

Toward Secure Web-Based Telemedicine: Detection of Relevant Cyberattacks

Andi Cahyono^{1*}, Ervira Dwiapriani As Syifa², Koko Harianto³, Ana Yulianti⁴, Liant Pratama Ranu Kumbolo⁵

^{1,2,5} Informatika Medis, Universitas Sains dan Teknologi Indonesia, Pekanbaru

³ Electrical Engineering and Computer Science IGP, National Yang Ming Chiao Tung University, Hsinchu

⁴ Fakulti Informatika dan Komputeran, Universiti Sultan Zainal Abidin, Terengganu

andicahyono@usti.ac.id¹, erviradwiapriani@usti.ac.id², koko.cs09@nycu.edu.tw³, SI5227@putra.unisza.edu.my⁴,
2417052801005@usti.ac.id⁵

Article Info

Article history:

Received 2026-08-30

Revised 2026-09-28

Accepted 2026-10-03

Keyword:

*Cyberattack detection,
Healthcare, Logistic
Regression, Telemedicine,
Web based healthcare.*

ABSTRACT

Web-based telemedicine systems are exposed to cyber threats that may affect the confidentiality, integrity, and availability of healthcare services. This study develops a domain oriented cybersecurity baseline by reorganizing selected CIC-IDS2017 traffic into five classes relevant to web-based telemedicine: Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS. After cleaning, 1,698,296 observations were retained. A leakage-aware pipeline was applied using an 80:20 stratified split, training-derived preprocessing, 63 locked predictors, standardization fitted only on the training partition, and Moderate Random Undersampling applied exclusively to training data. Logistic Regression served as the primary interpretable baseline, while Decision Tree, Random Forest, and Linear SVM were included only for comparative validation. Evaluation on an untouched test set of 339,660 observations showed that Logistic Regression achieved 94.62% accuracy, 58.70% Macro-F1, 99.40% macro ROC-AUC, and 61.73% macro PR-AUC. Strong performance was observed for Normal, Port Scan, and DoS/DDoS, whereas Web Application Attack remained difficult to distinguish and was frequently misclassified as Brute Force. These findings indicate that network-flow features provide a useful benchmark-level baseline, while application-oriented attacks require additional authentication, application-layer, and SIEM-correlated evidence.



This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.

I. INTRODUCTION

Telemedicine has become an increasingly important component of digital healthcare by enabling remote clinical consultation, exchange of medical information, and access to healthcare services through web-based platforms [1], [2]. The growing dependence on Internet-connected healthcare applications, however, also expands the attack surface of healthcare information systems [3]. Unlike conventional web services, telemedicine platforms may process sensitive clinical information, user credentials, consultation records, and other health-related data, making cybersecurity essential for maintaining the confidentiality, integrity, and availability of digital healthcare services [4]–[6]. A successful cyberattack against such systems may not only compromise information assets but also disrupt the continuity and

reliability of healthcare services delivered to patients [7]–[9]. Consequently, reliable detection of malicious activities is particularly important in web-based telemedicine, where cybersecurity incidents may affect both sensitive health information and the continuity of digital healthcare services.

Web-based telemedicine platforms are exposed to several categories of cyber threats commonly associated with Internet facing applications and servers. Port scanning may be used during reconnaissance to identify accessible services and potential entry points, while brute force attacks can target authentication mechanisms used by patients, healthcare professionals, and system administrators [10]–[12]. At the application layer, web attacks such as Cross-Site Scripting (XSS) and SQL injection may exploit weaknesses in user-facing interfaces or database driven services [13]. Denial of Service (DoS) and Distributed Denial of Service (DDoS)

attacks may also degrade or completely disrupt service availability by exhausting computational or network resources [14]. Collectively, these threats represent reconnaissance, credential-oriented attacks, web application exploitation, and availability disruption that are relevant to the attack surface of web-based telemedicine services [15]–[17]. However, these threat categories may not be equally distinguishable from network flow characteristics alone, particularly when authentication and application oriented attacks depend on contextual information that is not directly represented in network traffic.

Machine learning has increasingly been investigated as an approach for distinguishing legitimate network activity from malicious behavior using patterns extracted from network traffic [18]–[20]. Public cybersecurity datasets, including CIC-IDS2017, have been widely used to develop and evaluate detection models because they provide labeled benign and malicious traffic representing multiple attack scenarios [21]. However, many benchmark based studies formulate the problem primarily from a general network intrusion detection perspective, where broad attack collections and predictive performance are the main focus [22]. Such evaluations provide limited evidence regarding how well selected attack categories correspond to the security priorities of a specific application domain. This concern is particularly relevant to healthcare oriented environments, where the representativeness of a dataset and its correspondence with the target domain influence the interpretation of intrusion detection performance [23]. For web-based telemedicine, a focused evaluation of reconnaissance, authentication related attacks, web application exploitation, and availability disruption can therefore provide a more application oriented understanding of which threats are distinguishable using network flow characteristics and which may require additional security context [24], [25].

This study therefore adopts CIC-IDS2017 as a general cybersecurity benchmark rather than as a telemedicine-specific dataset. A domain oriented subset was constructed by selecting and reorganizing attack scenarios that are functionally relevant to web-based telemedicine systems. Five final classes were considered: Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS. The Web Application Attack category combines XSS and SQL Injection traffic to represent attacks directed toward the web application layer. Logistic Regression was employed as the primary lightweight and interpretable baseline for examining the extent to which these threat categories can be distinguished using network flow characteristics. To support a reproducible and leakage aware evaluation, the experimental design separates the test partition before training dependent preprocessing, derives feature cleaning and standardization parameters exclusively from the training partition, applies moderate random undersampling only to the training data, and preserves the naturally imbalanced test partition for final evaluation. Decision Tree, Random Forest,

and Linear Support Vector Machine were additionally included as representative conventional classifiers for comparative validation rather than for exhaustive model selection or optimization.

The contribution of this study is not the development of a new classification algorithm, but the establishment of a domain-oriented cybersecurity baseline that links benchmark based network flow classification with security concerns relevant to web-based telemedicine. The study contributes a reproducible evaluation framework that combines domain oriented attack mapping, leakage aware preprocessing, training only imbalance handling, and evaluation on an untouched test distribution. Logistic Regression remains the primary interpretable baseline, while representative conventional classifiers are used to contextualize whether observed class separability and error patterns are associated primarily with a linear decision boundary or persist across alternative classifier structures. Beyond aggregate performance, the analysis emphasizes class specific precision, recall, F1-score, ROC-AUC, PR-AUC, and misclassification behavior, with particular attention to the distinction between credential oriented and web application attacks. These findings provide an empirical basis for identifying the capabilities and limitations of network flow based detection and motivate subsequent validation using application, authentication, and security event data from a web-based telemedicine environment.

II. METHOD

This section describes the methodological framework used to evaluate cyberattack categories that are functionally relevant to web-based telemedicine systems. The study was designed as a reproducible and leakage aware experimental pipeline comprising domain oriented attack class mapping, data cleaning, fixed train test partitioning, training derived feature preprocessing, training only class rebalancing, model development, and evaluation on an untouched test partition. CIC-IDS2017 was used as a general cybersecurity benchmark, while Logistic Regression (LR) served as the primary lightweight and interpretable classifier. Decision Tree (DT), Random Forest (RF), and Linear Support Vector Machine (Linear SVM) were included as representative conventional classifiers to provide comparative validation of the class separability and error patterns observed with LR rather than to perform exhaustive model selection.

A. Research Design

This study employed a quantitative experimental design to examine the extent to which network flow characteristics can distinguish five cyberattack classes selected according to their functional relevance to web-based telemedicine: Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS. The overall research workflow is illustrated in Figure 1.

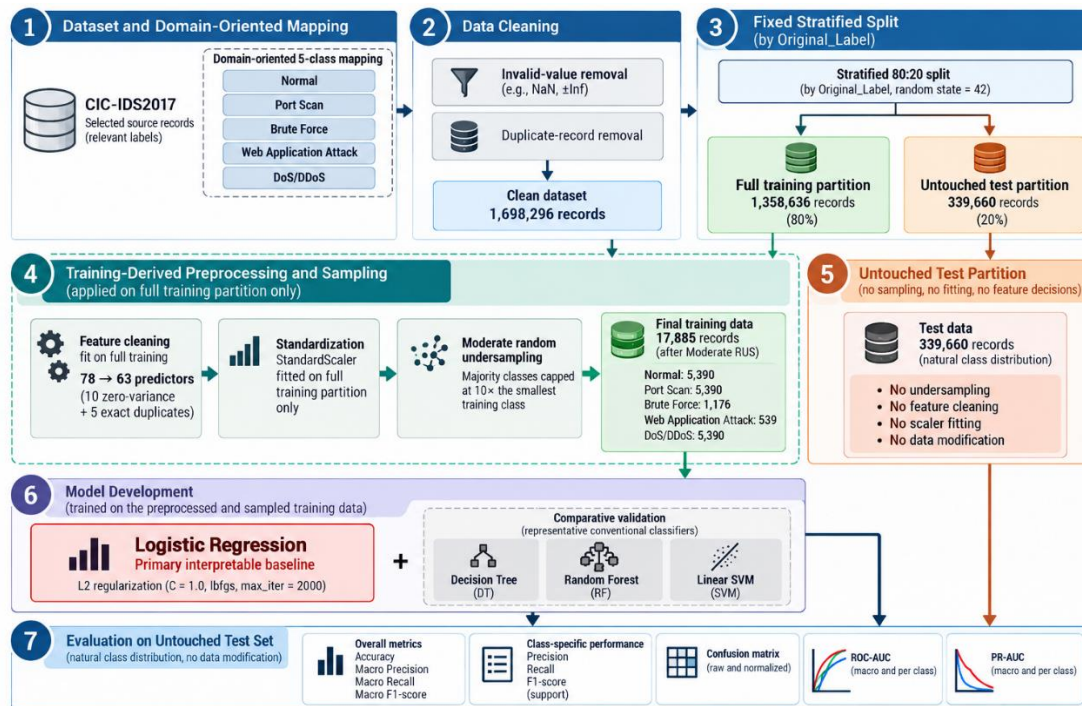


Figure 1 Research Workflow for Cyberattack Detection

The experimental procedure began with the selection and domain-oriented mapping of relevant CIC-IDS2017 traffic records into the five final classes. The mapped records were then cleaned by removing observations containing invalid numerical values and duplicate records. After cleaning, the complete dataset was divided into fixed training and testing partitions using an 80:20 stratified split with a random state of 42. Stratification was performed using the original CIC-IDS2017 attack labels to preserve the representation of the underlying attack subtypes across both partitions.

All training dependent preprocessing decisions were subsequently derived exclusively from the full training partition. Feature cleaning was first performed to remove zero-variance and exact duplicate predictors, after which the retained numerical predictors were standardized using parameters estimated from the full training data. These feature-selection and scaling parameters were then applied unchanged to the test partition, ensuring that no test information influenced preprocessing decisions.

Because the training data remained highly imbalanced after partitioning, moderate random undersampling was applied only to the training partition. Majority classes were capped relative to the smallest training class, while the available Brute Force and Web Application Attack observations were retained. The resulting training subset was used to develop the Logistic Regression baseline and the representative DT, RF, and Linear SVM comparative classifiers. In contrast, the test partition was preserved in its naturally imbalanced form and was not subjected to resampling or model-fitting operations.

The fitted models were finally evaluated on the same untouched test partition. Performance assessment included

overall accuracy, Macro Precision, Macro Recall, Macro-F1, class-specific precision, recall, and F1-score, multiclass confusion matrices, and one vs rest ROC-AUC and PR-AUC measures. Macro-F1 was treated as the primary summary metric because it assigns equal importance to each class and is therefore more informative than overall accuracy for evaluating performance under substantial class imbalance. The comparative classifiers were used to contextualize the behavior of the LR baseline, particularly the separability of credential oriented and web application attacks, rather than to identify an optimal classifier for telemedicine deployment.

B. Dataset and Attack Class Selection

The CIC-IDS2017 dataset was used as the benchmark data source because it provides labeled network flow observations covering benign activity and multiple categories of malicious traffic generated in a controlled experimental environment. The benchmark includes traffic associated with brute-force attacks, denial-of-service attacks, web attacks, infiltration, botnet activity, port scanning, and other intrusion scenarios. However, CIC-IDS2017 was not generated from a telemedicine infrastructure and was therefore not treated in this study as a healthcare-specific traffic dataset. Instead, it was used as a general cybersecurity benchmark from which a domain-oriented subset was constructed to examine threat categories that are functionally relevant to Internet-facing, web-based telemedicine services.

The attack-class selection was guided by the security functions represented in the attack surface of a web-based telemedicine system rather than by an attempt to reproduce all attack categories available in CIC-IDS2017. Four malicious

threat functions were considered: reconnaissance of exposed services, credential-oriented intrusion against authentication mechanisms, exploitation of the web-application layer, and disruption of service availability. Legitimate traffic was retained as a reference class. Based on this scope, the selected CIC-IDS2017 records were reorganized into five final classes: Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS.

BENIGN traffic was mapped to the Normal class to represent legitimate network and service activity. PortScan records were retained as Port Scan to represent reconnaissance activities directed at discovering accessible ports and services. Web Attack–Brute Force records were mapped to Brute Force because this scenario more closely represents repeated authentication attempts against a user-facing web service than service-oriented brute-force scenarios

such as FTP or SSH attacks. Web Attack XSS and Web Attack SQL Injection were combined into the Web Application Attack class because both represent exploitation attempts directed at the web-application layer. The selected records contained 652 XSS and 21 SQL Injection observations. Treating SQL Injection as a separate class would therefore have produced an extremely small class for independent class-level evaluation. Finally, DDoS, DoS Hulk, DoS GoldenEye, DoS slowloris, and DoS Slowhttptest were consolidated into the DoS/DDoS class because they share the functional objective of degrading or disrupting service availability.

The composition of the selected CIC-IDS2017 subset and its relationship to web based telemedicine security are summarized in Table 1.

TABLE 1
MAPPING OF SELECTED CIC-IDS2017 LABELS TO FINAL CYBERATTACK CLASSES

Original CIC-IDS2017 Label	Final Class	Threat Category	Web-Based Telemedicine Relevance
BENIGN	Normal	Legitimate activity	Represents normal network and service traffic
PortScan	Port Scan	Reconnaissance	Represents discovery of exposed ports and Internet-facing services
Web Attack – Brute Force	Brute Force	Credential oriented attack	Represents repeated attempts against user-facing authentication mechanisms
Web Attack – Sql Injection	Web application attack	Web application attack	Represents exploitation attempts against user-facing and database-driven application components
DoS Hulk, DoS GoldenEye, DoS slowloris, DoS Slowhttptest, DDoS	DoS/DDoS	Availability attack	Represents degradation or disruption of access to web-based telemedicine services.

The domain-oriented subset was intentionally narrower than the complete CIC-IDS2017 attack taxonomy. FTP-Patator and SSH-Patator, which represent brute-force attacks against FTP and SSH services, were not included because the present study focuses on credential-oriented behavior associated with user-facing web authentication; Web Attack Brute Force was therefore used as the representative class for this threat function. Bot and Infiltration scenarios were also excluded because they primarily represent post compromise, host level, or multi-stage intrusion behavior outside the four functional threat categories examined in this study. Heartbleed was not mapped to a final class because it represents a specific TLS vulnerability rather than one of the predefined functional categories and was represented by only 11 observations in the selected source data. These exclusions define the analytical scope of the study and should not be interpreted as implying that the excluded attacks are irrelevant to telemedicine security.

The resulting five-class structure therefore represents a domain-oriented abstraction of selected CIC-IDS2017 attack scenarios rather than a complete model of telemedicine cybersecurity threats. Its purpose is to establish a controlled benchmark problem for examining the extent to which network flow characteristics can separate threat categories

associated with different security functions of web-based telemedicine. This distinction is important because the subsequent analysis focuses not only on classification performance but also on where flow-level representations fail to distinguish functionally different threats. The selected records were subsequently cleaned and partitioned before any training-dependent feature preprocessing or class rebalancing was performed, as described in the following subsection.

C. Data Preprocessing and Dataset Construction

Following the domain oriented class mapping, the selected CIC-IDS2017 records contained 1,905,188 observations across the five final classes. The selected data comprised 1,363,390 Normal records, 158,930 Port Scan records, 1,507 Brute Force records, 673 Web Application Attack records, and 380,688 DoS/DDoS records. Because these observations were collected from multiple CIC-IDS2017 traffic scenarios and exhibited substantial class imbalance, data cleaning was performed before partitioning and model dependent preprocessing.

Records containing invalid numerical values, including missing values and positive or negative infinity, were removed. Duplicate observations were subsequently excluded to reduce repeated network-flow patterns that could otherwise

inflate the effective sample size and bias model evaluation. In total, 2,274 observations containing invalid values and 204,618 duplicate observations were removed. The resulting clean dataset contained 1,698,296 records distributed across the five final classes.

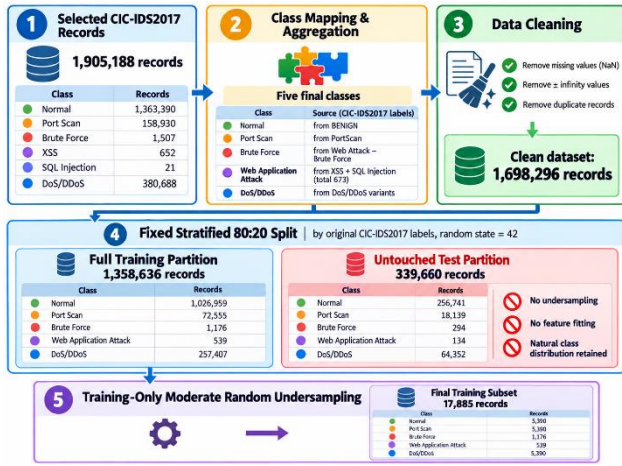


Figure 2 Dataset Construction and Balancing Process

After cleaning, the dataset was divided into fixed training and testing partitions using stratified random sampling with an 80:20 ratio and a random state of 42. Stratification was performed using the original CIC-IDS2017 labels rather than only the five aggregated final classes. This procedure was used to preserve the representation of the underlying attack subtypes, including the constituent web-attack and DoS/DDoS scenarios, across both partitions. The resulting full training partition contained 1,358,636 observations, while the test partition contained 339,660 observations.

The test partition was isolated at this stage and remained untouched throughout all subsequent training-dependent operations. No undersampling, feature-selection fitting, standardization fitting, or model training was performed using the test data. Consequently, the class distribution of the test partition retained the naturally imbalanced distribution inherited from the cleaned benchmark data. This design was adopted to provide a more realistic assessment of class-specific performance and false-positive behavior than an artificially balanced test set.

Table 2 summarizes the class distribution of the cleaned dataset, the full training partition, the final training subset used after training-only moderate random undersampling, and the untouched test partition. The undersampling procedure itself is described separately in Section II.E.

The partitioning strategy therefore separated model development from final evaluation before any training-dependent feature preprocessing or class rebalancing was performed. The full training partition was subsequently used to determine the retained feature space and standardization parameters, while the untouched test partition was reserved exclusively for final performance assessment.

TABLE 2
DATASET DISTRIBUTION BEFORE ACROSS EXPERIMENTAL PARTITIONS

Final Class	Selected Records	Full Training	Training After Moderate RUS	Untouched Test
Normal	1283700	1026959	5390	256741
Port Scan	90694	2555	5390	18139
Brute Force	1470	1176	1176	294
Web Application Attack	673	539	539	134
Dos/DDoS	321759	257407	5390	64352
Total	1698296	1358636	17885	339660

D. Training-Derived Feature Preprocessing

Feature preprocessing was performed exclusively using the full training partition obtained after the fixed train-test split. The original CIC-IDS2017 feature set contained 78 numerical predictors. To avoid introducing information from the test partition into the feature-preparation stage, all feature-removal decisions were determined using only the 1,358,636 training observations and were subsequently applied unchanged to the test data.

The first preprocessing step removed predictors that provided no discriminative information within the training partition. Ten zero-variance features were identified and excluded because their values remained constant across all training observations. The remaining predictors were then examined for exact duplication, and five additional features were removed because their values were identical to those of other retained predictors throughout the training partition. Consequently, the feature space was reduced from 78 to 63 numerical predictor.

Pairwise correlation among the retained predictors was additionally examined as a diagnostic assessment of feature redundancy. Highly correlated feature pairs were documented but were not automatically removed solely on the basis of correlation. This decision was intended to avoid additional feature-selection procedures that could unnecessarily alter the representation of the original network-flow characteristics. The same locked set of 63 predictors was therefore used consistently throughout the subsequent training and evaluation stages.

Because Logistic Regression and other linear classifiers are sensitive to differences in feature magnitude, the retained numerical predictors were standardized using StandardScaler. For each predictor, the standardized value was calculated as

$$z = \frac{x - \mu}{\sigma}$$

where z denotes the standardized feature value, x represents the original feature value, μ is the mean of the corresponding predictor estimated from the full training partition, and σ is its standard deviation.

Both the feature-space definition and the scaling parameters were derived before any random undersampling was performed. The fitted scaling parameters were then applied unchanged to the training subset used for model development and to the untouched test partition. Thus, neither the test observations nor the class-rebalancing procedure influenced the feature-removal criteria or the estimation of scaling parameters. This training-derived preprocessing strategy was adopted to maintain a consistent feature representation across all classifiers while minimizing the risk of information leakage during final performance evaluation.

E. Training-Only Random Undersampling

The full training partition remained highly imbalanced after the fixed train-test split, with substantially larger numbers of Normal, Port Scan, and DoS/DDoS observations than Brute Force and Web Application Attack records. To reduce the dominance of the majority classes during model development while avoiding the substantial information loss associated with fully equalizing all classes, a moderate random undersampling strategy was applied exclusively to the training data.

The smallest training class was Web Application Attack, containing 539 observations. Majority classes were therefore capped at ten times the size of this class, corresponding to a maximum of 5,390 observations per majority class. Random undersampling reduced the Normal, Port Scan, and DoS/DDoS classes to 5,390 observations each, while all available Brute Force and Web Application Attack records were retained. Consequently, the final training subset contained 5,390 Normal, 5,390 Port Scan, 1,176 Brute Force, 539 Web Application Attack, and 5,390 DoS/DDoS observations, resulting in a total of 17,885 training records.

No synthetic observations were generated during this procedure. Random undersampling was performed only after the feature space and scaling parameters had been determined from the complete training partition. The resulting training subset therefore used the same locked 63-feature representation and scaling parameters described in Section II.D. A fixed random state of 42 was used to support reproducibility.

Importantly, no resampling was applied to the test partition. The 339,660 test observations retained their original class distribution and were reserved exclusively for final evaluation. This separation ensured that class rebalancing influenced model development only, while reported performance reflected predictions on the naturally imbalanced test distribution.

F. Logistic Regression and Comparative Validation

Logistic Regression (LR) was employed as the primary classifier in this study because it provides a computationally efficient and interpretable linear baseline for examining the separability of the selected cyberattack categories using network-flow characteristics. Rather than seeking an optimized predictive model, the LR baseline was used to

determine the extent to which the five domain-oriented classes could be distinguished using a relatively simple decision boundary and to identify classes for which substantial overlap remained.

The LR model was trained using the moderate-RUS training subset containing 17,885 observations and the locked set of 63 standardized predictors. L2 regularization was applied with a regularization parameter of $C = 1.0$, the lbfgs solver was used for optimization, and the maximum number of iterations was set to 2,000. These parameters were fixed throughout the experiment, and no hyperparameter search or model-selection cross-validation was performed. This configuration was intentionally maintained as a conventional baseline rather than an extensively optimized classifier.

To determine whether the class-separability and error patterns observed with LR were specific to a linear decision boundary, three representative conventional classifiers were additionally evaluated: Decision Tree (DT), Random Forest (RF), and Linear Support Vector Machine (Linear SVM). DT used the Gini criterion with unrestricted maximum depth, a minimum of two samples for node splitting, and one sample per leaf. RF consisted of 200 trees using the Gini criterion, square-root feature selection, bootstrap sampling, and unrestricted maximum depth. Linear SVM was configured with $C = 1.0$, squared-hinge loss, and a maximum of 5,000 iterations. A random state of 42 was used where applicable.

All classifiers were trained using the same moderate-RUS training subset and the same locked 63-feature representation and were evaluated on the identical untouched test partition. No hyperparameter optimization was conducted for the comparative classifiers, and the comparison was not intended to identify an optimal classifier for telemedicine deployment. Instead, the additional models were used to contextualize the LR baseline by examining whether important class-specific behavior, particularly confusion between credential-oriented Brute Force and Web Application Attack traffic, persisted when alternative linear and nonlinear decision structures were applied.

Accordingly, LR remained the principal model for the subsequent analysis, while DT, RF, and Linear SVM were treated as supporting comparative evidence. The evaluation therefore emphasized not only aggregate predictive performance but also whether similar class-specific limitations emerged across classifier structures, providing additional insight into the capabilities and limitations of network flow features for distinguishing security threats relevant to web-based telemedicine.

G. Logistic Regression and Comparative Validation

Model performance was evaluated exclusively on the untouched test partition containing 339,660 observations. No test observations were used during feature cleaning, scaling-parameter estimation, random undersampling, or model training. The same test partition was used for Logistic Regression and all comparative classifiers. Performance was assessed using accuracy, class-specific precision, recall, and

F1-score, together with Macro Precision, Macro Recall, and Macro-F1. Because the test partition retained substantial class imbalance, Macro-F1 was used as the primary summary metric rather than accuracy alone.

Multiclass confusion matrices were used to examine the direction and concentration of classification errors across Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS traffic. Particular attention was given to confusion between Brute Force and Web Application Attack because these classes represent different credential-oriented and web-application threats that may exhibit overlapping network flow characteristics. Discrimination performance was further evaluated using one vs rest ROC-AUC and PR-AUC, including class level values and macro averages, to provide complementary assessments of ranking capability and precision-recall behavior under the naturally imbalanced test distribution.

Logistic Regression remained the primary focus of the detailed class-level and error analysis, while DT, RF, and Linear SVM were used only as comparative evidence to determine whether the principal class-separability and misclassification patterns observed with LR persisted across alternative classifier structures. Accordingly, the evaluation framework was intended to characterize the capabilities and limitations of network-flow-based classification rather than to identify an optimal classifier for telemedicine deployment.

III. RESULT AND DISCUSSION

This section presents and discusses the results obtained using the final leakage-aware experimental design. The analysis begins with the characteristics of the final training and test data, followed by the performance of the Logistic Regression baseline, class-specific prediction behavior, and misclassification patterns. Representative DT, RF, and Linear SVM results are subsequently used to contextualize the LR findings, followed by an interpretation of the implications and limitations of network flow based detection for web-based telemedicine security.

A. Final Dataset Characteristics

Following data cleaning, the domain-oriented CIC-IDS2017 subset contained 1,698,296 observations. The fixed stratified partition produced 1,358,636 training observations and 339,660 test observations. Training-only moderate random undersampling subsequently reduced the data used for model development to 17,885 observations, while the test partition remained unchanged. All classifiers used the same locked set of 63 numerical predictors obtained from the full training partition.

The final training subset retained all available Brute Force and Web Application Attack training observations while limiting the three majority classes to 5,390 observations each. In contrast, the untouched test partition preserved the original class imbalance, with Normal traffic accounting for approximately 75.59% of test observations and DoS/DDoS for 18.95%, while Brute Force and Web Application Attack

represented only approximately 0.09% and 0.04%, respectively. The final distributions used for model development and evaluation are summarized in Table 3.

TABLE 3
FINAL TRAINING AND TEST DISTRIBUTION

Class	Training After Moderate RUS	Untouched Test	Test Distribution (%)
Normal	5390	256741	75.59
Port Scan	5390	18139	5.34
Brute Force	1176	294	0.09
Web Application Attack	539	134	0.04
Dos/DDoS	5390	64352	18.95
Total	17885	339660	100.0

This configuration differs fundamentally from an artificially balanced evaluation setting. Random undersampling was used only to moderate majority-class dominance during model development, whereas the test data retained the distribution inherited from the cleaned benchmark. Consequently, overall accuracy must be interpreted together with macro-averaged and class-specific metrics, particularly for Brute Force and Web Application Attack, whose representation in the test partition is substantially smaller than that of Normal and DoS/DDoS traffic.

B. Classification Performance

The Logistic Regression baseline achieved an overall accuracy of 94.62% on the untouched test partition. However, because the test data were strongly imbalanced, accuracy alone provides an incomplete representation of performance across the five classes. The model obtained a Macro Precision of 56.39%, Macro Recall of 77.09%, Macro-F1 of 58.70%, and Weighted F1-score of 95.41%. The substantial difference between accuracy and Macro-F1 indicates that strong aggregate performance was influenced by the dominant classes, while detection performance varied considerably across the less frequent attack categories.

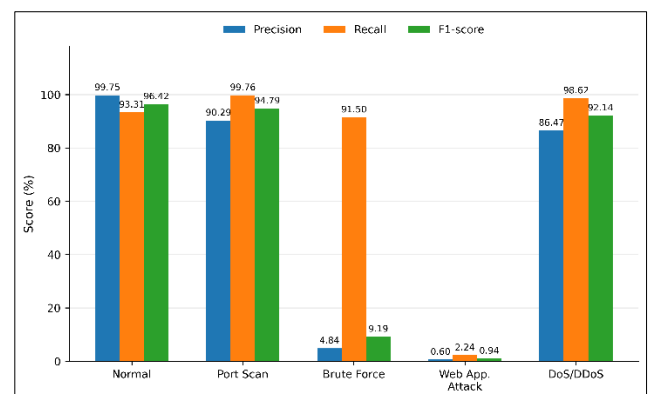


Figure 3 LR Class-Wise Performance

Figure 3 presents the class-specific precision, recall, and F1-score of the Logistic Regression baseline. Normal, Port Scan, and DoS/DDoS showed the strongest overall class-level performance, with F1-scores of 96.42%, 94.79%, and 92.14%, respectively. Brute Force exhibited a contrasting pattern, achieving a high recall of 91.50% but a very low precision of 4.84%, resulting in an F1-score of only 9.19%. This indicates that most actual Brute Force observations were detected, but a large number of observations from other classes were also assigned to the Brute Force category. Web Application Attack showed the weakest performance, with a precision of 0.60%, recall of 2.24%, and F1-score of 0.94%. These results demonstrate that the high overall accuracy of LR does not reflect uniformly strong discrimination across all threat categories.

The ranking based evaluation provided a complementary perspective. Logistic Regression achieved a macro one-vs-rest ROC-AUC of 99.40%, whereas the macro PR-AUC was 61.73%. The substantially lower PR-AUC is particularly relevant under the naturally imbalanced test distribution, in which Brute Force and Web Application Attack represent only a small fraction of the observations. Considered together with the Macro-F1 of 58.70% and the class-specific results in Figure 3, these findings indicate that the LR baseline provides strong separation for several broad threat categories but remains considerably less reliable for minority classes. The direction and concentration of these classification errors are examined further using the confusion matrix in the following subsection.

C. Confusion Matrix and Class Level Error Analysis

To examine the class-level behavior of the Logistic Regression baseline in greater detail, the prediction outcomes were further analyzed using a multiclass confusion matrix. This analysis complements the precision, recall, and F1-score results presented previously by showing not only how often each class was correctly identified, but also the direction in which misclassifications occurred. Such information is particularly important for the minority attack categories, whose aggregate metrics alone do not fully explain the underlying classification errors.

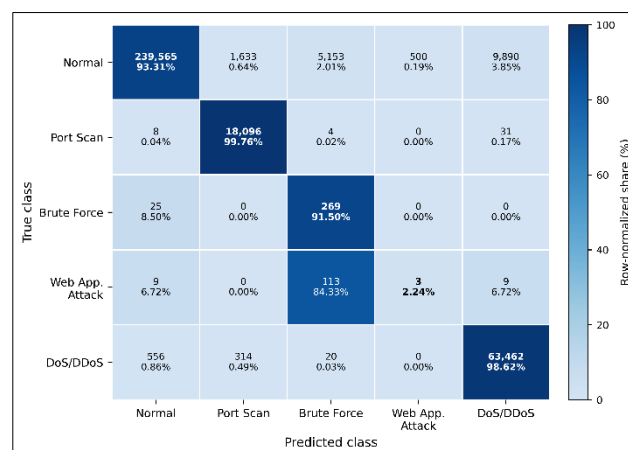


Figure 4 LR Confusion Matrix

Figure 4 shows that Normal, Port Scan, and DoS/DDoS were classified with strong class-level consistency. Specifically, 239,565 of 256,741 Normal observations (93.31%), 18,096 of 18,139 Port Scan observations (99.76%), and 63,462 of 64,352 DoS/DDoS observations (98.62%) were correctly classified. Brute Force also achieved high recall, with 269 of 294 observations correctly recognized (91.50%). These results indicate that the LR baseline was effective in identifying broad traffic patterns associated with benign activity, reconnaissance, and availability oriented attacks.

In contrast, Web Application Attack was the most problematic class. Only 3 of 134 observations were correctly classified, corresponding to a recall of 2.24%, while 113 observations (84.33%) were misclassified as Brute Force. This dominant error pattern explains the very low precision and F1-score observed for the two minority attack categories and shows that the main weakness of the LR baseline lies in separating attack classes with overlapping flow-level characteristics. From the perspective of web-based telemedicine security, this suggests that network flow features alone may be insufficient to reliably distinguish credential oriented attacks from application-layer exploitation, motivating the use of additional contextual information such as authentication logs, application events, or SIEM correlated evidence.

D. Comparative Validation

To contextualize the performance of the primary Logistic Regression baseline, three representative conventional classifiers Decision Tree, Random Forest, and Linear SVM were evaluated using the same locked feature set, Moderate RUS training partition, and untouched test partition. The comparison was intended to determine whether the performance characteristics observed with LR were primarily associated with its linear decision boundary or remained evident under alternative classifier structures. No hyperparameter optimization or model-selection procedure was performed for the comparative models.

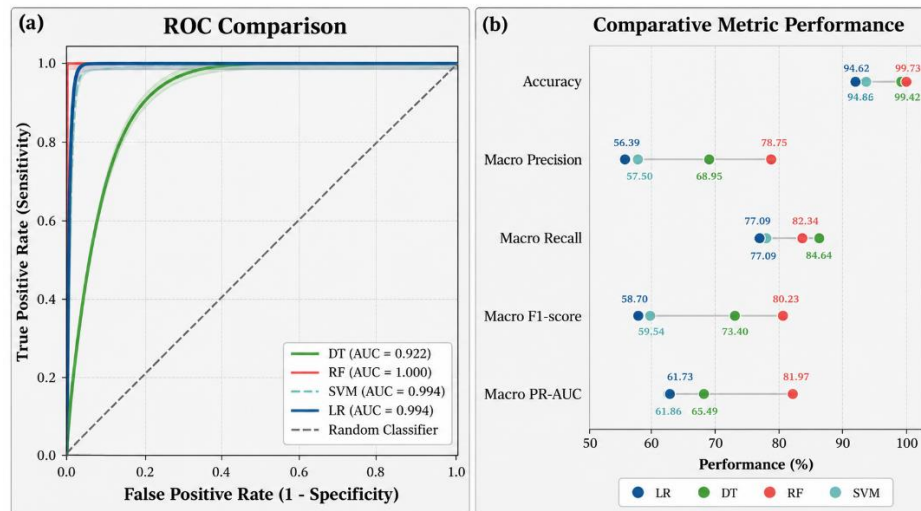


Figure 5 Comparative Performance of Logistic Regression and Representative Classifiers

Figure 5 shows that the four classifiers produced substantially different results when performance was considered beyond overall accuracy. LR achieved an accuracy of 94.62%, Macro-F1 of 58.70%, macro ROC-AUC of 99.40%, and macro PR-AUC of 61.73%. Linear SVM showed a closely comparable profile, with 94.86% accuracy, 59.54% Macro-F1, 99.36% macro ROC-AUC, and 61.86% macro PR-AUC. Decision Tree increased the observed accuracy to 99.42% and Macro-F1 to 73.40%, although its macro ROC-AUC was lower at 92.24%. Random Forest recorded the highest aggregate values in this experiment for accuracy (99.73%), Macro-F1 (80.23%), macro ROC-AUC (99.96%), and macro PR-AUC (81.97%). These results indicate that nonlinear tree-based decision structures improved aggregate class discrimination relative to the LR baseline, while the similar profiles of LR and Linear SVM suggest that part of the observed limitation is associated with linear class separation.

The comparative results nevertheless do not indicate uniformly reliable performance across all attack categories. Class-specific PR-AUC analysis shows that Web Application Attack remained substantially more difficult than the dominant classes across all four classifiers. Its PR-AUC was 4.30% for LR, 9.21% for Decision Tree, 26.00% for Random Forest, and 3.07% for Linear SVM. Brute Force also showed considerably lower PR-AUC than Normal, Port Scan, and DoS/DDoS for several classifiers. The improvement obtained by the tree-based models therefore reduced, but did not eliminate, the difficulty of distinguishing the minority credential oriented and web-application attack categories.

Taken together, the comparative validation supports the use of LR as an interpretable baseline while demonstrating that classifier structure influences the degree of separation achievable from the selected network flow representation. The results should not be interpreted as identifying an optimal classifier for telemedicine environments, since the

comparison was intentionally limited and no exhaustive hyperparameter optimization was conducted. Instead, the persistence of weaker minority-class discrimination particularly for Web Application Attack reinforces the need for subsequent validation using domain-specific information such as authentication events, application layer logs, and SIEM correlated security evidence.

E. Limitations and Research Implications

The findings of this study should be interpreted within the scope of a benchmark based network flow evaluation. CIC-IDS2017 provides reproducible attack traffic for cybersecurity research, but it does not represent traffic generated directly from a telemedicine application or healthcare information system. The five-class mapping used in this study therefore represents a functional abstraction of threat categories relevant to web-based telemedicine rather than a complete domain specific threat model. In particular, the Web Application Attack class contained only 673 cleaned observations and combined XSS with a very small number of SQL Injection observations, limiting the extent to which conclusions can be drawn for individual web-attack mechanisms.

A second limitation concerns the information contained in network flow features. Although these features provided strong discrimination for Normal, Port Scan, and DoS/DDoS traffic, they were considerably less effective for distinguishing Brute Force from Web Application Attack. Network flow records do not directly capture authentication outcomes, HTTP payload semantics, application sessions, user behavior, or security event context. Consequently, attacks that differ substantially at the application level may still exhibit overlapping flow characteristics. The use of Moderate RUS reduced the dominance of majority classes during training, while the untouched test partition retained the observed imbalance of the cleaned benchmark subset; nevertheless, the resulting performance should not be

interpreted as representing real-world telemedicine attack prevalence.

The comparative classifiers included in this study were used only to contextualize the behavior of the primary Logistic Regression baseline and to address whether the observed class-separation limitations persisted under alternative conventional classifier structures. The comparison was not designed as an exhaustive algorithm selection or optimization study. Future validation should therefore focus on transferring the identified threat categories to a domain-specific telemedicine environment and incorporating complementary evidence from authentication logs, application-layer events, and SIEM generated security records. Such validation would provide a more direct assessment of whether the class-level limitations identified in the benchmark setting persist when attacks are observed within an operational web-based telemedicine architecture.

IV. CONCLUSION

This study established a domain oriented cybersecurity baseline by mapping selected CIC-IDS2017 attack scenarios into five threat categories relevant to web-based telemedicine: Normal, Port Scan, Brute Force, Web Application Attack, and DoS/DDoS. Using a leakage aware experimental design with training-derived preprocessing, Moderate Random Undersampling applied only to the training partition, and evaluation on an untouched test set, the Logistic Regression baseline achieved an accuracy of 94.62%, Macro-F1 of 58.70%, macro ROC-AUC of 99.40%, and macro PR-AUC of 61.73%. Class-level analysis showed strong discrimination for Normal, Port Scan, and DoS/DDoS, whereas the principal limitation occurred in separating Brute Force from Web Application Attack. In particular, only 3 of 134 Web Application Attack observations were correctly identified by LR, with most misclassified as Brute Force.

Representative Decision Tree, Random Forest, and Linear SVM classifiers were additionally evaluated to contextualize these findings rather than to identify an optimal algorithm. The comparison showed that classifier structure influenced the degree of class separation, but difficulty in detecting the minority Web Application Attack class persisted across the evaluated models. These findings indicate that network flow information can provide a useful baseline for several telemedicine relevant cyberattack categories but is insufficient for consistently distinguishing threats whose differences are expressed primarily at the authentication or application layer. Future work should therefore validate these findings in a domain-specific web-based telemedicine environment by incorporating authentication records, application-layer logs, and SIEM-correlated security events, enabling a more direct assessment of cyberattack detection under healthcare-specific operational conditions.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia through the 2025 *Penelitian Dosen Pemula* (PDP) research grant, Contract No. 015/LL17/DT.05.00/PL/2025,057/C-01/USTI/VI/2025, from which this study was produced as one of the research outputs. The authors also thank the Canadian Institute for Cybersecurity (CIC), University of New Brunswick, for developing and making the CIC-IDS2017 dataset publicly available for cybersecurity research.

BIBLIOGRAPHY

- [1] "Telemedicine System Quality From Physicians' Perspectives: Key Determinants and Its Impact on Service Effectiveness via Interaction Quality | IEEE Journals & Magazine | IEEE Xplore." Accessed: Aug. 30, 2026. [Online]. Available: <https://ieeexplore.ieee.org/document/11021569>
- [2] "Communication during telemedicine consultations in general practice: perspectives from general practitioners and their patients | BMC Primary Care | Springer Nature Link." Accessed: Aug. 30, 2026. [Online]. Available: https://link.springer.com/article/10.1186/s12875-024-02576-1?utm_source=chatgpt.com
- [3] "Cyber threat assessment and management for securing healthcare ecosystems using natural language processing | International Journal of Information Security | Springer Nature Link." Accessed: Aug. 30, 2026. [Online]. Available: <https://link.springer.com/article/10.1007/s10207-023-00769-w>
- [4] Z. Wenhua *et al.*, "A lightweight security model for ensuring patient privacy and confidentiality in telehealth applications," *Comput. Hum. Behav.*, vol. 153, p. 108134, Apr. 2024, doi: 10.1016/j.chb.2024.108134.
- [5] K. Selvaraj and K. Kamatchi, "Hybrid Encryption for Privacy Preservation in Medical Image Transmission," in *2024 First International Conference on Data, Computation and Communication (ICDCC)*, Nov. 2024, pp. 723–729. doi: 10.1109/ICDCC62744.2024.10961779.
- [6] A. Prathik, S. P. Banu, B. S. Sagar, A. J. F. Prakash, R. Thamizhamuthu, and S. Velmurugan, "Telehealth Data Security and Privacy Solutions for Sensitive Health Records using Cloud Computing and Isolation Forest Algorithm," in *2024 2nd International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS)*, Oct. 2024, pp. 1199–1204. doi: 10.1109/ICSSAS64001.2024.10760862.
- [7] L. S. van Boven *et al.*, "Hacking Acute Care: A Qualitative Study on the Health Care Impacts of Ransomware Attacks Against Hospitals," *Ann. Emerg. Med.*, vol. 83, no. 1, pp. 46–56, Jan. 2024, doi: 10.1016/j.annemergmed.2023.04.025.
- [8] Y. Angler, S. Flessa, E. Grass, and O. Goetz, "Assessing the impact of technology partners on the level of cyberattack damage in hospitals," *Health Policy Technol.*, vol. 14, no. 1, p. 100955, Feb. 2025, doi: 10.1016/j.hlpt.2024.100955.
- [9] "Towards technology-driven business continuity management in the saudi healthcare sector through security and operational risk management - ScienceDirect." Accessed: Aug. 30, 2026. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2666188825009505?via%3Dihub>
- [10] "Dual layer security framework for privacy preserving AI driven healthcare edge cloud analytics | Discover Networks | Springer Nature Link." Accessed: Aug. 30, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s44354-025-00010-1?utm_source=chatgpt.com
- [11] O. F. Jeelani, "Anomaly Detection in Metaverse Healthcare for 6G-Enabled Internet of Things: An Approach Based on Feature

- Selection,” in *2025 13th International Symposium on Digital Forensics and Security (ISDFS)*, Apr. 2025, pp. 1–6. doi: 10.1109/ISDFS65363.2025.11012020.
- [12] Z. Wenhua *et al.*, “A lightweight security model for ensuring patient privacy and confidentiality in telehealth applications,” *Comput. Hum. Behav.*, vol. 153, p. 108134, Apr. 2024, doi: 10.1016/j.chb.2024.108134.
- [13] A. Hannousse, S. Yahiouche, and M. C. Nait-Hamoud, “Twenty-two years since revealing cross-site scripting attacks: A systematic mapping and a comprehensive survey,” *Comput. Sci. Rev.*, vol. 52, p. 100634, May 2024, doi: 10.1016/j.cosrev.2024.100634.
- [14] E. Owusu *et al.*, “Online Network DoS/DDoS Detection: Sampling, Change Point Detection, and Machine Learning Methods,” *IEEE Commun. Surv. Tutor.*, vol. 27, no. 4, pp. 2543–2580, Aug. 2025, doi: 10.1109/COMST.2024.3488580.
- [15] M. Smyrlis *et al.*, “RAMA: a risk assessment solution for healthcare organizations,” *Int. J. Inf. Secur.*, vol. 23, no. 3, pp. 1821–1838, Jun. 2024, doi: 10.1007/s10207-024-00820-4.
- [16] “Cyber threats in mobile healthcare applications: systematic review of enabling technologies, threat models, detection approaches, and future directions | Discover Computing | Springer Nature Link.” Accessed: Aug. 30, 2026. [Online]. Available: https://link.springer.com/article/10.1007/s10791-025-09686-z?utm_source=chatgpt.com
- [17] S. Albnhar, S. H. Bakry, H. Alaidaros, and M. Arafah, “Development of a Cybersecurity Assessment Framework for E-Health Critical Information Infrastructure Systems,” in *Proceedings of the 1st International Conference on Creativity, Technology, and Sustainability*, D. Berkouk, U. Chatterjee, T. A. K. Bouzir, and I. B. Dhaou, Eds., Singapore: Springer Nature, 2025, pp. 143–153. doi: 10.1007/978-981-97-8588-9_14.
- [18] “Machine learning-based network intrusion detection for big and imbalanced data using oversampling, stacking feature embedding and feature extraction | Journal of Big Data | Springer Nature Link.” Accessed: Aug. 30, 2026. [Online]. Available: <https://link.springer.com/article/10.1186/s40537-024-00886-w>
- [19] J. Talpini, F. Sartori, and M. Savi, “Enhancing trustworthiness in ML-based network intrusion detection with uncertainty quantification,” *J. Reliab. Intell. Environ.*, vol. 10, no. 4, pp. 501–520, Dec. 2024, doi: 10.1007/s40860-024-00238-8.
- [20] S. Elsayed, K. Mohamed, and M. A. Madkour, “A Comparative Study of Using Deep Learning Algorithms in Network Intrusion Detection,” *IEEE Access*, vol. 12, pp. 58851–58870, 2024, doi: 10.1109/ACCESS.2024.3389096.
- [21] “Enhancing Network Threat Detection with Random Forest-Based NIDS and Permutation Feature Importance | Journal of Network and Systems Management | Springer Nature Link.” Accessed: Aug. 31, 2026. [Online]. Available: <https://link.springer.com/article/10.1007/s10922-024-09874-0>
- [22] A. Shebl, E. I. Elsedimy, A. Ismail, A. A. Salama, and M. Herajy, “DCNN: a novel binary and multi-class network intrusion detection model via deep convolutional neural network,” *EURASIP J. Inf. Secur.*, vol. 2024, no. 1, p. 36, Dec. 2024, doi: 10.1186/s13635-024-00184-1.
- [23] J. Doménech, O. León, M. S. Siddiqui, and J. Pegueroles, “Evaluating and enhancing intrusion detection systems in IoMT: The importance of domain-specific datasets,” *Internet Things*, vol. 32, p. 101631, Jul. 2025, doi: 10.1016/j.iot.2025.101631.
- [24] M. Yacoubi, O. Moussaoui, and C. Drocourt, “AI for IoMT security: a comprehensive survey of intrusion detection and system architectures,” *Internet Things*, vol. 36, p. 101869, Mar. 2026, doi: 10.1016/j.iot.2025.101869.
- [25] C. M. Mejía-Granda, J. L. Fernández-Alemán, J. M. Carrillo de Gea, and J. A. García-Berná, “A method and validation for auditing e-Health applications based on reusable software security requirements specifications,” *Int. J. Med. Inf.*, vol. 194, p. 105699, Feb. 2025, doi: 10.1016/j.ijmedinf.2024.105699.