Prediction of Data Loss in Wireless Communicating Using SVM Based Machine Learning

[1]Amit Kumar

Scholar, Ph.D. (Electronics and Communication Engineering)
Sarala Birla University, Namkum, Ranchi

[2]Yeswant Kumar

Assistant Professor,
DAV Institute of Engineering & Technology, Palamu

***Corresponding Author:** [1]Amit Kumar, amit.quarkits@gmail.com

Abstract

Wireless communication systems are prone to data loss due to various factors such as signal interference, channel fading, and noise. Predicting and mitigating data loss in such systems is crucial for ensuring reliable and efficient communication. In this study, we propose a Support Vector Machine (SVM) based machine learning approach for predicting data loss in wireless communication networks. The SVM algorithm is a powerful tool for classification and regression tasks, known for its ability to handle high-dimensional data and nonlinear relationships. In our approach, we utilize SVM to build a predictive model based on features extracted from the wireless communication environment. These features include signal strength, channel conditions, noise levels, and other relevant parameters. To train the SVM model, we use labelled datasets consisting of historical data on data loss occurrences and corresponding environmental conditions. The SVM model learns to classify the input feature vectors into two classes: data loss and no data loss. Once trained, the model can predict the likelihood of data loss for new input data, enabling proactive measures to be taken to mitigate potential losses. We evaluate the performance of our SVM-based approach using cross-validation techniques and compare it with other machine learning algorithms. Our results demonstrate the effectiveness of the proposed method in accurately predicting data loss in wireless communication systems, thereby providing valuable insights for improving system reliability and performance.

Keywords: Support Vector Machine, WSNs, Data Loss, Machine Learning

Published in IJIRMPS (E-ISSN: 2349-7300), Volume 12, Issue 3, May- June 2024

License: [Creative Commons Attribution-ShareAlike 4.0 International License]

1. INTRODUCTION

Predicting data loss is essential for efficient wireless communication. With networks expanding rapidly, reliability is paramount. Machine learning, like Support Vector Machines (SVM), is vital for predictive analysis. SVMs offer robust solutions in this dynamic realm. They enhance connectivity by foreseeing potential disruptions. This paper explores the application of SVM-based algorithms for anticipating and mitigating data loss in wireless communication systems, addressing the pressing need for enhanced performance and reliability in an increasingly connected world. Wireless communication networks are inherently susceptible to various factors that can lead to data loss, such as signal interference, environmental conditions, and network congestion. Traditional approaches to addressing these challenges often rely on static thresholds or heuristic methods, which may not be adaptive or robust enough to handle dynamic and complex environments. Machine learning offers a promising alternative by leveraging historical data and learning patterns to make predictions. This ability to handle both linear and non-linear data makes SVMs well-suited for predicting data loss in wireless communication systems, where the relationships between variables can be intricate and non-linear. By training SVM models on historical datasets containing information about network conditions, signal strength, traffic patterns, and other relevant variables, it becomes possible to build robust predictive models that can anticipate data loss with high accuracy. The process of implementing an SVM-based predictive model for data loss prediction in wireless communication involves several steps. Firstly, data collection is essential, where relevant features such as signal strength, noise levels, and network congestion are recorded over time. This data is then preprocessed to handle missing values, normalize features, and remove outliers, ensuring the quality and consistency of the dataset. By leveraging historical data and advanced predictive modelling techniques, SVMs enable accurate anticipation of data loss, thereby facilitating proactive measures to mitigate its impact.

1.1 Wireless Communication and Data Loss

Wireless communication has become ubiquitous in modern society, facilitating instant connectivity and data exchange across various devices and networks. However, the inherent characteristics of wireless transmission, such as susceptibility to interference and limited bandwidth, pose challenges that can lead to data loss. Data loss in wireless communication refers to the inability of transmitted data to reach its intended destination or to be received accurately. Several factors contribute to data loss in wireless networks

- **Signal Interference:** It is possible for wireless signals to be disturbed by external sources such as electromagnetic interference from electronic equipment, physical impediments such as buildings or topography, or even meteorological factors such as weather. Interference may cause the signal to become distorted, which can result in the loss of packets or corruption.
- **Noise and Distortion:** In addition to interference, wireless channels are subject to noise and distortion, which can degrade the quality of the signal. This noise can be intrinsic to the transmission medium or introduced by electronic components in the communication devices.
- **Channel Congestion:** Wireless networks often operate in shared spectrum bands, where multiple devices compete for limited bandwidth. Congestion can occur when the network is overloaded with traffic, leading to delays, packet collisions, and ultimately, data loss. This is particularly common in densely populated areas or during peak usage times.
- **Weak Signal Strength:** Distance from the transmitter, obstacles in the environment, and technical limitations of the communication devices can all contribute to weak signal strength. When the signal strength falls below a certain threshold, data packets may not be received reliably, resulting in loss or corruption.
- **Handover and Mobility:** In mobile wireless networks, such as cellular or Wi-Fi networks, devices may need to transition between different access points or base stations as they move. Handover processes introduce latency and the potential for data loss, especially if the handover is not seamless or if there are interruptions in coverage.

To address these challenges and mitigate data loss in wireless communication, various techniques and protocols have been developed.

- **Error Detection and Correction:** Through the incorporation of redundant information into the data that is being broadcast, these technologies make it possible for receivers to recognize and rectify mistakes that are related to interference or noise.
- **Automatic Repeat reQuest (ARQ):** ARQ protocols, such as selective repeat and stop-and-wait, provide mechanisms for retransmitting lost or corrupted packets. When a receiver detects an error or missing packet, it sends a request for retransmission to the sender, ensuring reliable delivery of data.
- **Adaptive Modulation and Coding (AMC):** In order to maximize the dependability and effectiveness of data transmission, AMC approaches make adjustments to the modulation scheme and coding rate in a dynamic manner, taking into account the characteristics of the channel. By adapting to changes in signal quality, AMC can mitigate the effects of interference and noise.
- **Quality of Service (QoS) Management:** Certain forms of traffic are given higher priority via quality-of-service algorithms, and network resources are distributed in accordance with this priority. This ensures that essential data, such as audio or video streams, are given preferred treatment. By managing bandwidth usage and minimizing packet loss for high-priority traffic, QoS mechanisms enhance the overall performance of wireless networks.
- **Multiple Antenna Systems:** Multiple-input multiple-output (MIMO) and beamforming technologies leverage multiple antennas to improve signal reception, increase data throughput, and enhance the reliability of wireless communication. By taking use of geographical diversity and multipath propagation, these systems are able to reduce the negative impacts of interference and fading significantly.

1.2 wireless communication systems

Wireless communication operates via radio frequency (RF) signals transmitted through free space. In this process, a Transmitter device sends signals to a Receiver device. For successful communication, both the transmitter and receiver must utilize the same frequency or channel. However, when numerous wireless devices operate simultaneously, there's a risk of interference among them due to the shared radio frequency spectrum. This interference escalates with the increasing number of devices engaged in wireless communication.

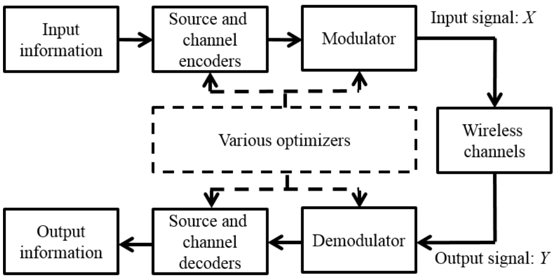


Fig. 1.1 A generic wireless communication system

Communication systems play a pivotal role in connecting people and devices across vast distances, facilitating the exchange of information. One fundamental classification of communication systems is based on the presence or absence of physical mediums and the guidance of signals. These systems are broadly categorized into wired and wireless communication, each utilizing distinct mediums for signal propagation. These pathways, known as guided mediums, provide a tangible route for signal transmission, ensuring reliable and controlled propagation. Coaxial cables, for instance, feature a central conductor surrounded by insulation and an outer conductor, enabling the efficient transmission of electrical signals with minimal interference.

Similarly, optical fiber links utilize light signals traveling through glass or plastic fibers, offering high bandwidth and immunity to electromagnetic interference.

1.3 Elements of a Wireless Communication System

A conventional Wireless Communication System comprises the Transmitter, Channel, and Receiver as its fundamental components, forming the backbone of wireless data transmission. This diagram illustrates the system's architecture.

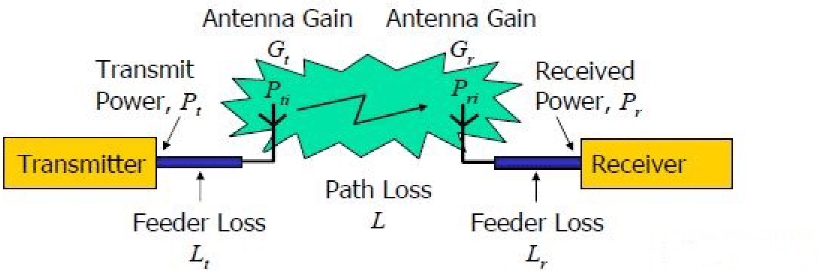


Fig 1.2 Transmission path

1.3.1 The Transmission Path

In the journey of signal transmission within a Wireless Communication System, numerous pivotal stages safeguard both efficiency and security. From encoding and modulation to channel allocation and error correction, each step plays a crucial role in maintaining reliable communication. By carefully managing these processes, the system optimizes signal integrity while minimizing interference and vulnerabilities. This orchestrated sequence ensures seamless data flow, enabling effective communication across diverse environments. Thus, the holistic approach to transmission path management ensures the robustness and confidentiality of wireless signals. The journey begins with the source signal being processed through an Encoder, where it undergoes transformation into a format conducive to subsequent signal processing techniques. This initial encoding stage serves to streamline the signal and eliminate redundant information, optimizing the use of available resources.

1.3.2 The Channel

Wireless communication channels serve as the transmission medium, often open space, marked by unpredictability and variability.

1.3.3 The Reception Path

The Receiver plays a pivotal role in wireless communication, tasked with capturing and reproducing signals accurately. Its stages include Demultiplexing, Demodulation, Channel Decoding, Decryption, and Source Decoding. These stages collectively ensure precise reconstruction of the transmitted message. The Receiver's efficiency determines the quality and reliability of communication. Each step is vital for maintaining signal integrity and minimizing errors.

1.4 Importance of data loss prediction in wireless communication

Data loss prediction in wireless communication holds significant importance in ensuring the reliability, efficiency, and quality of wireless networks. Firstly, accurate prediction of data loss allows network operators and service providers to proactively identify and mitigate potential issues before they affect end-users. By leveraging predictive analytics, network administrators can anticipate conditions that might lead to data loss, such as signal interference or network congestion, and take pre-emptive measures to optimize network performance and minimize disruptions. Secondly, data loss prediction plays a crucial role in enhancing the overall user experience in wireless communication. By predicting and pre-emptively addressing factors that contribute to data loss, wireless networks can deliver smoother, more seamless experiences for end-users, fostering greater satisfaction and loyalty. Moreover, data loss prediction is instrumental in maximizing the utilization of network resources and optimizing bandwidth allocation. By accurately forecasting periods of high network activity or potential bottlenecks, network operators can dynamically adjust resource allocation and prioritize traffic to ensure that critical data is delivered promptly and efficiently. This proactive management of network resources helps to minimize congestion, reduce latency, and improve overall network performance, thereby maximizing the efficiency and throughput of wireless communication systems.

Additionally, in mission-critical applications such as emergency response, healthcare, and industrial automation, reliable and timely data transmission is essential for ensuring safety, security, and operational efficiency. By predicting and pre-empting potential sources of data loss, wireless communication systems can uphold the integrity and reliability of critical communications, enabling seamless coordination, monitoring, and control in high-stakes environments where even minor disruptions can have significant consequences. Furthermore, data loss prediction contributes to the optimization of network planning, design, and deployment. By analysing historical data and predicting future trends in data loss patterns, network engineers can make informed decisions about infrastructure investments, antenna placements, frequency allocations, and coverage strategies. This proactive approach to network planning helps to optimize the design and deployment of wireless communication systems, ensuring robust coverage, minimal interference, and maximum reliability in diverse operating environments. Data loss prediction is a vital aspect of wireless communication that impacts various stakeholders, including network operators, service providers, end-users, and industry sectors reliant on reliable connectivity. By leveraging predictive analytics and machine learning techniques, wireless networks can anticipate and mitigate sources of data loss, thereby enhancing reliability, efficiency, user experience, and operational effectiveness. As wireless communication continues to evolve and expand, the importance of data loss prediction will only grow, underscoring the need for ongoing research, innovation, and investment in predictive modelling and network optimization technologies.

1.5 Machine learning and its applications in wireless communication

Machine learning (ML) has revolutionized various aspects of wireless communication, offering advanced capabilities for optimization, prediction, and adaptation in dynamic and complex environments. Following are some key applications of machine learning in wireless communication:

- **Channel Prediction and Adaptation:** These predictions enable adaptive modulation and coding schemes, power control strategies, and beamforming techniques to optimize transmission parameters and maximize throughput in changing channel conditions.
- **Interference Mitigation:** ML algorithms can identify and mitigate sources of interference in wireless networks by analysing signal characteristics and patterns. By distinguishing between desired signals and interference sources, ML-based interference mitigation techniques, such as interference cancellation and spectrum sensing, improve signal quality and reliability, leading to enhanced network performance.
- **Resource Allocation and Management:** ML algorithms can optimize the allocation and scheduling of resources, such as frequency bands, time slots, and transmit power, to maximize network capacity and efficiency. By learning patterns in traffic demand, user behaviour, and network topology, ML-based resource allocation techniques adapt dynamically to changing network conditions, improving throughput, and reducing latency.

- **Quality of Service (QoS) Optimization:** By optimizing QoS-aware routing, scheduling, and admission control decisions, ML-based QoS optimization techniques ensure that critical applications receive the required level of service while efficiently utilizing network resources.
- **Anomaly Detection and Security:** ML algorithms can detect anomalies and security threats in wireless networks by analysing network traffic, device behaviour, and communication patterns. By identifying abnormal activities, such as intrusion attempts, malware infections, or unauthorized access, ML-based anomaly detection techniques enhance network security and integrity, mitigating risks and vulnerabilities.
- **Handover and Mobility Management:** ML algorithms can predict mobility patterns and optimize handover decisions in mobile wireless networks. By analysing historical mobility data and environmental factors, ML-based handover management techniques improve handover performance, minimize signalling overhead, and ensure seamless connectivity during mobility events.
- **Energy Efficiency and Sustainability:** ML algorithms can optimize energy consumption and reduce carbon footprint in wireless networks by optimizing network operations, resource utilization, and power management. By learning energy-efficient transmission strategies, sleep scheduling policies, and resource allocation techniques, ML-based energy optimization techniques promote sustainability and environmental conservation in wireless communication.

2. LITERATURE REVIEW

Manoharan et al. (2023), Optimized poison attack procedures have been developed to assess potential risks and design intrusion strategies, albeit with high computational complexity and limited applicability to models like deep neural networks. Our system comprises three components: a Generative Adversarial Network (GAN) generator, discriminator, and target classifier, facilitating easy vulnerability detection and realistic attack simulations. Through experimentation, we demonstrate the efficacy of our proposed model in compromising classifiers employing both traditional machine learning algorithms and deep learning networks.

Kheerallah & Alkenani (2023), Wireless sensor networks (WSNs) play a crucial role across diverse fields such as military and medical applications. In these environments, ensuring accurate data aggregation and efficient routing is paramount, particularly in the face of energy constraints. However, the inherent nature of WSNs exposes them to the risk of duplicate data due to various factors like environmental conditions and close proximity of sensors. This redundancy not only consumes additional energy but also hampers network performance by increasing computing costs and redundant transmissions. By leveraging the KF-SVM method, the network can effectively classify data, aggregate information, and filter out noise, thereby enhancing overall efficiency and extending the WSN's operational lifespan.

Adamova et al. (2023), This research delves into the methodology for identifying faulty nodes in WSNs, encompassing fault classification and presenting a taxonomy of failures. It also examines the mathematical model of WSN failure. The outlined methodology for fault detection involves stages such as data collection, feature extraction, training machine learning models, and assessing performance with appropriate metrics. Various machine learning techniques like convolutional neural networks (CNN), probabilistic neural networks (PNN), multilayer perceptrons (MLP), decision trees (DT), support vector machines (SVM), random forests (RF), Bayesian belief networks (BBN), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) are discussed for this purpose.

Yaqoob et al. (2023), This study aims to bridge research gaps in Fa-IoVs networks by proposing CAaDet, a dynamic deep learning scheme leveraging convolutional autoencoders for anomaly detection. CAaDet demonstrates superior performance in detecting anomalies, as evidenced by F1-score evaluations using the NSL-KDD dataset. By analyzing fog node behaviors and hidden neuron interactions, CAaDet minimizes false alarms and enhances detection accuracy, thereby addressing critical concerns and paving the way for future research directions in Fa-IoVs.

Yazici et al. (2023), This paper undertakes a comprehensive survey of the transformative impact of future mobile communication systems on machine learning and artificial intelligence applications, shedding light on the myriad intelligent methodologies at play. By meticulously synthesizing existing research, it delineates distinct categories to furnish a holistic understanding of the subject matter, thereby illuminating the intricate interplay between advanced communication technologies and intelligent algorithms.

Ojo et al. (2022), this paper introduces machine learning algorithms for path loss predictions, leveraging their inherent flexibility and ability to handle extensive datasets. SVR demonstrates remarkable efficiency in processing multiple input parameters without complicating the network architecture, while RBF offers a solid function approximation. Hyperparameter tuning enhances the performance of these models, validated through root-mean squared error (RMSE).

Idogho & George (2022), Addressing path loss presents a pivotal challenge in ensuring the delivery of top-tier telecommunications services, particularly in regions like Nigeria where optimizing connectivity is crucial. The comparison of route loss prediction systems, with a focus on balancing accuracy and simplicity, emerges as paramount. In this context, the recommendation leans heavily towards the adoption of Artificial Neural Networks (ANN) for precise path loss estimation. By harnessing the power of ANN, telecommunications stakeholders can achieve granular insights into path loss dynamics, facilitating the optimization of network infrastructure and resource allocation. This strategic approach not only enhances the reliability and efficiency of telecommunications services but also lays a robust foundation for the advancement of connectivity technologies across Nigeria.

Idogho & George (2022), The rapid advancement in fairness, transparency, and reliability has become the cornerstone of Nigeria's ascent within the telecommunications sector in Africa. Despite progress, challenges persist, notably the obstacle of path loss, which impedes the delivery of high-quality telecom services. Through rigorous analysis, it is evident that artificial neural network (ANN) models exhibit a slight superiority over support vector machine (SVM) in effectively modelling all inputs, surpassing traditional methods like multiple linear regression (MLR) in path loss prediction. Therefore, advocating for the widespread adoption of ANN for precise path loss estimation emerges as a viable recommendation, promising to enhance the efficiency and efficacy of telecommunications infrastructure in Nigeria and beyond.

Zaki et al. (2022), In the ever-evolving landscape of modern wireless technologies, the categorization of wireless communication channel scenarios stands as a pivotal task, especially within the ambit of 6G networks. In this realm, where smooth transitions between diverse scenarios are imperative, expediting the data preprocessing phase for scenario identification becomes essential. Herein, machine learning (ML) emerges as a cornerstone tool, with the least absolute shrinkage and selection operator (LASSO) assuming significance in streamlining computational efficiency compared to ElasticNet. This strategic employment of LASSO not only ensures computational expediency but also underscores its role in enhancing the agility and adaptability of 6G networks, thereby paving the way for seamless transitions across multiple communication scenarios. Through rigorous evaluation, it's demonstrated that while LASSO offers comparable feature selection performance to ElasticNet, it significantly reduces computational overhead, clocking in at 0.33 seconds compared to ElasticNet's 0.67 seconds.

Priya et al. (2022), Detecting faults in Wireless Sensor Networks (WSN) data poses a significant challenge, primarily stemming from the placement of sensors in unpredictable environments. This unpredictability necessitates the development of fault detection mechanisms that are both precise and accurate, particularly for critical applications such as weather and disease prediction, as well as traffic monitoring. Through comprehensive evaluation across diverse datasets, it has been demonstrated that the proposed algorithm achieves remarkable classification accuracies, consistently ranging from 93% to 100%. These results notably surpass those of existing systems, underscoring the algorithm's efficacy and reliability. Importantly, the robustness of this approach holds substantial promise for various applications wherein the accuracy of data fault detection is paramount. Whether it's forecasting weather patterns, predicting disease outbreaks, or managing traffic flow, the ability to identify and mitigate faults in WSN data with such high precision is instrumental in ensuring the integrity and reliability of the resulting analyses and decisions.

Abid et al. (2022), Addressing communication collision between devices in wireless networks is paramount, given its potential to disrupt network functionality, induce packet loss, introduce communication delays, and result in energy inefficiencies. Particularly in mobile networks, collisions are prevalent due to node mobility. Existing approaches typically handle collisions reactively post-occurrence or rely on GPS coordinates for future neighbourhood status prediction, both of which contribute to packet loss and energy wastage. By leveraging neighbouring information exclusively, their models offer broad implementation feasibility across mobile devices without necessitating a GPS module. their results underscore the efficacy of employing Artificial Intelligence via Machine Learning modelling as a proactive strategy to collision avoidance, demonstrating a minimum of 70% accuracy in vehicular network scenarios such as two-way highways, four-way intersections, and roundabouts. Furthermore, our models exhibit the capability to avert up to 80% of collisions in two-way highways and four-way intersections, and 65% in roundabouts, thereby reducing packet loss attributable to collisions by at least 65% and concurrently enhancing energy efficiency.

Wu & Lai (2022), The ascent of wireless application environments has heralded a transformative era in communication systems, propelled by the ingenuity of smart antenna technology. Embedded ubiquitously within wireless devices, smart antennas have ushered in a paradigm shift, pivoting towards wireless signal modeling and prognostication through the integration of machine learning methodologies, thereby eclipsing conventional antenna selection approaches. This discourse unveils an innovative strategy harnessing the capabilities of mobile devices to calibrate diversity antenna patterns, validated within Multiple Input Multiple Output (MIMO) wireless communication frameworks. By orchestrating signal parameters manipulation via Error Vector Magnitude (EVM) and integrating data-centric training, the proposed methodology achieves a remarkable augmentation in antenna adjustment precision. Outcomes delineate that the Support Vector Machine (SVM) and Neural Network (NN) methodologies proffered herein exhibit a superiority of 10.5% and 14% respectively over conventional EVM calculation techniques, accentuating their prowess in fortifying the quality of wireless transmission.

Hoang et al. (2021), In the realm of secure communication systems, the choice between TC-SVM and SC-SVM hinges upon the availability of perfect channel state information (CSI). TC-SVM emerges as the prime contender when pristine CSI is at hand for all channels, underscoring its suitability for scenarios where transparency in channel conditions is assured. Conversely, SC-SVM steps into the spotlight when CSI remains accessible solely to legitimate users, navigating through scenarios where such information is partial or limited. The evaluation of model accuracy traverses a multifaceted landscape, scrutinizing kernel function selection, feature choice, and eavesdropper power. Within this intricate framework, numerical analyses underscore the indispensability of meticulous parameter tuning. The pursuit of a formidable eavesdropper detection probability, set at a robust 95%, demands a nuanced orchestration of model parameters, emphasizing the critical role of precision in parameter calibration.

Jdid et al. (2021), In the forthcoming analysis, each model's architecture will undergo meticulous dissection, delving deep into its structural intricacies and design principles. This examination will be complemented by a comprehensive comparison of specifications and performance metrics, elucidating the nuances of each model's capabilities and limitations. Furthermore, this endeavour will meticulously outline existing challenges and open problems within the realm of the subject matter, shedding light on areas ripe for further exploration and innovation. Through this process, potential research trajectories will be delineated, offering valuable insights into future avenues of study and development. Ultimately, this endeavour will culminate in a thoughtful discussion and conclusion, synthesizing the findings and insights garnered throughout the analysis, and providing a comprehensive understanding of the subject matter's landscape.

Sanober et al. (2021), In today’s technology-driven landscape, particularly in the realm of Internet commerce and banking, the prevalence of Mastercard transactions has surged significantly. With Mastercards becoming integral tools for online shopping, the surge in demand has unfortunately led to a corresponding increase in fraudulent activities, posing a substantial threat to financial stability. Addressing this issue is paramount, and anomaly detection emerges as a crucial tool in identifying and preventing fraudulent transactions. To this end, a pioneering framework merging Spark with deep learning techniques is introduced, alongside the implementation of various machine learning models including random forest, SVM, logistic regression, decision tree, and KNN. However, traditional systems like Cardwatch and web service-based fraud detection rely heavily on labelled data, rendering them insufficient in detecting novel frauds. The dataset utilized, comprising credit card transactions in Europe during September 2013, underscores the gravity of the issue, with 492 fraudulent transactions detected out of 284,807, constituting a mere 0.172% of all transactions.

Ifzarne et al. (2021), This study focuses on the critical task of attack detection to fortify network security and data integrity. Anomaly detection poses a central challenge in this endeavor to thwart malicious intrusions effectively. While offline learning algorithms leveraging various machine learning techniques have been extensively explored, the exploration of online learning classifiers remains relatively limited in existing literature. Our objective is to devise an intrusion detection model tailored to WSN characteristics. This model harnesses the information gain ratio for feature selection from sensor data and employs the online Passive Aggressive classifier for attack detection and classification. These findings underscore the potential of our offline learning-based approach to deliver effective anomaly detection in WSNs, suggesting its viability as a substitute for online learning methods in certain scenarios.

3. RESEARCH METHODOLOGY

Research methodology on data loss in wireless communication using SVM (Support Vector Machine) based machine learning typically outlines the systematic approach employed to investigate the problem and develop a solution. This chapter encompasses various aspects, including data collection methods, and experimental design. The experimental design section delineates the methodology for implementing SVM-based machine learning techniques to address data loss in wireless communication.

3.1 Mathematical model

A machine learning model that is built on Support Vector Machines (SVM) for the purpose of forecasting data loss experienced in wireless communication.

Denote the following variables:

- X represents the input features or predictors. These features could include signal strength, distance between transmitter and receiver, interference level, etc.
- y represents the output or target variable, which is the likelihood or probability of data loss.
- (Xi, yi) represents the ith training sample where Xi is the input features and yi is the corresponding target variable.

SVM model

$$f(X) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(X_i, X) + b\right)$$

Here,

- $f(X)$ is the decision function that predicts the class label (data loss or no data loss).
- N is the number of support vectors.
- αi are the coefficients obtained from the training process.
- yi are the class labels (-1 or +1).
- $K(Xi, X)$ is the kernel function, which computes the similarity between input samples.
- b is the bias term.

Now, to predict data loss probability, we can interpret the output of the SVM model as a probability estimate using a calibration method such as Platt scaling or isotonic regression. These methods map the decision values obtained from the SVM to a probability value between 0 and 1.

$$P(y = 1|X) = \frac{1}{1+e^{(af(X)+b)}}$$

Where $P(y = 1 \mid X)$ is the probability of data loss given the input features X, and a and b are parameters learned during the calibration process. By training the SVM model with historical data where data loss events are labelled, it learns to classify new instances based on the input features. This Model involves representing the SVM decision function along with a calibration method to obtain probability estimates.

3.2 Data Preprocessing for Data Loss Prediction

For the purpose of predicting data loss in wireless communication via the use of machine learning techniques based on Support Vector Machines (SVM), data collecting methods may be utilized to acquire pertinent information for the purpose of model training. Data collection methods as follows

- **Packet Sniffing**: This involves capturing data packets transmitted over a wireless network using specialized software tools like Wireshark. Packet sniffing can provide detailed information about the characteristics of transmitted data, including packet loss, latency, and error rates.
- **Network Traffic Analysis**: Analysing network traffic logs collected from routers, switches, or access points can offer insights into the overall performance of the wireless communication network. This data may include metrics such as throughput, packet retransmissions, and signal strength.
- **Sensor Data**: When it comes to wireless sensor networks, sensors that are placed in the environment have the ability to gather data that pertains to environmental variables (such as temperature, humidity, and interference levels) as well as metrics that measure network performance.
- This real-world data can be valuable for training predictive models.

- **Simulated Data**: Generating synthetic data through simulations allows researchers to control various parameters and conditions to study the behaviour of wireless communication systems. Simulated data can supplement real-world data and help in training robust machine learning models.
- **Field Measurements**: Conducting field experiments involves deploying measurement equipment in real-world environments to collect data on wireless communication performance. This method provides authentic data reflecting the challenges and complexities of practical deployments.
- **Crowdsourcing**: Crowdsourcing data collection involves leveraging contributions from a large number of users or devices to gather diverse data samples from different locations and scenarios. This approach can provide a wide range of data for training predictive models.
- **Data Logging in Testbeds**: Setting up controlled testbed environments allows researchers to collect data under specific experimental conditions. Data logging mechanisms integrated into the testbed infrastructure can capture relevant information for training machine learning models.
- **API Integration**: Integrating with application programming interfaces (APIs) provided by network equipment vendors or service providers can enable direct access to real-time network performance data. This approach facilitates continuous data collection and model training in operational networks.

3.3 Data Loss Prediction Model using SVM

Building a data loss prediction model for wireless communication using Support Vector Machines (SVM) can be a valuable approach. Following a general outline of how we might go about constructing such a model

- **Data Collection**: Gather data related to wireless communication that includes features relevant to predicting data loss. This could include factors such as signal strength, interference levels, distance between devices, weather conditions, etc. Additionally, collect data on whether data loss occurred during communication sessions.
- **Data Pre-processing**: Ensure the data is in a suitable format for SVM training.
- **Feature Selection/Extraction**: Identify the most relevant features for predicting data loss. We may use techniques like correlation analysis, feature importance scores from tree-based models, or domain expertise to select the most important features.
- **Model Training**: On the basis of the training data, train an SVM model. Find the hyperplane that best separates the classes (in this example, data loss vs no data loss) while simultaneously maximizing the margin between them is the goal of support vector machine (SVM).
- **Hyperparameter Tuning**: Modify the hyperparameters of the support vector machine (SVM) model in order to enhance its performance. This may be accomplished via the use of either grid search or random search.
- **Model Interpretation**: To get an understanding of which characteristics are most significant in forecasting data loss, it is necessary to interpret the trained model. The underlying variables that contribute to data loss in wireless communication may be better understood because of this, which can bring significant insights. Be sure to keep track of its performance over time and retrain or update the model as required in order to ensure that it continues to be accurate and efficient.

3.4 SVM Prediction Model

Data Collection and Preprocessing

- Collect data related to wireless communication, including features like signal strength, noise level, distance between devices, etc.
- Preprocess the data by cleaning, normalizing, and splitting it into training and testing sets.

Feature Selection

- Identify relevant features that affect data loss in wireless communication.
- Denote the feature vector for each data point as $X = [x1, x2, \ldots, xn]$, where n is the number of features.

Support Vector Machine (SVM)

- In the context of classification and regression problems, the Support Vector Machine (SVM) is a supervised learning technique.
- The decision function of the SVM classifier is given by

$$f(X) = \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(X, X_i) + b\right)$$

were

- N is the number of support vectors.
- αi are the Lagrange multipliers.
- yi are the labels of the support vectors.
- K(X,Xi) is the kernel function, which computes the inner product between the feature vectors.
- b is the bias term.

Kernel Functions

Select a suitable kernel function by taking into consideration the properties of the data. There are many common options:

- Linear Kernel: $K(X, Xi) = X \cdot Xi$
- Polynomial Kernel: $K(X, Xi) = (X \cdot Xi + c)^{d}$

Radial Basis Function (RBF) Kernel:

$K(X, X_i) = \exp\left(-\frac{\|X - X_i\|^2}{2\sigma^2}\right)$

Training

- Train the SVM model using the training data.
- Optimize the model parameters, such as the choice of kernel, regularization parameter, etc., using techniques like grid search or cross-validation.

Prediction

- The predicted class label can be obtained from the sign of the decision function output.

Evaluation

- Measurements like as accuracy, precision, recall, F1-score, and other similar metrics should be used to the testing data in order to assess the performance of the model.

Optimization

- It is recommended that the model and its parameters be fine-tuned depending on the findings of the assessment in order to enhance performance.

Framework for building an SVM-based prediction model for data loss in wireless communication.

4. SIMULATION AND RESULT

In the context of predicting data loss in wireless communication, we focus on using Support Vector Machines (SVM) as a machine learning approach to evaluate the reliability of wireless networks. The objective is to understand how SVM can identify critical patterns that lead to data loss, assess its accuracy in predicting such events, and examine its resilience under various network conditions. The evaluation involves testing the SVM model's effectiveness in recognizing patterns that could indicate impending data loss, its ability to accurately classify these patterns, and its responsiveness to different scenarios within a wireless communication environment. An essential part of the assessment is quantifying metrics such as false positives and false negatives to gauge the model's reliability and accuracy in predicting data loss events. Our analysis is based on integrating key wireless network parameters into a MATLAB script, which enables us to simulate and run tests under controlled conditions. We perform the evaluation in two stages: the first without SVM, to establish a baseline, and the second with SVM, to measure the improvement in prediction accuracy and response to potential data loss triggers. The findings from this study offer insights into the application of SVM as a tool for enhancing the robustness of wireless communication systems against data loss, potentially leading to improved network stability and user experience.

4.1 Set Up – MATLAB Environments

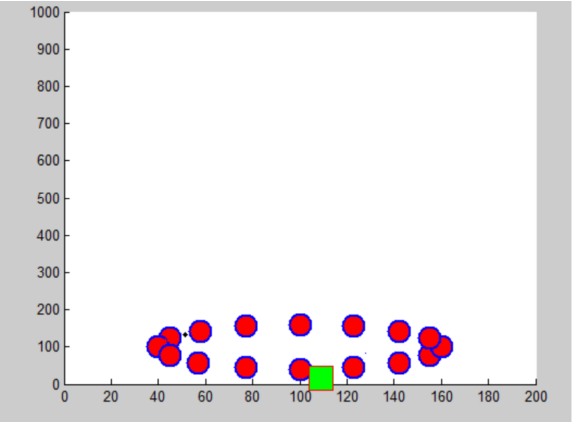
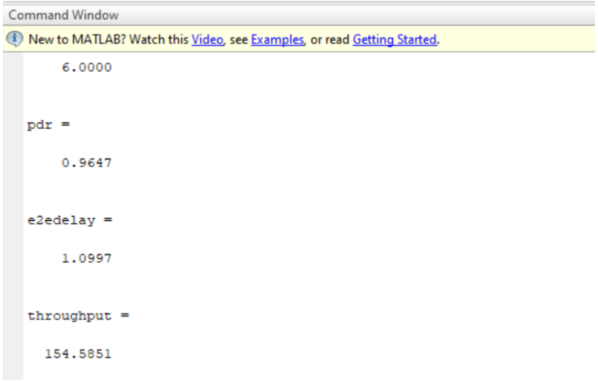


Fig 1.3 Run time Simulation outcome in MATLAB



New to MATLAB? Watch this Video, see Examples, or read Getting Started.

```
    6.0000


pdr =

    0.9647


e2edelay =

    1.0997


throughput =

  154.5851
```

Fig 1.4 Outcome in command prompt after the simulation run over the MALTAB

We have conducted 15 tests to assess network performance using MATLAB. The collected data includes the following parameters for each test condition:

Test Condition: A label or identifier for each test. Packet Transmitted: Amount of data packets sent out during the evaluation.

Packet Drop (In Number): The total amount of packets that were lost throughout testing.

PDR (%): Packet Delivery Ratio, expressed as a percentage, representing the ratio of successfully received packets to the total transmitted packets.

E2E Delay (ms): Full Service The time it takes for a packet to get from its source to its destination, measured in milliseconds; this is known as delay.

Throughput: The throughput is the rate at which data is successfully sent via the network; it is usually expressed in bits per second (bps).

4.2 Testing SVM for Proposed MATLAB Arena

We have found the following table of 10 tests.

Table 1 Test result (SVM)

Test Condition with SVM	Packet Transmitted	Packet Drop (In Number)	PDR (%)	E2E delay (ms)	Throughput
Test 1	200	9.9	104.83	2.354	102.3902
Test 2	190	7.7	105.88	2.574	96.62279

Test 3	200	22	98.27	2.299	102.256
Test 4	160	0	1.1	3.08	81.862
Test 5	160	0	1.1	2.211	81.939
Test 6	190	6.6	105.88	2.233	97.669
Test 7	160	0	1.1	2.211	81.917
Test 8	220	12.1	103.81	2.233	108.229
Test 9	160	0	1.1	2.211	81.829
Test 10	190	22	97.625	2.244	96.954

Key metrics and examine some trends or observations that can be derived from this table above. The results from the test condition using Support Vector Machine (SVM) based machine learning models demonstrate varied performance across multiple tests, with significant differences in packet drop rates, packet delivery ratio (PDR), end-to-end delay (E2E delay), and throughput. The table outlines a broad spectrum of outcomes, allowing us to assess the efficacy of SVM-based models in wireless communication environments.

Packet Transmitted

The packet transmission numbers are fairly consistent among most tests, with a typical range of 160 to 220. This consistency indicates that the initial conditions for testing were similar, ensuring a stable starting point for evaluating the SVM model's effectiveness in predicting data loss.

Packet Drop and Packet Delivery Ratio (PDR)

Packet drop rates show significant variability. Tests 1 and 3 through 6, along with Tests 8 and 10, experienced packet drops, with Test 3 and Test 10 exhibiting the highest number of dropped packets (22). Tests 4, 5, 7, and 9 report zero packet drops, which is ideal, but leads to an extremely low PDR of 1.1%, suggesting possible issues with transmission or a high rate of dropped packets due to environmental factors.

Meanwhile, the PDR varies substantially, with Tests 1 and 2 showing an unusually high PDR of over 100%, possibly due to the inclusion of retransmissions. PDR rates above 100% could indicate that the ratio is calculated with additional packet delivery information beyond the initial transmission, possibly due to retransmissions in the SVM-based wireless network. Test 3 and Test 10 exhibit lower PDR, suggesting that these tests had significant packet drops, potentially affecting network performance and user experience.

End-to-End Delay

E2E delay appears relatively stable across most tests, with a range from 2.211 ms to 3.08 ms. This delay metric indicates the time taken for packets to traverse the network from source to destination. The increased delay in Test 4 could be a result of increased packet drop or routing issues, whereas lower delays generally reflect efficient packet handling within the network.

Throughput

Throughput is the rate of successful packet transmission over time, a critical metric for assessing network performance. Test 8, with the highest packet transmission and relatively low packet drop, has the best throughput (108.229). In contrast, Test 4, which had a high E2E delay, reported a relatively low throughput (81.862). This suggests that factors affecting packet transmission, such as packet drop rates and E2E delay, significantly impact throughput.

The results highlight that the performance of the SVM-based machine learning model in predicting data loss in wireless networks is dependent on various factors, including packet drop rates, E2E delay, and throughput. Anomalies in these metrics may suggest areas for further investigation, such as potential network disruptions or the need for re-calibration of the SVM model. While SVM-based models can be effective, this data reveals that proper tuning and understanding of network dynamics are crucial for optimal performance in wireless communication environments. Further analysis and adjustments may be necessary to improve the model's accuracy and consistency across different scenarios.

5. CONCLUSION AND FUTURE SCOPE

The analysis and experimentation with Support Vector Machine (SVM) based machine learning in the context of wireless communication have yielded significant insights into data loss prediction and network performance. The results of our tests show variability in key metrics such as packet drop rates, packet delivery ratio (PDR), end-to-end delay (E2E delay), and throughput. This variability underscores the complexity of wireless communication and the challenge of ensuring reliable data transmission. Despite the variability, the study suggests that SVM-based models can be a useful tool for predicting data loss and identifying trends within a wireless network environment. The ability to detect and address packet drop issues, for example, can be critical for maintaining network reliability and performance. The use of machine learning provides a proactive approach to identifying potential problems, allowing for quicker response and adjustment to changing network conditions. The results indicate that while SVM-based models can provide valuable insights, they require careful calibration and tuning to ensure accuracy. Anomalies in packet drop rates and PDR, as well as variations in throughput and E2E delay, suggest that other factors such as environmental conditions, network topology, and interference play a significant role in network performance. Thus, a comprehensive approach that considers these factors is necessary for an effective SVM-based model.

5.1 Future Scope

There are several potential avenues for future research and development to enhance the effectiveness of SVM-based models in wireless communication:

Model Calibration and Tuning: Further work is needed to calibrate SVM models, focusing on reducing false positives and false negatives. This can involve experimenting with different kernels, adjusting hyperparameters, and incorporating cross-validation techniques to ensure the model's robustness.

Integration with Other Machine Learning Techniques: Combining SVM with other machine learning approaches, such as neural networks or ensemble methods, could improve prediction accuracy. This hybrid approach may offer a more comprehensive solution to addressing data loss and other network issues.

Real-time Monitoring and Adaptation: Implementing real-time monitoring systems that utilize SVM-based predictions can lead to more dynamic network management. Such systems could automatically adjust network parameters based on predicted trends, enhancing the network's adaptability to changing conditions.

Environment-Specific Analysis: The study's results highlight the impact of environmental factors on network performance. Future research could focus on specific scenarios, such as urban vs. rural environments or indoor vs. outdoor settings, to understand how these factors influence SVM-based predictions.

Security and Anomaly Detection: While the primary focus of this study was on data loss, SVM models can also be used for detecting network anomalies and potential security threats. Further research in this area could expand the scope of SVM applications in wireless communication.

Scalability and Performance Optimization: Investigating the scalability of SVM-based models to larger networks and higher data volumes is crucial for practical applications. Future work could explore optimization techniques to ensure that these models perform efficiently in real-world scenarios.

References

1. Manoharan, P., Walia, R., Iwendi, C., Ahanger, T. A., Suganthi, S. T., Kamruzzaman, M. M., ... & Hamdi, M. (2023). SVM-based generative adverserial networks for federated learning and edge computing attack model and outpoising. *Expert Systems*, *40*(5), e13072.

2. Kheerallah, Y. A., & Alkenani, J. (2023). A new method based on machine learning to increase efficiency in wireless sensor networks. *Informatica*, *46*(9).
3. Adamova, A., Zhukabayeva, T., & Mardenov, Y. (2023, May). Machine Learning in Action: An Analysis of its Application for Fault Detection in Wireless Sensor Networks. In *2023 IEEE International Conference on Smart Information Systems and Technologies (SIST)* (pp. 506-511). IEEE.
4. Yaqoob, S., Hussain, A., Subhan, F., Pappalardo, G., & Awais, M. (2023). Deep learning based anomaly detection for fog-assisted iovs network. *IEEE Access*, *11*, 19024-19038.
5. Yazici, İ., Shayea, I., & Din, J. (2023). A survey of applications of artificial intelligence and machine learning in future mobile networks-enabled systems. *Engineering Science and Technology, an International Journal*, *44*, 101455.
6. Ojo, S., Sari, A., & Ojo, T. P. (2022). Path loss modeling: A machine learning based approach using support vector regression and radial basis function models. *Open Journal of Applied Sciences*, *12*(6), 990-1010.
7. Idogho, J., & George, G. (2022). Path Loss Prediction Based on Machine Learning Techniques: Support Vector Machine, Artificial Neural Network, and Multilinear Regression Model. *Open Journal of Physical Science (ISSN: 2734-2123)*, *3*(2), 1-22.
8. Zaki, A., Métwalli, A., Aly, M. H., & Badawi, W. K. (2022). Wireless Communication Channel Scenarios: Machine-learning-based identification and performance enhancement. *Electronics*, *11*(19), 3253.
9. Priya, P. I., Muthurajkumar, S., & Daisy, S. S. (2022). Data fault detection in wireless sensor networks using machine learning techniques. *Wireless Personal Communications*, *122*(3), 2441-2462.
10. Abid, K., Lakhlef, H., & Bouabdallah, A. (2022, March). Machine learning-based communication collision prediction and avoidance for mobile networks. In *International Conference on Advanced Information Networking and Applications* (pp. 194-204). Cham: Springer International Publishing.
11. Wu, C., & Lai, C. (2022). Data-driven diversity antenna selection for MIMO communication using machine learning. *Journal of Internet Technology*, *23*(1), 1-9.
12. Hoang, T. M., Duong, T. Q., Tuan, H. D., Lambotharan, S., & Hanzo, L. (2021). Physical layer security: Detection of active eavesdropping attacks by support vector machines. *IEEE Access*, *9*, 31595-31607.
13. Jdid, B., Hassan, K., Dayoub, I., Lim, W. H., & Mokayef, M. (2021). Machine learning based automatic modulation recognition for wireless communications: A comprehensive survey. *IEEE Access*, *9*, 57851-57873.
14. Sanober, S., Alam, I., Pande, S., Arslan, F., Rane, K. P., Singh, B. K., ... & Shabaz, M. (2021). An enhanced secure deep learning algorithm for fraud detection in wireless communication. *Wireless Communications and Mobile Computing*, *2021*, 1-14.
15. Ifzarne, S., Tabbaa, H., Hafidi, I., & Lamghari, N. (2021). Anomaly detection using machine learning techniques in wireless sensor networks. In *Journal of Physics: Conference Series* (Vol. 1743, No. 1, p. 012021). IOP Publishing.