ISSN: 2182-2069 / E-ISSN: 2182-2077

Novel ML-based Detection System for Identifying Malicious Pdf Documents in Financial Cloud

Dr. Prerna Mahajan[1*], Madhur Taneja[2], Dr. Arun Kumar Marandi[3], Sohel Das[4], Madhur Grover[5], and S. Aswath[6]

[1*]Professor, Department of Computer Science and Information Technology, Jain (Deemed to be University), Bangalore, Karnataka, India. prerna.m@jainuniversity.ac.in, https://orcid.org/0000-0002-2408-024X

[2]Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India. madhur.taneja.orp@chitkara.edu.in, https://orcid.org/0009-0000-2931-7122

[3]Associate Professor, Department of Computer Science & IT, ARKA Jain University, Jamshedpur, Jharkhand, India. dr.arun@arkajainuniversity.ac.in, https://orcid.org/0000-0003-1507-7580

[4]Assistant Professor, Department of uGDX, ATLAS SkillTech University, Mumbai, Maharashtra, India. sohel.das@atlasuniversity.edu.in, https://orcid.org/0000-0001-9111-6524

[5]Chitkara Centre for Research and Development, Chitkara University, Himachal Pradesh, India. madhur.grover.orp@chitkara.edu.in, https://orcid.org/0009-0008-3520-4667

[6]Assistant Professor, ECE Department, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. professoraswath@gmail.com, https://orcid.org/0000-0001-8170-6862

Received: April 23, 2025; Revised: June 06, 2025; Accepted: July 25, 2025; Published: August 30, 2025

Abstract

The growing dependability of financial institutions on cloud-based services for document management has given rise to significant safety concerns over the potential for malicious PDF documents. In response, this research suggests a Machine Learning (ML) based detection technique for spotting erroneous PDF files in a financial cloud environment. The recommended solution leverages modern technologies to enhance the security posture of financial institutions by efficiently recognizing and mitigating potential hazards. The methodology analyzes the attributes of malicious PDF documents and the final relevant qualities that indicate malicious intent using the Sparse K-Nearest Neighbor (SKNN) algorithm. The PDF files are pre-processed to ascertain whether the samples have errors or duplicate content. Principal Component Analysis (PCA) is used to extract the features of PDF files. A user employs files that can avoid detection by a security system and send a variety of payloads that have the potential to do serious damage. Comprehensive testing and assessment in terms of ROC-AUC (98.7%), F1-score (98%), recall (97%) and detection accuracy (98.5%) demonstrate the effectiveness of the recommended strategy. The finding's outstanding performance demonstrates how ML techniques can be used to improve cloud financial system cybersecurity safeguards.

Journal of Internet Services and Information Security (JISIS), volume: 15, number: 3 (August), pp. 218-229. DOI: *10.58346/JISIS.2025.I3.015*

*Corresponding author: Professor, Department of Computer Science and Information Technology, Jain (Deemed to be University), Bangalore, Karnataka, India.

Keywords: Artificial Intelligence (AI), Software Detection, Malicious, Sparse K-Nearest Neighbour (SKNN), Financial Institutions, Cloud Security.

1 Introduction

The phrase was first used in 1972 to describe the nexus between information technology and finance. It serves as a catch-all for any financial services technology that is connected digitally as a consequence. Since then, fintech has been widely used by clients, and this trend is anticipated to continue. The developments in the finance industry enable services like digital payments and banking. However, the suppliers of fintech applications depend on a suitable technological infrastructure to implement them. Fintech solutions can profit from developments in the cloud computing space. Large data processing capacities are provided by cloud systems, along with improved security and availability (Stojanovic & Božic, 2022). The Information and Communication Technology (ICT) industry is very concerned about Distributed Denial-of-Service (DDoS) attacks. It is obvious that what happens when banking, telecommunications, relaxation, and commercial services are attacked. DDoS attacks are becoming more common and have a direct impact on the dependability of modern society's operations. The increasing centralization of cloud-based architectures' services and applications dealing with DDoS is getting increasingly challenging (Corrêa et al., 2021). Human resources (HR) analytics has become a crucial component of all HR operations, including recruitment, orientation, training, planning for succession, benefits, wages, engagement, and retention. Analytics has been used to examine hiring costs and attrition rates. HR analytics gives businesses the ability to forecast the future of the workforce and quantify the commercial effect of people policy using sophisticated statistical analysis. This makes it possible for managers to evaluate whether the organization's financial success and its HR policies relate to one another (Sharma et al., 2019).

The use of innovation increases the risk to financial institutions' financial technology and innovation-related activities. The complexity of financial services and products creates a knowledge gap between providers and clients, creating distinct dangers from recognized threats. Tightening the controls is necessary to ensure that the management and evaluation procedures are used (Mishchenko et al., 2021). In today's environment, financial cybersecurity is incredibly useful and plays an important part in safeguarding prevent financial loss for consumers and corporations. Moreover, economic internet safety is an essential part of the security of information that guards against malicious actors and inadvertent acts, along with safeguarding digital financial transactions, data, and networks. To safeguard client data, banking organizations should guarantee the security of their computer networks, thwart deception, and preserve their outstanding morals (Koibichuk & Dotsenko, 2023). Analogous information must be transformed into digital form as part of the process of digitization. The financial sector has been significantly impacted by these procedures. The current relationship between statistics and finance is significant due to the integration of increasingly resource-intensive technology, including artificial intelligence (Choudhary & Verma, 2025). (AI) and cryptocurrencies. Data is sent during financial transactions, and financial infrastructures such as payment and stock exchange systems can be thought of as data networks (Kun, 2024). Adobe Systems created the PDF file format in 1993 to facilitate the storing of several kinds of electronic data, including form fields, buttons, and links. PDF files may be viewed or printed by anyone, irrespective of their technology, program, or operating system. To execute Spear Phishing attacks, Remote Code Execution (RCE) attacks, and other attacks, attackers typically insert malicious code inside normal PDF files. They distribute this code via email or other channels (Jiang et al., 2023). Figure 1 depicts the process of malicious contaminants constituting hazards (Narayanan & Rajan, 2024).

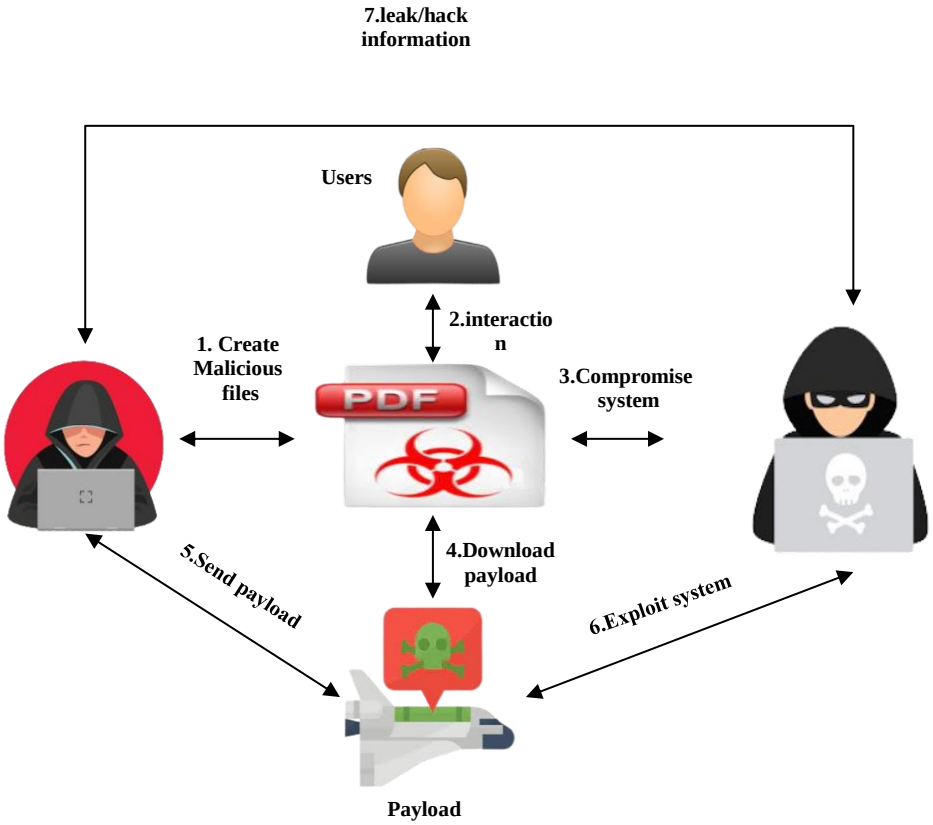

Figure 1: Process of Malicious Contaminants Constituting Hazards

The sophisticated tools and strategies employed in these types of security breaches have also increased. Such sophisticated assaults cannot be stopped or detected by conventional cybersecurity breach prevention measures like firewalls and anti-virus software. Installing hardware and/or software to continually check the network for attacks and intrusions is one technique to handle this issue (Dina & Manivannan, 2021). Thwarting intrusions and cyberattacks on connected Internet of Things (IoT) devices, machine learning techniques may identify infiltration and cyberattacks in the industrial IoT, enhancing durability and ensuring uninterrupted services. Devices connected to detectors linked to large network servers inside an IoT-based application framework are vulnerable to threats and malicious assaults (Khatkar et al., 2022). The study aims to improve financial cloud security by identifying and addressing threats from malicious PDF documents, safeguarding sensitive data, and ensuring regulatory compliance.

2 Related Works

Ahmad et al., (2024) described the secure policies of the cloud security framework (SPCSF), which has a secure application management framework for cloud security. SPCSF's basis was the establishment of fine-grained control over application permissions, as well as the encryption of Representational State Transfer Application Programming Interface (REST API) connections, to enhance defence against malicious assaults. BN & SH, (2023) advocated employing semantic similarity algorithms to evaluate attack intensity and identify attack relationships, to create a complete multiclass ransomware detection model capable of automatic or semi-automated responses (Soy & Balkrishna, 2024). Tripathy et al., (2020) investigated the possibility of employing ML methods to identify Structured Query Language

(SQL) injections at the application level. A range of moderate and hazardous payloads was used to train the predictive techniques under evaluation. The organization ascertains whether or not the input contains malware by gazing at the payload (Aswathy, 2024).

Gupta et al., (2020) investigated the models of ML and deep learning (DL) based strategies for identifying and countering any new or emerging risks. By employing training and test datasets derived from traffic movements across multiple fields, ML and DL-based algorithms can make informed judgments on threat identification and mitigation. Falah et al., (2021) identified the most important feature set required for accurate PDF maldoc categorization, and the classification of PDF maldocs was accomplished. To extract features and present a set of data-supported characteristics to improve subsequent classification with the help of two widely used PDF analysis tools. Next, it assesses each attribute using an abstraction procedure. Nguyen et al., (2023) evaluated the efficacy of ML techniques, provided approaches for cloud-based malware recognition, and suggested a feasible framework for wireless recognition of malware. ML techniques have been employed to solve this problem and have proven to be successful in identifying malware in a variety of scenarios. Al Haque et al., (2023) investigated the efficacy of several ML subfields in malware detection (Udayakumar et al., 2023). Malicious attackers utilize files that can avoid detection systems and deliver a variety of payloads that have the potential to do serious damage.

3 Methodology

The sample of cloud financial PDF documents, pre-processing of the malicious PDF document, feature extraction using principal component analysis (PCA), detection and identification of malicious documents using sparse k-nearest neighbor (SKNN) are included in the methods section. Figure 2 illustrates the suggested methodology's flow.

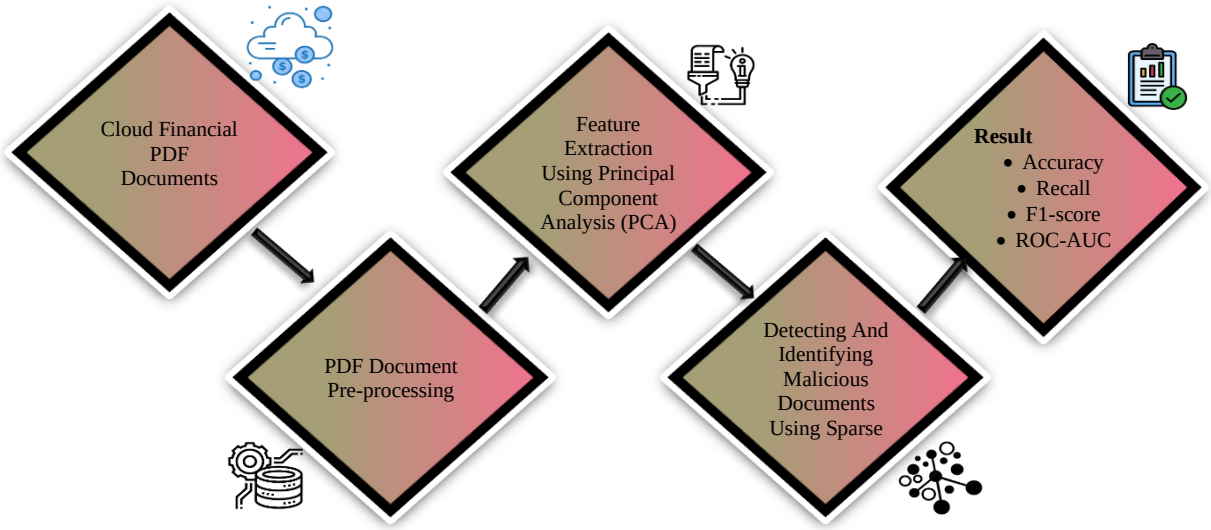

Figure 2: Flow of Proposed Methodology

3.1 Dataset

Datasets were collected from several websites, such as 1565 Cloud, financial Malicious PDF documents. Research utilized a supervised classification approach to detect samples as malware. The samples were

divided into two datasets: training and testing. The training set contained 925 samples, while the testing set consisted of 640 samples from a shared malware library.

3.2 Data Pre-processing

For our purpose, pre-processing the document into free text (.txt) and using Cloud Financial Malicious PDF to arrange the data in a manner consistent with the PDF layout is sufficient. Unlike PDF to HTML, PDF to Text does not produce extra information to identify the original location of the material. Line by line, content is arranged in a text document that maintains the preliminary file's structure, so aligning the content's vertical and horizontal placements. Equation (1) shows the empty regions that appear in the Cloud financial Malicious PDF versions. Whitespace characters and blank lines are utilized in the written content to substitute for them. According to this perspective, a text file is represented as a matrixC_{ji}, in which the width i represents the maximum number of characters that may fit on a single line, and the height j represents the number of lines:

$$C = \begin{bmatrix} This & is \\ an & example \end{bmatrix} \quad (1)$$

As a result, every character in the document space has coordinates. The decision to make a matrix for every page can be taken to optimize the procedure. Preprocessing was performed on the gathered Cloud finance malicious PDF instances.

3.3 Principal Component Analysis is Used to Extract Features (PCA)

After preprocessing, we extract the features using PCA. A linear dimensionality reduction technique called PCA intends to find translation in the directions of the highest variability. The information distribution matrix's eigenvalues and eigenvectors are calculated throughout this procedure. Eigenvectors are sorted according to Eigenvalue in descending order, and next, the categorized eigenvectors' orientations are transferred onto the actual dataset.

Below is a step-by-step guide to the PCA procedure:

Step 1: For the M-dimensional signal samples, compute the covariance matrix of order $M \times M$ as

Here x is the mean vector of the provided signal matrix, which is made up of N data points of dimension M each.

Step 2: Determine the diagonal elements of the matrix $C$and matrix U of eigenvectors as the provided eigenvalues of the covariance matrix$Q$.

Step 3: Arrange the eigenvectors according to $C's$ eigenvalues in decreasing order.

Step 4: Applying a dot product between the information supplied and the Eigenvectors, project the data into the directions of the sorted Eigenvectors.

Step 5: Determine whether the problem has a specific percentage of variability (e.g., 95% or 98%) and then choose the first few primary components. The cloud financial malicious PDF files features and dimensionality are effectively extracted and reduced by PCA.

3.4 Detecting and Identifying Malicious PDF Documents by using Sparse K-Nearest Neighbour (SKNN)

The characteristics are used to create a sparse representation of the Cloud financial malicious PDF document once they have been extracted. The technique we employed to detect and determine malicious

PDF documents was Sparse K-Nearest Neighbor (SKNN). When categorizing materials using the KNN rule, an evaluation text is allocated to the class with the highest membership among the KNN. The majority voting discrimination criteria have no specific limitations. In rare instances, when there is a tie between the competing groupings, the researcher randomly gives the test material to the competing category. A document is categorized into a class when it contains material from previously published papers in that class. There are frequently a lot of unnecessary words in every document in a collection. As a result, the content similarity of the two publications may not be indicative of their genuine similarity. The SKNN rule may determine a document's class label incorrectly when the number of members in the contending classes is equal or different. Thus, our confidence in our judgment would be higher if the variance were just 1 or 2, and research could distribute the evaluation documents to a group with the greatest number of SKNN.

There are two phases in the recommended method. As it may take a few files from the sample used for the training set to determine a specific document's class, ensuring that every neighbor has at least some content similarity with the test file is possible with this initial stage. The process determines if the test document's neighborhood resemblance is above a specific threshold, represented by θ. If such documents are not found in the current set of neighbors, the procedure generates a new Set of Pruning Neighbors (SPN) and deletes them from the current set. Let the collection of N training documents should be represented by $(d_1, d_2 \ldots d_N)$ and the test document by d_t. Subsequently, SPN is developed as follows.

The second step of the process involves choosing the initial$L$papers from SPN. The document is allocated to the strongest contending class if, across these L neighbors, the variation in the number of competitors in the highest and distant second categories is larger than the value of the criterionβ. If not, the identical procedure is carried out again with L incremented by one. This procedure continues until it gets to the final SPN document, at which point it stops. If certain test set documents don't match the requirements, they won't be categorized after the test set ends.

The program starts by calculating the separations between each document in the training set and the test document. The documents in the collection of neighbors are filtered out if their distances exceed a predetermined thresholdθ. The remaining documents are subsequently preserved in S. Next, S is sorted in increasing dissimilarity order to form SPN. To conduct a majority vote, β documents are first gathered from SPN under the requirement that the total amount of documentation that is different between the two groups in competition ($Lx1\ and\ Lx2$) be equal to β among SPN. The document determination is given to the class ($cx1$) with the maximum number of SPN members if this is the case. If not, it proceeds to carry out the same operation and look through the next text in SPN. The process is carried out again till it reaches the final SPN document. When the algorithm reaches the final document in the SPN, it fails to assign an instance of an attribute if $(Lx1 - Lx2) < \beta$.

Examine the subsequent instance for $\beta = 2$. Given two classes, c_1, andc_2 and there are eleven papers in SPN, and these are their class labels:

At the beginning, $Lx1 = 1, Lx2 = 1$, and $(Lx1 - Lx2 < 2)$.

Consequently, to do the same procedure, L needs to be increased by one and so on. Consequently, $Lx1 = 6$ and $Lx2 = 5$ indicates that no categories can be applied to the test document in the end. However, using SKNN, a document with $k = 3$can be categorized as c_1It can be a draw for$k = 4$, a document with $k = 5$determination is classed as c_2, and so on. It makes clear that voting by simple majority could not be suitable. In general, we feel that the test document shouldn't be placed in one of the contending classes when there is almost equal representation from those classes among the

neighbors. It makes sense that the suggested system has a higher chance of accurate categorization than SKNN. To qualify for SKNN to classify a document, it must have a sufficient number of neighbors who endorse the relevant class; the document won't be categorized if it doesn't. Having sufficient relevant content for the test document and its neighbors is the goal of the SKNN reduced search procedure. On the other hand, a notable discrepancy in the quantity of papers belonging to the majority class and its rival classes is bolstering trust in the resolution process. The following SKNN methods were used to identify the malicious PDF document related to cloud finance and provide more accurate assessments.

4 4 Results and Discussion

In this section, research evaluates the performance of SKNN with existing approaches, Random Forest (RF) and Multilayer Perceptron (MLP) (Torres and De Los Santos, 2018). The suggested model was created with Python software and evaluated according to several criteria, including accuracy, ROC-AUC, recall, and F1 score. Research evaluated each algorithm's performance using four metrics to identify its advantages and disadvantages. Table 1 shows the overall numerical findings of accuracy, recall, F1-score, and ROC-AUC.

Table 1: Numerical Findings of Accuracy, F1-score, Recall, and ROC-AUC

Methods	Accuracy (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
RF	92	94	92	98
MLP	96	96	96	98
SKNN (Proposed)	98.5	97	98	98.7

a) Accuracy

The accuracy of a classifier can be used to gauge its efficiency; it displays the proportion of properly classified malicious PDF document samples to all samples in the test dataset. Figure 3 shows the accuracy comparison with various approaches. Equation (2) obtains the calculation of accuracy. The suggested technique has an SKNN value of 98.5 %, which is superior to the accuracy of other existing approaches like RF (92%) and MLP (96%).

$$Accuracy = \frac{TP+TF}{TP+TN+FP+FN} \quad (2)$$

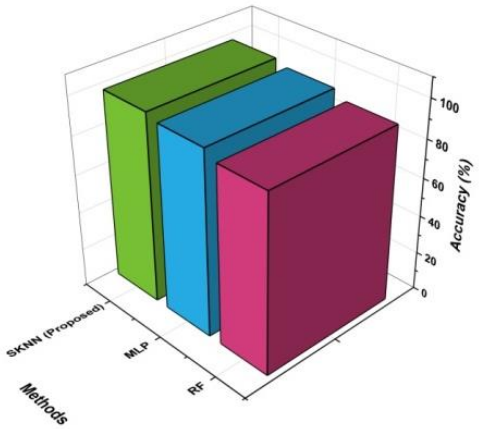

Figure 3: Comparison of Accuracy

b) Recall

Recalls evaluate a model's capacity to precisely predict whether a particular malicious PDF document sample in a dataset corresponds to a target sample or not. Based on the study findings, it is calculated as the proportion of positive results to the sum of all true benefits and erroneous negatives. Equation (3) may be used to obtain the recall rate. The relative recall measures are shown in Figure 4. It reveals that the RF contains 94%, and the MLP obtains 96%. In terms of recall, the suggested method, SKNN, outperformed the state-of-the-art, coming in at 97%.

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

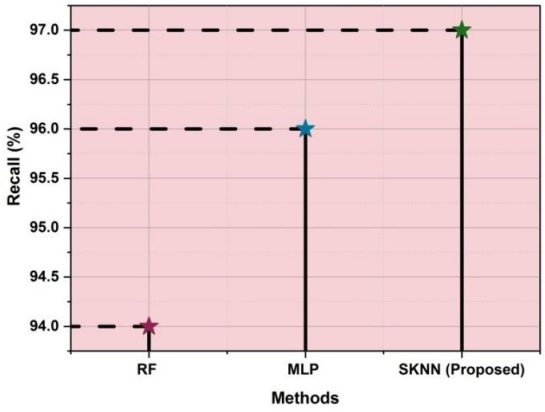

Figure 4: Evaluation of Recall

c) F1-score

In the evaluation process, the F1 score is frequently employed. Precision is given a higher weight than recall, or vice versa, in the F1-score. The recommended approach receives a higher F1 score than standard procedures. Someone can calculate their F1-score by figuring out equation (4). The comparison of the F1 score is displayed in Figure 5. The proposed technique has an SKNN value of 98%, which is much higher than the F1 score of other existing methods, such as RF (92%) and MLP (96%).

$$F1 - Score = \frac{precision \cdot recall}{precision+recall} \quad (4)$$

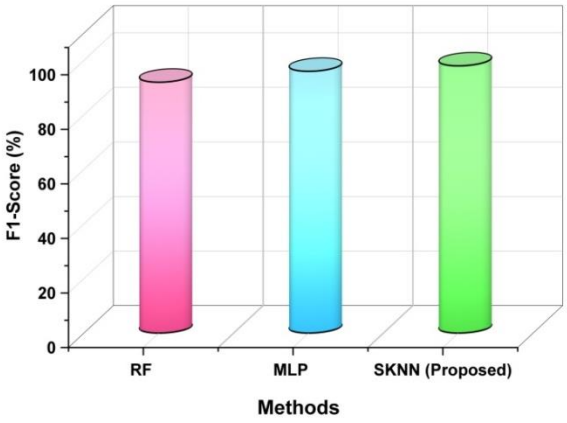

Figure 5: Evaluation of F1-score

d) ROC-AUC

The detection system's overall performance is measured across all potential threshold settings using the area under the ROC curve (AUC). Greater discriminating capacity is shown by a higher AUC value, which means the detection system maintains lower false positive rates while achieving greater true positive rates of malicious PDF documents. The comparison of the ROC-AUC is displayed in Figure 6. The suggested method's SKNN obtained 98.7%, which is significantly higher than the ROC-AUC of other methods presently in consumption, such as MLP and RF contain 98%.

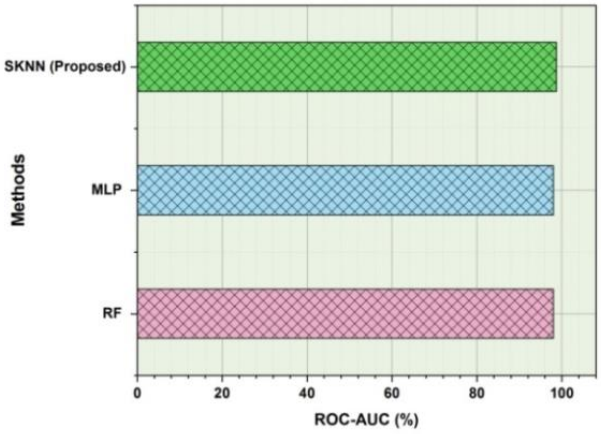

Figure 6: Comparison of ROC-AUC

4.1 Discussion

The potential of present methods, such as RF and MLP, to identify fake PDF documents in financial clouds is extremely constrained. Although RF has limits with high-dimensional feature spaces, it is dependable and scalable to huge datasets. However, MLP is more prone to overfitting and necessitates extensive hyperparameter adjustment, particularly in situations when training data is scarce. SKNN is advantageous for these types of research as it can evaluate large amounts of data promptly and effectively without demanding extensive variable evaluation. In particular, in financial cloud systems, SKNN is useful for detecting suspicious patterns in PDF files as it can handle sparse data representations rapidly from a security perspective.

5 Conclusion

In this study, the malicious PDF documents in financial cloud establishment and consumption of an ML-driven Detection System SKNN, intended to recognize fraudulent PDF documents in financial cloud settings, represent an important advance in cybersecurity protocols. The system has shown encouraging consequences, in particular identifying and mitigating the threats posed by malicious PDF files by utilizing cutting-edge ML methods and strategies, including DL and natural language processing. The techniques section includes a sample of cloud financial PDF documents, pre-processing of the malicious PDF document, and principal component analysis (PCA) for feature extraction. A user utilizes files that can deliver a range of payloads with the ability to do significant harm and evade detection by security systems. Extensive testing and evaluation concerning detection ROC-AUC (98.7%), F1-score (98%), recall (97%) and accuracy (98.5%) show that the suggested SKNN is more effective compared to existing approaches. Furthermore, using mathematical formulas and models improves the identification

system's quantitative and forecasting capacity. The system consistently distinguishes between malicious and harmless PDF files using mathematical concepts and algorithms for anomaly detection, classification, and feature extraction.

References

[1] Ahmad, S., Mehfuz, S., Urooj, S., & Alsubaie, N. (2024). Machine learning-based intelligent security framework for secure cloud key management. *Cluster Computing*, 1-27. https://doi.org/10.1007/s10586-024-04288-8

[2] Al Haque, A. F. (2023). Optimizing the Performance of Machine Learning Algorithms in Detecting Malicious Files using Hybrid Models.

[3] Anaya Menon, A., & Srinivas, K. (2023). Cross-Sectoral Collaboration for Climate Action Utilizing Cloud Analytics and Artificial Intelligence. In *Cloud-Driven Policy Systems* (pp. 1-6). Periodic Series in Multidisciplinary Studies.

[4] Aswathy, S. (2024). Bibliometric Analysis of Sustainability in Business Management Policies Using Artificial Intelligence. *Global Perspectives in Management*, *2*(1), 44-54.

[5] BN, C., & SH, B. (2023). Revolutionizing ransomware detection and criticality assessment: Multiclass hybrid machine learning and semantic similarity-based end-to-end solution. *Multimedia Tools and Applications*, 1-34. https://doi.org/10.1007/s11042-023-16946-x

[6] Choudhary, N., & Verma, M. (2025). Artificial Intelligence-Enabled Analytical Framework for Optimizing Medical Billing Processes in Healthcare Applications. *Global Journal of Medical Terminology Research and Informatics*, *3*(1), 1-7.

[7] Corrêa, J. H., Ciarelli, P. M., Ribeiro, M. R., & Villaça, R. S. (2021). Ml-based ddos detection and identification using native cloud telemetry macroscopic monitoring. *Journal of Network and Systems Management*, *29*, 1-28. https://doi.org/10.1007/s10922-020-09578-1

[8] Dina, A. S., & Manivannan, D. (2021). Intrusion detection based on machine learning techniques in computer networks. *Internet of Things*, *16*, 100462. https://doi.org/10.1016/j.iot.2021.100462

[9] Falah, A., Pan, L., Huda, S., Pokhrel, S. R., & Anwar, A. (2021). Improving malicious PDF classifier with feature engineering: A data-driven approach. *Future Generation Computer Systems*, *115*, 314-326. https://doi.org/10.1016/j.future.2020.09.015

[10] Gupta, R., Tanwar, S., Tyagi, S., & Kumar, N. (2020). Machine learning models for secure data analytics: A taxonomy and threat model. *Computer Communications*, *153*, 406-440. https://doi.org/10.1016/j.comcom.2020.02.008

[11] Jiang, T., Liu, Y., Wu, X., Xu, M., & Cui, X. (2023). Application of deep reinforcement learning in attacking and protecting structural features-based malicious PDF detector. *Future Generation Computer Systems*, *141*, 325-338. https://doi.org/10.1016/j.future.2022.11.015

[12] Khatkar, M., Kumar, K., & Kumar, B. (2022). Unfolding the network dataset to understand the contribution of features for detecting malicious activities using AI/ML. *Materials Today: Proceedings*, *59*, 1824-1830.https://doi.org/10.1016/j.matpr.2022.04.391

[13] Koibichuk, V. V., & Dotsenko, T. V. (2023). Content and Meaning of Financial Cyber Security: a Bibliometric Analysis. http://doi.org/10.21272/fmir.7(1).145-153.202

[14] Kun, E. (2024). Challenges in regulating cloud service providers in EU financial regulation: From operational to systemic risks, and examining challenges of the new oversight regime for critical cloud service providers under the Digital Operational Resilience Act. *Computer Law & Security Review*, *52*, 105931. https://doi.org/10.1016/j.clsr.2023.105931

[15] Mishchenko, S., Naumenkova, S., Mishchenko, V., & Dorofeiev, D. (2021). Innovation risk management in financial institutions. *Investment Management and Financial Innovations*, *18*(1), 191-203. https://doi.org/10.21511/imfi.18(1).2021.16

[16] Narayanan, L., & Rajan, A. (2024). Artificial Intelligence for Sustainable Agriculture: Balancing Efficiency and Equity. *International Journal of SDG's Prospects and Breakthroughs, 2*(1), 4-6.
[17] Nguyen, P. S., Cuong, N. N., & Long, H. V. (2023). Cloud-Based Malware Detection Using Machine Learning Methods. In *Risk Detection and Cyber Security for the Success of Contemporary Computing* (pp. 171-197). IGI Global. https://doi.org/ 10.4018/978-1-6684-9317-5.ch009
[18] Sharma, N., Chakrabarti, A., & Balas, V. E. (2019). Data management, analytics and innovation. *Proceedings of ICDMAI*, *1*. https://doi.org/10.1007/978-981-32-9949-8
[19] Soy, A., & Balkrishna, S. M. (2024). Automated detection of aquatic animals using deep learning techniques. *International Journal of Aquatic Research and Environmental Studies*, *4*(S1), 1-6. https://doi.org/10.70102/IJARES/V4S1/1
[20] Stojanović, B., & Božić, J. (2022). Robust financial fraud alerting system based in the cloud environment. *Sensors*, *22*(23), 9461. https://doi.org/10.3390/s22239461
[21] Torres, J., & De Los Santos, S. (2018). Malicious PDF documents detection using machine learning techniques. In *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP 2018)* (pp. 337-344).
[22] Tripathy, D., Gohil, R., & Halabi, T. (2020, May). Detecting SQL injection attacks in cloud SaaS using machine learning. In *2020 IEEE 6th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing, (HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS)* (pp. 145-150). IEEE. https://doi.org/10.1109/BigDataSecurity-HPSC-IDS49724.2020.00035
[23] Udayakumar, R., Mahesh, B., Sathiyakala, R., Thandapani, K., Choubey, A., Khurramov, A., ... & Sravanthi, J. (2023, November). An integrated deep learning and edge computing framework for intelligent energy management in IoT-based smart cities. In *2023 International Conference for Technological Engineering and its Applications in Sustainable Development (ICTEASD)* (pp. 32-38). IEEE.

Authors Biography

Dr. Prerna Mahajan is a Professor in the Department of Computer Science and Information Technology at Jain (Deemed-to-be University), Bangalore, Karnataka, India. With extensive academic and research experience, her areas of expertise include artificial intelligence, data science, and advanced computing technologies. Dr. Mahajan is actively involved in mentoring students, leading research initiatives, and contributing to the advancement of computer science education.

Madhur Taneja is affiliated with the Centre of Research Impact and Outcome at Chitkara University, Rajpura, Punjab, India. His work primarily focuses on enhancing research quality, impact measurement, and supporting collaborative research initiatives. Madhur actively contributes to building a robust academic research environment through data-driven strategies and innovation.

Dr. Arun Kumar Marandi is an Associate Professor in the Department of Computer Science & IT at ARKA JAIN University, Jamshedpur, Jharkhand, India. He specializes in computer science education, with key interests in software engineering, data analytics, and technological innovation in education. Dr. Marandi is dedicated to advancing research and fostering technical excellence among students.

Sohel Das is an Assistant Professor in the Department of uGDX at ATLAS SkillTech University, Mumbai, Maharashtra, India. His academic and research interests lie in the areas of user experience design, digital transformation, and interdisciplinary learning. He is committed to mentoring students and contributing to innovative curriculum development in the field of design and technology.

Madhur Grover is affiliated with the Chitkara Centre for Research and Development at Chitkara University, Himachal Pradesh, India. His work focuses on supporting institutional research initiatives, improving research quality, and enhancing scholarly outcomes. Madhur is involved in research impact strategies, academic publishing support, and fostering interdisciplinary collaboration.

S. Aswath is an Assistant Professor in the Department of Electronics and Communication Engineering at Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India. His academic focus includes embedded systems, communication technologies, and VLSI design. He is actively involved in research and teaching, guiding students toward innovation in electronics and communication domains.