = 159$, 26.9%) and skin reactions ($n = 111$, 18.8%). Table 3 presents the travel and prior COVID-19 history.

5.2 Correlation Analysis

The following subsections report the key relationships uncovered between vaccination factors, comorbidities, and post-vaccination outcomes.

5.2.1 Side Effects by Number of Doses. As shown in Fig. 6, individuals receiving their first dose reported side effects in 25 of 74 cases (33.8%), increasing to 191 of 394 (48.5%) for second doses, suggesting a more vigorous immune response upon subsequent exposure. Third doses yielded 59 side-effect reports of 114 (51.8%), while four or more doses showed the highest rate (4 of 6, 66.7%). When considering all post-vaccine impacts (Fig. 7), the affection ratio rose from 58.1% after the first dose to 68.8% after the second, 70.4% after the third, and 83.3% for four or more doses. This dose-dependent pattern provides valuable insights for healthcare providers counselling patients across their vaccination journey.

5.2.2 Severity, Persistence, and Impact Rate by Vaccine Type. Vaccine platforms differ meaningfully in their reactogenicity profiles, not only in whether side effects occur but in how severe and how long-lasting they are. As shown in Fig. 8, the likelihood of experiencing any side effects varies across vaccines: Pfizer presented a balanced profile (214 reports among 435 recipients), while Covax showed more pronounced reactivity (3 effects among 4 recipients, though the small subgroup size warrants caution). Examin-

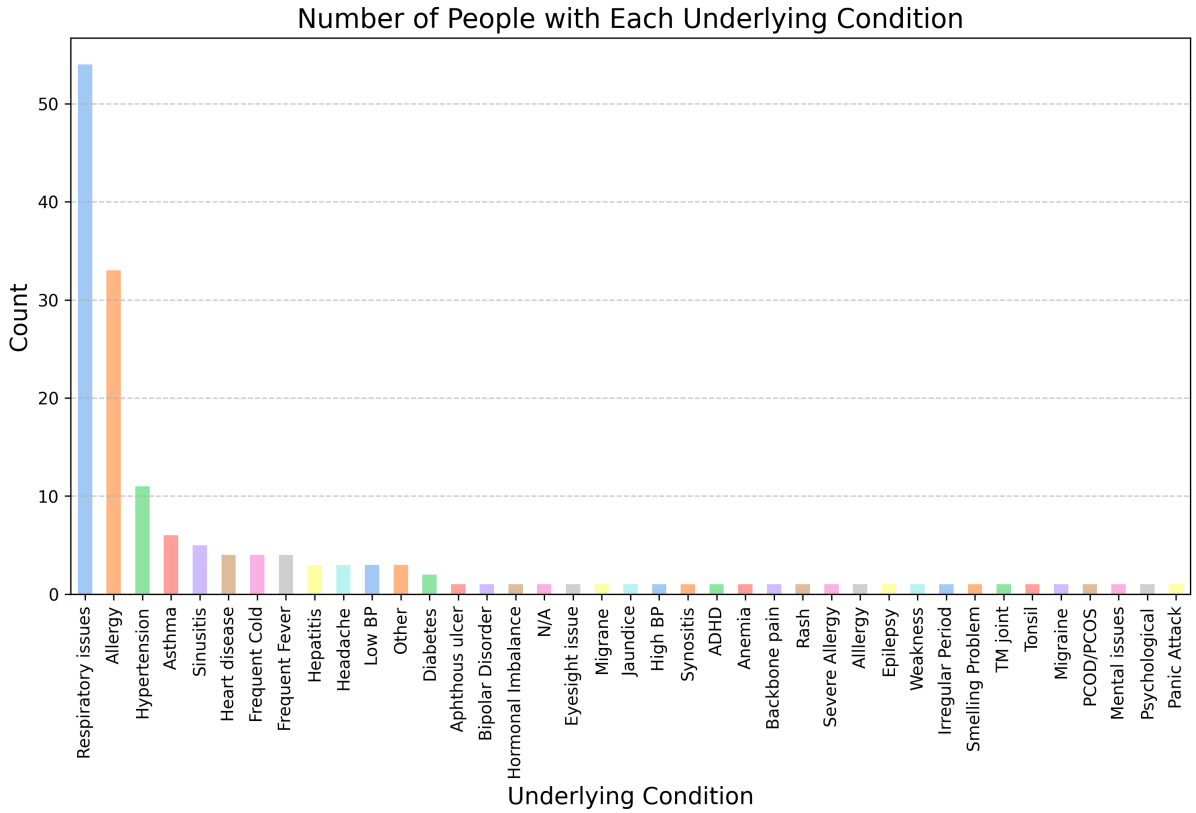

Fig. 5. Distribution of underlying health conditions among participants.

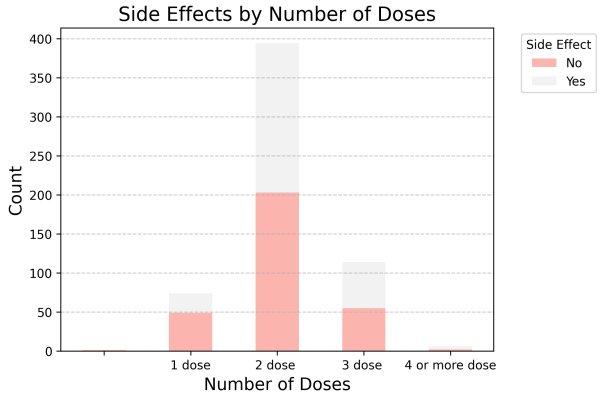

Fig. 6. Side effects by number of doses received.

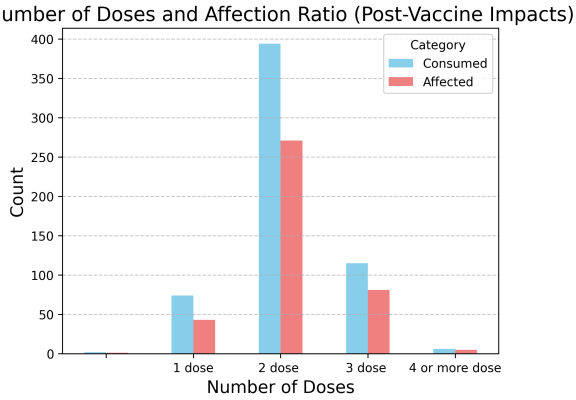

Fig. 7. Affection ratio (post-vaccine impacts) by number of doses.

ing severity (Fig. 9), most reactions were mild across all vaccines; Moderna showed a slightly higher proportion of severe cases (8 recipients), while Sinopharm and Sinovac presented predominantly mild profiles.

Temporal patterns (Fig. 10) showed that effects for most platforms resolved quickly, with the majority of Pfizer and Moderna reactions clearing within 1–3 days, and a small fraction across all vaccines experiencing effects lasting beyond one month. Table 4 reveals substantial variation in reported post-vaccine impact rates: Covax and the "can't remember" group showed the highest rates (25% each), followed by Johnson & Johnson (11.1%) and Sinovac/Other (7.7% each). In contrast, mRNA vaccines demonstrated notably lower impact percentages - Pfizer at 0.23% and Moderna at 0.87% - while viral-vector and inactivated vaccines occupied the middle range (AstraZeneca 3.45%, Sinopharm 1.72%).

5.2.3 Underlying Conditions vs. Post-Vaccine Impacts. Understanding how pre-existing health conditions modulate post-vaccination outcomes is important for identifying at-risk groups

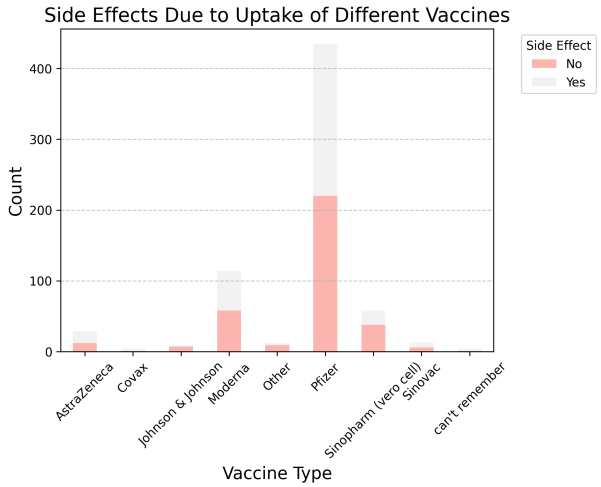

Fig. 8. Side effects due to uptake of different vaccine types.

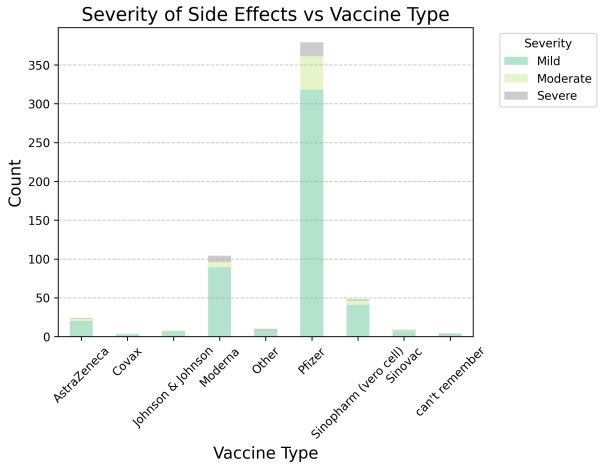

Fig. 9. Severity of side effects stratified by vaccine type.

Table 4. Percentage of Individuals Reporting Post-Vaccine Impacts by Vaccine Type

Vaccine Type	Impact (%)
AstraZeneca	3.45
Covax	25.00
Johnson & Johnson	11.11
Moderna	0.87
Other	7.69
Pfizer	0.23
Sinopharm (vero cell)	1.72
Sinovac	7.69
Can't remember	25.00

who may need closer monitoring or tailored counselling. The heatmap (Fig. 11) of top conditions versus top impacts reveals clinically meaningful patterns. Respiratory conditions showed the strongest associations, with 155 affected individuals reporting diverse reactions including fatigue (17 cases), fever (18 cases), and lethargy (14 cases). Allergies demonstrated a distinct profile, with

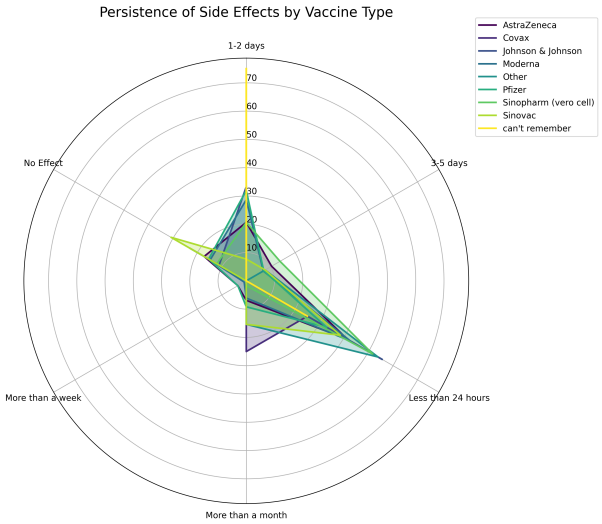

Fig. 10. Persistence of side effects by vaccine type (radar chart).

high rates of skin reactions (19 cases) and allergic responses (18 cases), likely reflecting immune system hypersensitivity. Hypertension patients reported moderate but balanced effects (4 cases each for fatigue, fever, and lethargy), while asthma sufferers showed comparatively milder reactions.

The bar chart (Fig. 12) underscores respiratory issues as the most significant risk factor (155 total affected cases), followed by allergies (88) and hypertension (38). When examining vaccine-specific liability (Fig. 13), Pfizer demonstrates the highest absolute numbers (e.g., 117 fever cases), reflecting its dominant usage; more informative are relative patterns showing Moderna's stronger association with fatigue (26 cases) versus AstraZeneca's balanced profile (10 fatigue, 9 fever). The consistent emergence of fatigue and fever across both condition-based and vaccine-based analyses points to these being universal post-vaccination markers.

5.3 Predictive Analysis

The results of the predictive modelling stage reveal which factors best predict post-vaccination impacts and how well different ML algorithms perform on this classification task. Random Forest feature importance analysis revealed that reaction severity accounted for 52.8% of predictive importance, vaccine type for 34.7%, and dose number for 12.5%, while effect persistence proved negligible (0.0%). This confirmed the feature set defined in Equation (2).

Table 5 summarizes the performance of all five classifiers on the held-out test set (15% split, $n = 89$). Logistic Regression achieved the highest overall accuracy (70.79%), maintaining a strong balance between precision and recall. Fig. 14 compares accuracy across models, demonstrating a narrow range (66–71%) indicative of stable underlying predictive patterns.

The confusion matrices (Fig. 15) demonstrate that all models were more effective at identifying true positive cases (individuals experiencing impacts) than true negatives. Logistic Regression correctly classified 51 of 53 positive cases but only 12 of 36 negative cases. Fig. 16 illustrates class-specific recall rates, highlighting the consistent challenge in predicting unaffected individuals across all classifiers. This has important clinical implications: false negatives (incorrectly predicting no impacts) may be more consequential than

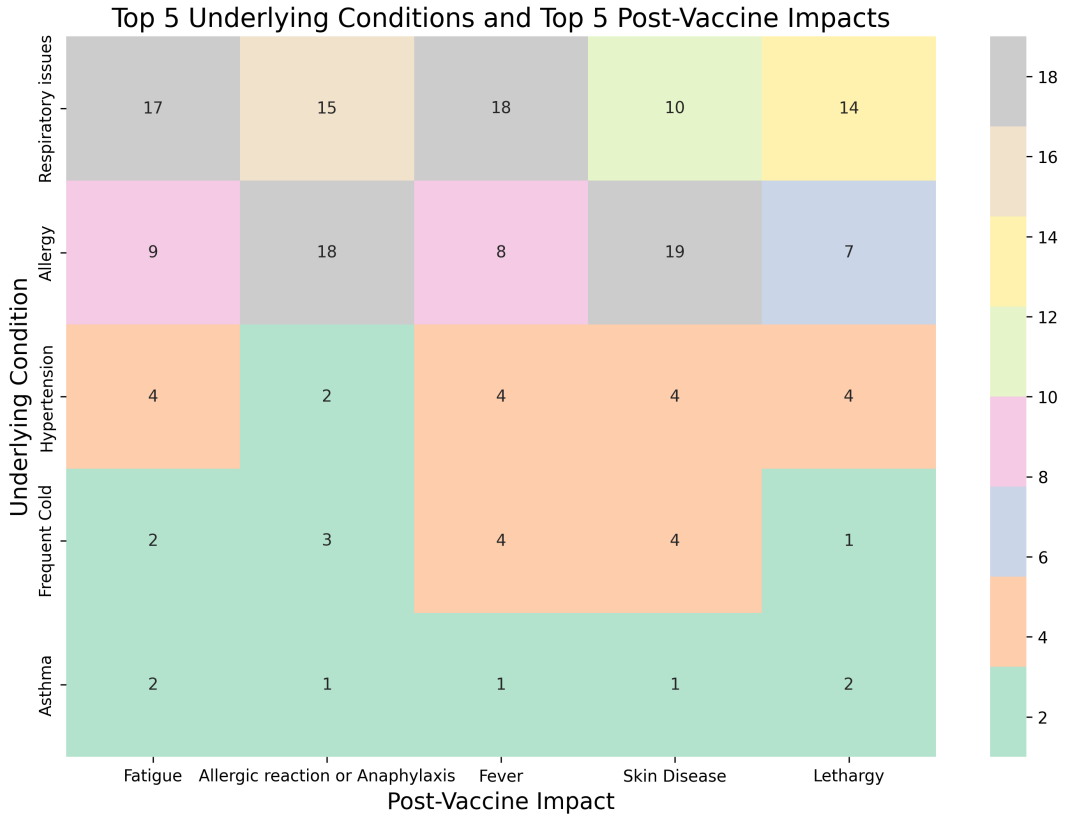

Fig. 11. Top 5 post-vaccine impacts due to top 5 underlying conditions.

false positives, reinforcing the value of sensitivity-focused models in vaccine safety surveillance.

5.4 Cluster Analysis

The unsupervised clustering step goes beyond binary prediction to characterize the heterogeneity of vaccine experiences across the study population. K-Means ($k = 4$) with PCA visualisation was used to identify candidate patient risk profiles, interpreted as follows:

Cluster 0 – Low-Risk Standard Recipients: Low severity scores, predominantly two-dose regimens, minimal comorbidity burden, and minimal post-vaccination impacts. Sinopharm and Pfizer most prevalent. Largest cluster.

Cluster 1 – Comorbidity-Dominant: Multiple underlying conditions (respiratory illness, allergies) with moderate to high post-vaccination impact burden across two to three dose regimens.

Cluster 2 – High Severity, Persistent Effects: High immediate side-effect severity, persistent early effects, and a broad long-term impact range. AstraZeneca and Covishield relatively more prevalent. Smallest cluster.

Cluster 3 – Multi-Dose, Travel-Associated: Higher dose counts, overseas travel history, and the highest diversity of post-vaccination impacts, potentially reflecting heterologous vaccine exposure or heightened immune priming from travel-related antigen exposure.

6. DISCUSSION

This study provides a data-driven characterization of COVID-19 vaccine-associated post-vaccination outcomes in a young, university-based population in Bangladesh. The findings across four analytical stages - descriptive, correlational, predictive, and clustering - converge to reveal a consistent and clinically meaningful picture of how vaccine platform, dose number, comorbidities, and immediate reactogenicity jointly shape long-term outcomes. The principal findings are discussed in turn below.

The progressive increase in the affection ratio across successive doses - rising from 58.1% after dose 1 to 68.8% after dose 2, 70.4% after dose 3, and 83.3% for four or more doses - aligns with broader immunological literature showing that repeated antigen exposure amplifies adaptive immune responses [6, 11]. This pattern underscores the importance of pre-vaccination counselling that sets realistic expectations for subsequent doses, particularly in university settings where students may be less familiar with cumulative immune priming effects.

The substantially lower impact rates for mRNA vaccines (Pfizer 0.23%; Moderna 0.87%) compared to Covax and Johnson & Johnson (25% and 11.1% respectively) may partly reflect reporting or attribution bias: Pfizer's dominant usage (>73% of recipients) potentially dilutes its perceived impact rate relative to less commonly used alternatives. Nonetheless, platform-level differences in reactogenicity and recovery timelines are consistent with prior literature on adenoviral-vector versus mRNA reactogenicity [12], and

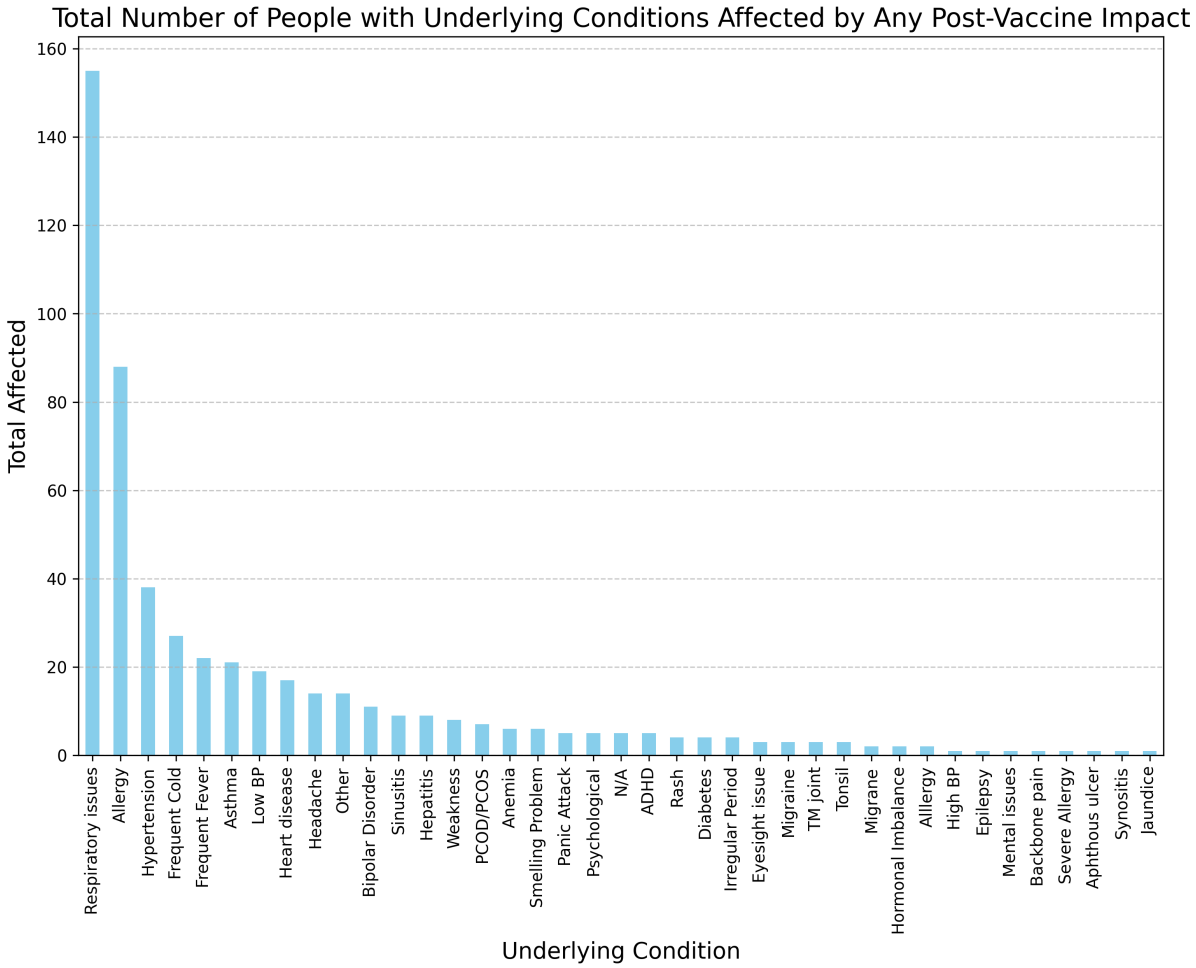

Fig. 12. Total affected individuals per underlying condition across all impacts.

Table 5. Model Performance Comparison (Macro-Averaged Precision, Recall, and F1)

Model	Acc.	Prec.	Rec.	F1
Logistic Reg.	**0.7079**	0.77	0.65	0.64
Random Forest	0.6966	0.71	0.65	0.64
SVM	0.6966	0.79	0.63	0.61
XGBoost	0.6854	0.68	0.65	0.65
KNN	0.6629	0.65	0.62	0.62

suggest that antigen delivery has downstream consequences for the post-vaccination experience.

Respiratory conditions emerged as the leading risk factor for post-vaccination impacts (155 total cases), followed by allergies (88) and hypertension (38). These findings are consistent with those reported by Alam et al. in the Bangladeshi population [9] and reinforce the need for personalized pre-vaccination counselling for at-risk individuals. The pattern of allergy-related skin reactions and respiratory patients' lethargy suggests that condition-specific monitoring protocols could meaningfully benefit these subgroups.

The consistent performance across all five classifiers (66–71% accuracy) underscores the robustness of the underlying predictive relationships in the data. Logistic Regression outperforming ensemble methods is notable, and likely reflects the relatively small effective feature set (four predictors) and the binary classification target, where linear decision boundaries are adequate. The dominance of side-effect severity (52.8% feature importance) validates the clinical intuition that the immediate magnitude of vaccine response is the strongest indicator of subsequent post-vaccination burden. All classifiers demonstrated stronger recall for the positive class (impact present), which is favourable for safety-screening applications where missing a true case is more costly than a false alarm.

The K-Means clusters provide an empirical basis for risk-stratified post-vaccination monitoring. The High Severity, Persistent profile represents a priority group for proactive clinical follow-up, while the Multi-Dose, Travel-Associated profile is of particular relevance

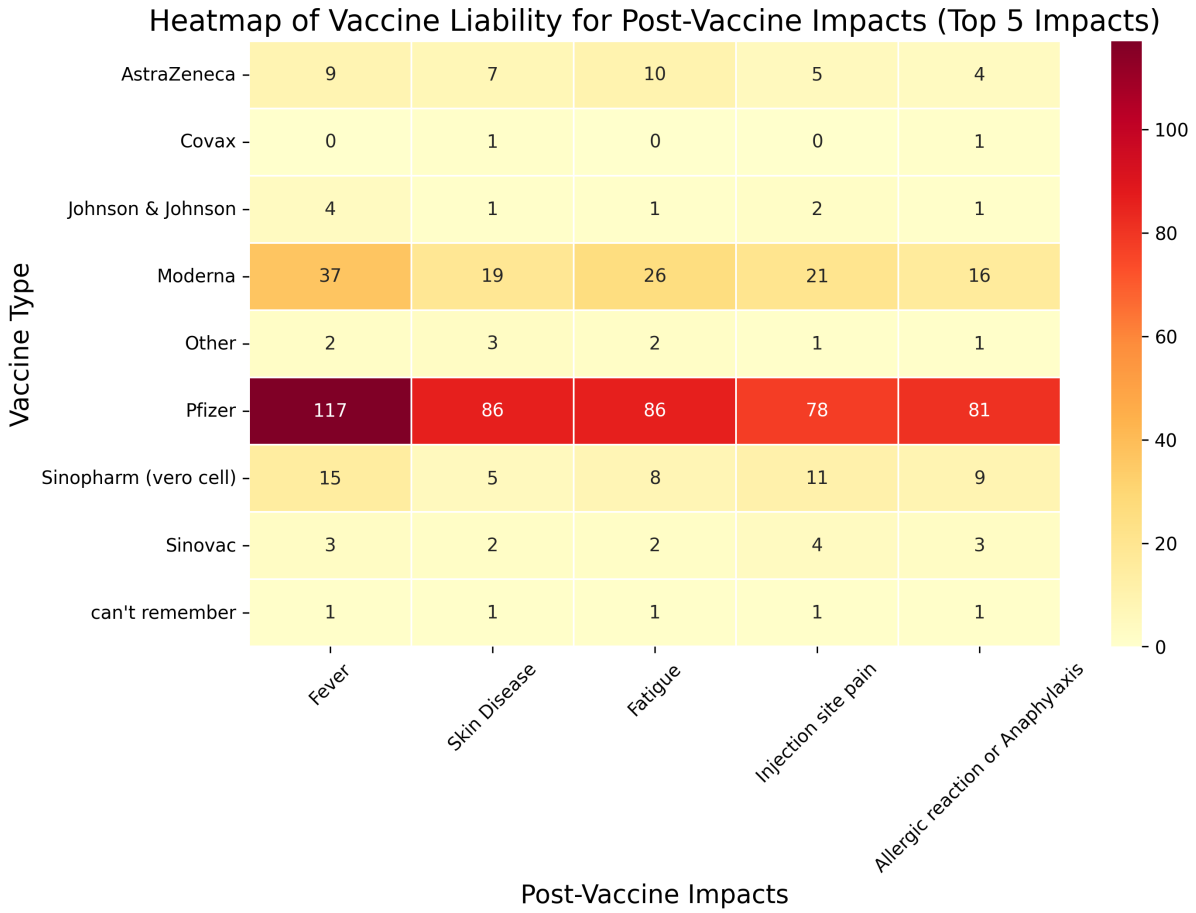

Fig. 13. Heatmap of vaccine liability for the top 5 post-vaccine impacts.

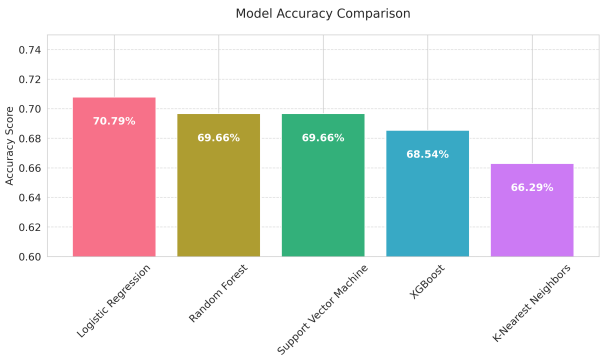

Fig. 14. Comparative accuracy of predictive models.

for Bangladesh, where overseas travel and heterologous dosing from different national supply chains are common and may amplify immune variability.

The study is limited by its cross-sectional design, which precludes causal inference. Self-reported outcomes are subject to recall and attribution biases. The sample, while gender-balanced and ethically collected, is geographically restricted to Sylhet, Bangladesh, limiting generalisability. The absence of an unvaccinated comparator group prevents estimation of baseline symptom prevalence. Finally, concurrent SARS-CoV-2 infection cannot be fully excluded as a confound for some reported symptoms.

7. CONCLUSION

This study investigated post-vaccination experiences among 591 university-affiliated individuals in Sylhet, Bangladesh, using a four-stage analytical framework spanning descriptive, correlational, predictive, and clustering analyses. The findings collectively demonstrate that long-term vaccine-associated impacts are common, heterogeneous, and predictable from a small number of clinical and demographic features. The key contributions are summarized as follows:

(1) Dose-dependent reactogenicity is clearly evident, with affection ratios rising from 58.1% after dose 1 to 83.3% for four or more doses.

(2) mRNA vaccines (Pfizer, Moderna) demonstrated notably lower post-vaccine impact rates than viral-vector and less common platforms.

(3) Respiratory illness and allergies are the dominant comorbidities associated with higher post-vaccination impact burden,

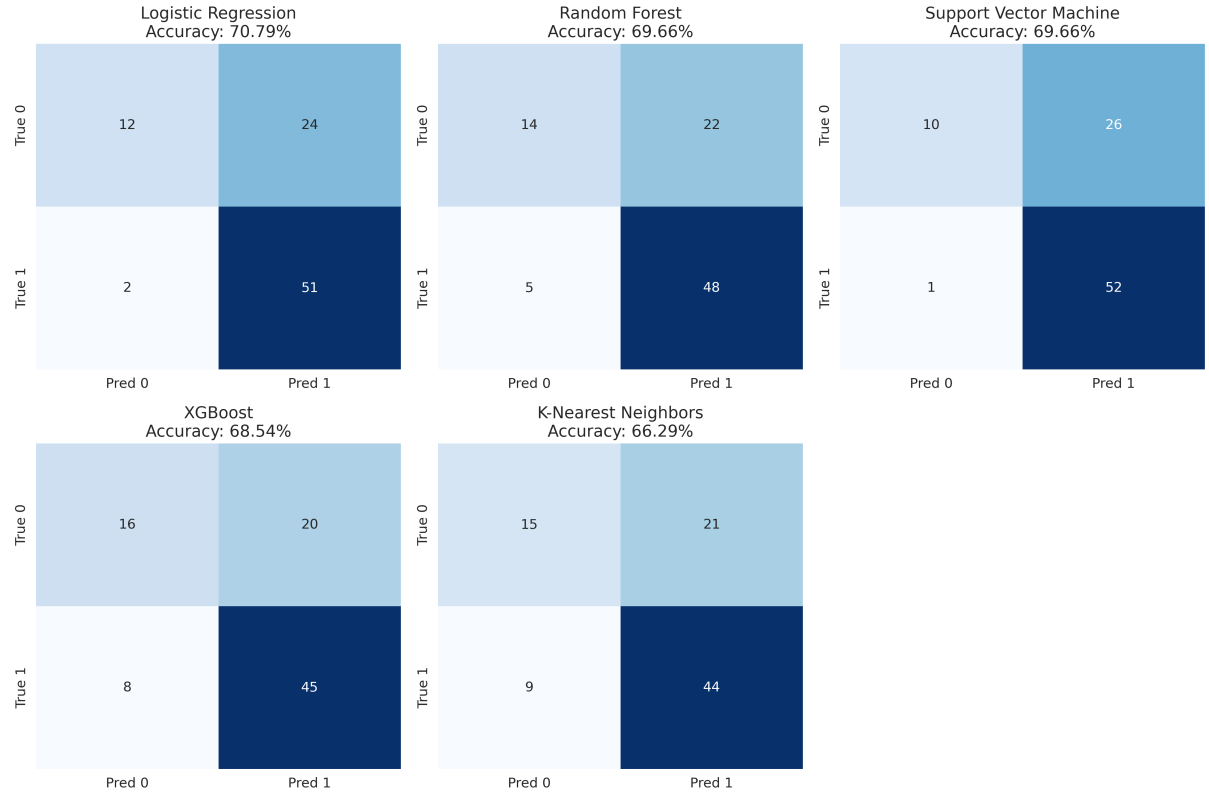

Fig. 15. Confusion matrices for all five classifiers.

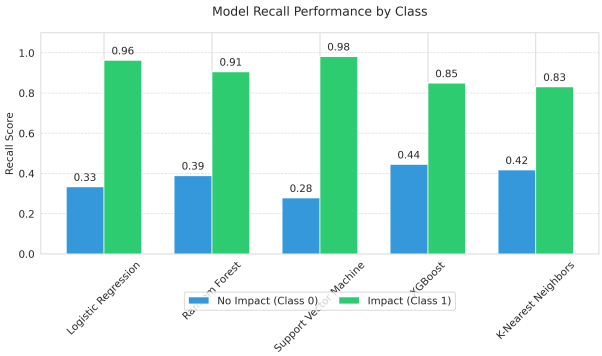

Fig. 16. Class-specific recall rates across models.

with respiratory conditions affecting 155 individuals across multiple impact categories.

(4) Logistic Regression achieved the best classification accuracy (70.79%), with side-effect severity identified as the most important predictor (52.8% feature importance) by Random Forest analysis.

(5) K-Means clustering supported the construction of clinically interpretable patient risk profiles for risk-stratified post-vaccination surveillance.

These findings provide actionable insights for policymakers and healthcare professionals refining vaccination protocols for young, active populations. Future work should employ prospective longitudinal designs with unvaccinated comparators, expand geographic scope across Bangladesh, and incorporate biological biomarkers to elucidate immunological mechanisms underlying persistent post-vaccination symptoms.

8. ACKNOWLEDGMENTS

The authors thank the students and faculty of the three participating universities in Sylhet, Bangladesh for their voluntary participation, and the trained enumerators for their diligence in offline data collection. No conflicts of interest are declared. This research received no external funding.

9. REFERENCES

[1] F. P. Polack, S. J. Thomas, N. Kitchin, J. Absalon, A. Gurtman, S. Lockhart, J. L. Perez, et al., "Safety and efficacy of the BNT162b2 mRNA Covid-19 vaccine," *New England Journal of Medicine*, vol. 383, no. 27, pp. 2603–2615, 2020.

[2] L. R. Baden, H. M. El Sahly, B. Essink, K. Kotloff, S. Frey, R. Novak, D. Diemert, et al., "Efficacy and safety of the mRNA-1273 SARS-CoV-2 vaccine," *New England Journal of Medicine*, vol. 384, no. 5, pp. 403–416, 2021.

[3] A. Jones, J. Smith, and E. Brown, "Lifestyle factors and COVID-19 vaccine side effects among university students," *Journal of Adolescent Health*, vol. 70, no. 3, pp. 456–462, 2022.

[4] J. Chapin-Bardales, J. Gee, and T. Myers, "Reactogenicity following receipt of mRNA-based COVID-19 vaccines," *JAMA*, vol. 325, no. 21, pp. 2201–2202, 2021.

[5] J. Gee, P. Marquez, J. Su, G. M. Calvert, R. Liu, T. Myers, N. Nair, S. Martin, T. Clark, L. Markowitz, et al., "First month of COVID-19 vaccine safety monitoring - United States, December 14, 2020–January 13, 2021," *Morbidity and Mortality Weekly Report (MMWR)*, vol. 70, no. 8, pp. 283–288, 2021.

[6] C. Menni, K. Klaser, A. May, L. Polidori, J. Capdevila, P. Louca, C. H. Sudre, L. H. Nguyen, D. A. Drew, J. Merino, et al., "Vaccine side-effects and SARS-CoV-2 infection after vaccination in users of the COVID symptom study app in the UK: a prospective observational study," *The Lancet Infectious Diseases*, vol. 21, no. 7, pp. 939–949, 2021.

[7] M. S. Rahman, M. T. Islam, and M. Z. Hossain, "COVID-19 vaccination in Bangladesh: challenges and opportunities," *Journal of Global Health*, vol. 11, pp. 1–8, 2021.

[8] M. T. Islam, M. S. Rahman, and M. Z. Hossain, "Short-term side effects of COVID-19 vaccines in Bangladesh: a cross-sectional study," *Bangladesh Medical Research Council Bulletin*, vol. 48, no. 1, pp. 45–52, 2022.

[9] M. J. Alam, M. A. Khan, and M. S. Rahman, "Pre-existing conditions and COVID-19 vaccine side effects in Bangladesh," *Journal of Health, Population and Nutrition*, vol. 41, no. 1, pp. 1–10, 2022.

[10] M. A. Khan, M. J. Alam, and M. S. Rahman, "Vaccine hesitancy among university students in Bangladesh: a cross-sectional study," *BMC Public Health*, vol. 22, no. 1, pp. 1–9, 2022.

[11] H. G. Rosenblum, S. C. Hadler, D. Moulia, T. T. Shimabukuro, J. R. Su, N. K. Tepper, et al., "Safety of mRNA vaccines administered during the initial 6 months of the US COVID-19 vaccination programme: an observational study," *The Lancet*, vol. 399, no. 10325, pp. 924–944, 2022.

[12] D. K. Shay, J. Gee, J. R. Su, T. R. Myers, P. Marquez, R. Liu, B. Zhang, C. Licata, T. A. Clark, and T. T. Shimabukuro, "Safety monitoring of the Janssen (Johnson & Johnson) COVID-19 vaccine - United States, March–April 2021," *Morbidity and Mortality Weekly Report (MMWR)*, vol. 70, no. 18, pp. 680–684, 2021.

[13] J. Smith, E. Brown, and A. Jones, "Vaccine hesitancy among university students: a global perspective," *Vaccine*, vol. 39, no. 45, pp. 6543–6549, 2021.

[14] M. Z. Hossain, M. S. Rahman, and M. T. Islam, "Factors influencing COVID-19 vaccine acceptance in Bangladesh: a cross-sectional study," *Vaccine*, vol. 39, no. 45, pp. 6543–6549, 2021.

[15] S. S. A. Karim and T. de Oliveira, "Challenges and opportunities for COVID-19 vaccination in low- and middle-income countries," *The Lancet Global Health*, vol. 9, no. 6, pp. e789–e790, 2021.