International Journal of Emerging Research in Engineering and Technology
Pearl Blue Research Group | Volume 7 Issue 1 PP 171-179, 2026
ISSN: 3050-922X | https://doi.org/10.63282/3050-922X.IJERET-V7I1P123

Original Article

Automated Regulatory (RegTech) Compliance Monitoring in the Financial Landscape Through Machine Learning Models

Dr. Prashant Kumar Srivastava
PhD CSE, SOCT,Associate Professor, Sanjeev Agrawal Global Educational (SAGE) University Bhopal.

Received On: 06/01/2025 ***Revised On: 08/02/2026*** ***Accepted On: 13/02/2026*** ***Published On: 17/02/2026***

Abstract *- Regulatory Technology (RegTech) has become an imperative solution in automated compliance monitoring in the financial sector, especially in detecting fraud in imbalanced data of transactions. This paper presents a machine learning-based Regulatory (RegTech) compliance monitoring system based on the Credit Card Fraud Detection (CCFD) dataset. Data integrity and consistency were ensured by conducting comprehensive data preprocessing solutions, such as missing value management, duplicate elimination, and categorical transformation, feature selection, and normalization using StandardScaler. Class imbalance was compensated using Synthetic Minority Over-sampling Technique (SMOTE). The data was split into training and testing subsets. Convolutional Neural Network (CNN) and Random Forest (RF) model have been built and tested on the basis of accuracy, precision, recall, and F 1 -score. The experimental findings prove suggested CNN to be better in comparison to RF, as it only shows 99.8% accuracy and 99.9% precision, recall, and F1-score. The results prove the efficiency of deep learning-based solutions in improving automated financial regulatory compliance monitoring systems.*

Keywords *- Regulatory Governance, Administrative Innovation, Digital Transformation, Regtech, Supervisory Technology, Financial Regulation.*

1. Introduction

The compliance with the regulatory frameworks is a cornerstone of the contemporary financial industry to maintain the integrity of the market, consumer protection, transparency, and financial stability[1]. The financial institutions must meet the ever-changing laws and regulations, as well as industry standards brought about by globalization, digital transformation, and new emerging financial risks[2][3]. Nevertheless, there has been the increased complexity of the regulation frameworks that has rendered compliance management more difficult[4][5]. Manual compliance processes are inflexible, consumptive of resources and subject to human error and therefore unsuitable to the dynamic and data-intensive nature of the modern financial ecosystem.

Regulatory Technology (RegTech) is the promising solution to these issues, as it can improve the process of monitoring compliance and risk management through the application of new digital technologies[6][7]. RegTech is a combination of artificial intelligence (AI), machine learning (ML), data analytics, cloud computing, robotic process automation (RPA), and blockchain to automate regulatory processes[8][9]. Through these technologies, real-time monitoring of the transactions can be done, regulatory reporting can be performed automatically, and anomalies can be identified, and the security of the data can be enhanced[10][11]. RegTech improves transparency and accuracy in financial institutions, as well as proactive mitigation of risks, by alleviating the operational burden and compliance expenses.

Machine Learning (ML) models are at the heart of the automated regulatory compliance monitoring offering adaptive, intelligent, and scalable analytical solutions[12]. ML algorithms analyze the past in order to make predictions and classify transactions to detect suspicious activities as well as compliance risks and predict them with significant accuracy[13]. Higher-order methods of learning, such as deep learning models, can bring much higher accuracy in detection by discovering non-linear relationships in large financial data[14][15]. These features are specifically important in fraud detection, anti-money laundering (AML) monitoring, credit risk analysis, and anomalies detection in real-time[16]. Artificial Intelligence (AI) has thus emerged as a revolution within the financial industry and has increased the efficiency of operations and process planning by using machine learning, deep learning, natural language processing, and predictive analytics[17]. With the growing digitalization and data-based nature of financial systems, AI-based RegTech is an important driver of resilience, innovation, and regulatory compliance[18][19]. The machine learning-based compliance systems can guarantee scalable, accurate, and automated regulatory activity within the current financial environment by constantly learning based on adapting data trends.

1.1. Motivation and Contribution

The motivated by the fact that digital financial transactions are rapidly growing and fraudsters are increasingly becoming more complex, posing a serious challenge to regulatory compliance and risk management. Old compliance monitoring systems tend to be manual, rule-based, and are usually not very good working with large scale, high-dimensional, and imbalanced transaction data.

The fact is that there is a high demand of smart, automated and scalable solutions capable of helping to detect suspicious activities in real-time with high accuracy and low false alarms. The study is hence inspired by the need to raise the efficiency of regulating bodies to efficient and effective data-driven methods that guarantee better financial security, transparency, and trust in the contemporary financial ecosystems. This study has a number of main contributions as can be listed as follows:

- Used a real financial transactions data set with extreme imbalance between fraud and legitimate records to provide practical significance and strength of the research.
- Cleaned the data, including missing values, deleted duplicates and noise to attain the quality of the dataset. Applied extensive preprocessing, feature selection, normalization and SMOTE-based data balancing of data to improve model reliability and performance.
- An Intelligent Regulatory (RegTech) compliance monitoring framework was developed with machine learning and deep learning algorithms to identify fraud.
- Introduced a superior machine learning and deep learning-based system incorporating the models of the Random Forest and Convolutional Neural Network to successfully monitor Regulatory (RegTech) compliance.
- Evaluated the suggested model based on extensive evaluation indices, such as accuracy, precision, recall, F1-score, and has a comprehensive and dependable analysis of performance.

1.2. justification and novelty

The justification of the proposed research is rooted in the fact that there exists the growing need of the intelligent and automated Regulatory (RegTech) based compliance monitoring systems that are able to effectively identify fraud activities in a highly imbalanced financial transactions setting. Traditional systems that are based on rules do not typically identify nonlinear and complex behavioral patterns and therefore restrict the detection capabilities. The originality in this work is that a complete, data-oriented structure is created, in which complex preprocessing, feature optimization, and the management of class imbalances are placed into one architecture. This systematic and organized method increases the accuracy of detection, resistance and generalization, which is a scalable and efficient system of contemporary financial regulatory compliance tracking.

1.3. Organization of the Paper

The remaining part of this paper will be structured in the following way: Section II is a review of Regulatory (RegTech) compliance monitoring and fraud detection methods. Section III is a description of the dataset, preprocessing, and application of the suggested models, and Section IV is an experimental result and a comparative analysis. Lastly, Section V presents a conclusion of the study in terms of important findings and future research directions.

2. Literature Review

To develop this study, a detailed study and analysis of major research reports on Regulatory (RegTech) Compliance Monitoring were reviewed and analyzed to inform and enhance this research study.

Malali and Madugula, (2025) study applies RandomForest (RF), Gradient Boosting (GB), and AdaBoost (AB) to predict compliance risks using the bank marketing dataset. SMOTE and cost-sensitive learning addressed class imbalance, while Grid Search Cross-Validation optimized model performance. Accuracy, precision, recall, and F1-score were utilized for evaluation. RF achieved the best performance with 93.98% accuracy, 92.03% precision, 96.42% recall, and a 94.17% F1-score, followed by GB and AB with slightly lower results. Traditional models like K-Nearest Neighbors (KNN) and Extreme Gradient Boosting (XGB) performed less effectively. Conclusion[20].

Modalavalasa, (2025) evaluates all aspects of RegTech solutions designed for financial industry regulatory compliance and discusses the implementation of AI and machine learning approaches in detail. Accounting for its special implementation an Artificial Neural Network (ANN) model detects fraudulent credit card activity within European dataset. The model shows outstanding results through its 97.86% accuracy combined with 99.9% precision and recall and F1-score and 1.00 AUC from the ROC curve verification[21].

Al Jameel *et al.*, (2025) demonstrate that ensemble models, particularly XGBoost, achieve a remarkable accuracy of 88.7%, along with superior precision (86.5%), recall (84.5%), and AUC-ROC (0.90), highlighting their robustness in financial distress prediction. Feature importance analysis reveals that key financial indicators, including the Debt-to-Equity Ratio (25% importance) and Return on Assets (ROA) (20 % importance), play a critical role in influencing bankruptcy risk[22].

Bhasin *et al.*, (2024) proposed Quantum Support Vector Machine (QSVM) demonstrates unparalleled success, with a remarkable Portfolio performance over time of 89.65%. This result significantly surpasses existing quantum algorithms, including Quantum Principal Component Analysis (QPCA), Quantum Boltzmann Machines (QBM), and Quantum K-Means Clustering (QKC), by an impressive margin of 25.15%[23].

Agrawal *et al.*, (2024) Algorithms like K-Nearest Neighbors, Random Forest, Support Vector Machines, and Logistic Regression, in this paper prove effective in classifying decision-making problems within the finance sector with great recall, precision and F1 Score above 0.8 in majority cases. They have well utilized stacking classifier ensemble technique for better results. A comprehensive approach to AI integration encompasses all aspects of the business, including product development, risk management, compliance, and customer service[24].

Table 1 summarizes the recent studies on the Regulatory (RegTech) Compliance Monitoring and explains the proposed models, datasets used, key findings, and challenges.

Table 1: Recent Studies On Regulatory (Regtech) Compliance Monitoring Using Machine Learning Technqiues

Author	Proposed Work	Results	Key Findings	Limitations & Future Work
Malali & Madugula (2025)	Addressed class imbalance using resampling techniques and applied Grid Search Cross-Validation for model optimization across ML classifiers.	RF achieved 93.98% accuracy, 92.03% precision, 96.42% recall, and 94.17% F1-score; GB and AB slightly lower; KNN and XGB less effective.	Random Forest performed best among tested models; class balancing significantly improved fraud detection performance.	Limited comparison with deep learning models; future work can explore hybrid and real-time fraud detection systems.
Modalavalasa (2025)	Evaluated RegTech solutions using AI/ML; implemented ANN model for fraud detection on European credit card dataset.	97.86% accuracy, 99.9% precision, recall, F1-score, and AUC = 1.00.	ANN demonstrated outstanding fraud detection capability with near-perfect classification metrics.	Needs validation on diverse and large-scale real-time datasets; scalability assessment required.
Al Jameel et al. (2025)	Used ensemble models (especially XGBoost) for financial distress prediction with feature importance analysis.	Accuracy 88.7%, Precision 86.5%, Recall 84.5%, AUC-ROC 0.90.	Debt-to-Equity Ratio (25%) and ROA (20%) were identified as key bankruptcy risk indicators.	Focused mainly on financial distress prediction; future work may extend to multi-sector financial datasets.
Bhasin et al. (2024)	Proposed Quantum Support Vector Machine (QSVM) for financial portfolio optimization and prediction.	Achieved 89.65% portfolio performance, outperforming QPCA, QBM, and QKC by 25.15%.	QSVM significantly outperformed existing quantum algorithms in financial modelling tasks.	Practical implementation challenges due to quantum hardware limitations; future research on scalability and real-world deployment.
Agrawal et al. (2024)	Applied ML algorithms (KNN, RF, SVM, LR) with stacking ensemble for financial decision-making and compliance.	Most models achieved recall, precision, and F1-score above 0.8.	Stacking ensemble improved classification performance; AI integration is beneficial across compliance and risk management.	Limited dataset diversity mentioned; future work can include advanced deep learning and cross-domain validation.

Research gaps: Despite the high achievements of AI-based fraud detection and RegTech compliance systems, some research gaps can still be identified. Majority of the studies lay more emphasis on accuracy and little on model interpretability, scalability and real-time implementation in dynamic financial settings. Also, most models are validated using a single or small dataset and this limits their applicability in different financial environments. The following directions should be sought in future research: explainable AI, cross-dataset validation, computational efficiency, and strong and efficient real-time compliance monitoring systems.

3. Research Methodology

The proposed methodology from the CCFD Dataset that has 284,807 records with high levels of class imbalance in order to establish a Regulatory (RegTech) compliance monitoring system. First, the steps of preprocessing the data, involving the treatment of missing values, elimination of duplicates, categorical data conversion, and normalization with StandardScaler, were undertaken to ascertain data reliability and uniformity. The most relevant attributes were selected by applying feature selection techniques, then SMOTE was used to balance the fraud classes imbalance. The data was further divided into training and testing sets using an 80:20 stratified split to maintain class distribution. Lastly, the CNN and Random Forest models, which were proposed, were tested and trained based on accuracy, precision, recall, F1-score, analysis to evaluate the performance of the models on detecting fraud. The flowchart of Regulatory (RegTech) Compliance Monitoring using machine learning is proposed in Fig. 1.

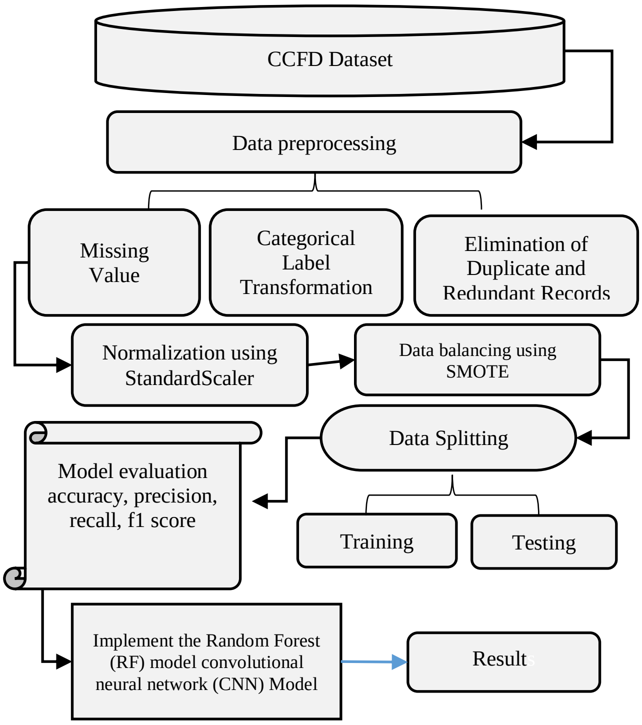
CCFD Dataset
Data preprocessing
Missing Value
Categorical Label Transformation
Elimination of Duplicate and Redundant Records
Normalization using StandardScaler
Data balancing using SMOTE
Data Splitting
Model evaluation accuracy, precision, recall, f1 score
Training
Testing
Implement the Random Forest (RF) model convolutional neural network (CNN) Model
Result

Fig 1 : Proposed Flowchart For Regulatory (Regtech) Compliance Monitoring Using Machine Learning

The next section provides an in-depth account of every step that is a part of the proposed methodology:

3.1. Data Gathering and Analysis

In dataset, Credit card fraud (CCFD) Dataset is used, and the one used to be downloaded on kaggle.com. The dataset consists of 284,807 records, and 492 fraudulent and 284,315 legitimate records out of the total records were found to be fraudulent, only 0.172% of total transactions were identified as fraudulent. The analysis of the distribution of the attacks, the correlations between features, etc. with the data in the form of different data visualizations (bar plots, heatmaps, etc.) are provided below:

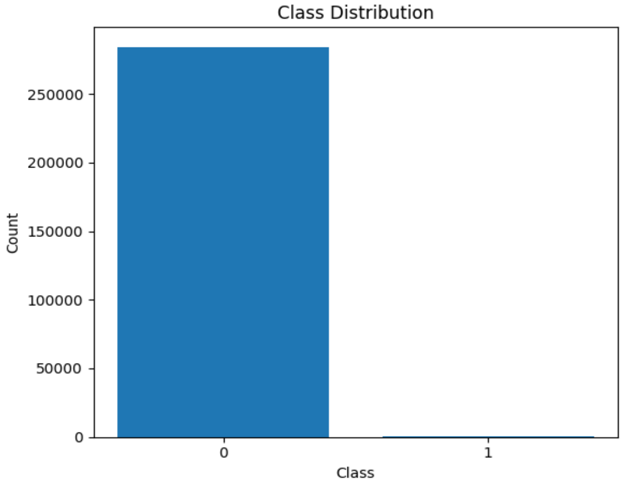

Fig 2: Bar Graph Of Class Distribution Of CCFD Dataset

Figure 2 demonstrates the distribution of classes of the CCFD Dataset with a high number of legitimate and fraudulent transactions. Class 0 (legitimate) prevails in the dataset consisting of approximately 284,000 transactions whereas Class 1 (fraud) consists of 492. Such a drastic imbalance renders it difficult to detect fraud and necessitates data balancing practices to bring the model to work.

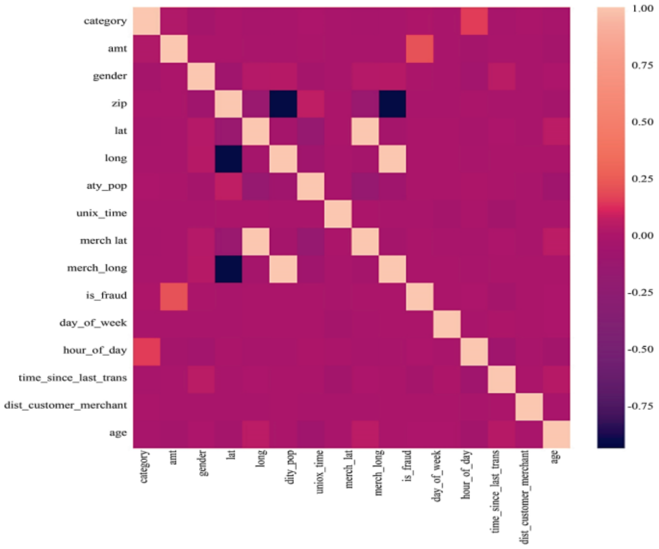

Fig 3: Correlation Matrix Heatmap On CCFD Dataset

Figure 3 presents the correlation matrix heatmap of the CCFD dataset used for Regulatory (RegTech) compliance monitoring. The heatmap shows the correlation between transaction characteristics with the light colors depicting strong positive links and darker colors showing negative links. The analysis assists in establishing the feature dependencies that are important and facilitate the effective selection of features and the level of performance of the fraud detection models.

3.2. Data Pre-processing

Data preparation, such as concatenation, cleansing, feature engineering, and normalization, was done by using the Transactions Dataset. Preprocessing procedures included missing values, removal of duplicate and redundant records, changing the categorical labels in order to convert the non-numeric data to numerical format and normalization of features to normalize the data to a consistent range. Such steps guaranteed better data regularity, less bias, decreased impact of outliers, and better model performance. These are the key preprocessing processes to note:

Missing Value: Missing value deals with identifying and processing incomplete or null values in a dataset. This will make the data consistent and enhance the reliability and accuracy of the machine learning model. Elimination of Duplicate and Redundant Records: There were duplicate and redundant records that were detected and eliminated to avoid duplication and bias during the analysis. This measure helps to increase the data integrity and overall model performance and reliability.

Categorical Label Transformation: Categorical label transformation transforms categorical variables into a numerical representation to run the machine learning algorithms successfully. This is required to make sure non-numeric data is modelled in a well-organised and valuable format to train the model.

3.3. Normalization using Standard Scaler

Since the data sets are of different magnitude, the data were standardized by the StandardScaler () procedure to reshape the data to the extent that the average of the resulting distribution is set to zero and the standard deviation equals one. The transformation is done through calculating the difference between the average of all the observations and dividing them by the standard deviation, like in Equation (1):

$$z = \frac{x - \mu}{\sigma} \ldots \ldots . . (1)$$

where z is the transformed value of the feature, x is the original value of each descriptor, μ is the mean, and σ is the standard deviation of the feature in the dataset.

3.4. Feature Selection

The process of feature selection is crucial where the required items are selected out of an available dataset according to the knowledge. The process of choosing and prioritizing the most useful features in a dataset that will enhance the performance of the model is known as feature selection. It can be used to reduce dimensionality, remove redundant data, compute more efficiently, and increase prediction quality.

3.5. Data balancing using SMOTE

Data balancing is done to solve the issue of imbalance in the dataset in terms of classes by balancing the minority and majority classes. This is done to make models fair and increase the accuracy and reliability of classification. To control the problem of class imbalance, data balancing with SMOTE (Synthetic Minority Over-sampling Technique) is used, whereby artificial samples of the small group are produced. The method enhances the performance of models by establishing a better class balance and less bias on the majority class.

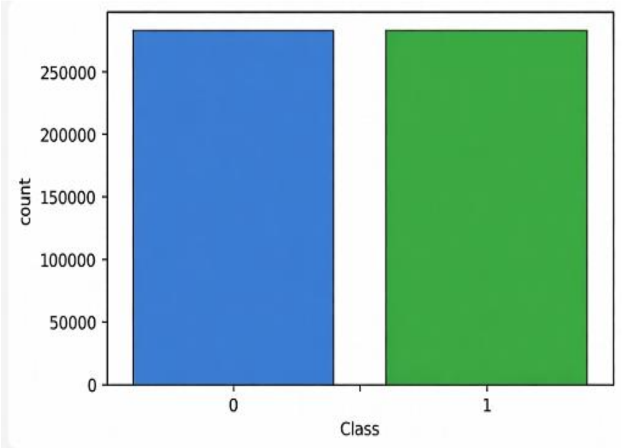

Fig 4: Bar Graph Of Class Distribution After SMOTE Of CCFD Dataset

Fig. 4 presents how Synthetic Minority Over-sampling Technique (SMOTE) was used to achieve balance in the dataset. Known to be able to produce artificial cases of fraudulent transactions, SMOTE is used to balance the ratio of classes by producing synthetic transactions by this selected approach.

3.6. Data Splitting

The data was stratified and separated into training and testing set (80:20). This strategy made sure that both subsets retained the same proportion of classes as the initial dataset, thus they had a proportional representation in both sets.

3.7. Proposed Convolutional Neural Network (CNN) Model

A Convolutional Neural Network (CNN) is a deep learning architecture widely used for feature extraction and classification tasks[25]. It automatically learns hierarchical feature representations from input data through convolutional operations, making it highly effective for pattern recognition in structured datasets such as images and network traffic data. The fundamental operation in a CNN is the convolution process, where a filter (kernel) slides over the input to extract important features. The convolution operation is mathematically expressed (2) as:

$$Z_{i,j}^{(l)} = (X * W^{(l)})_{i,j} + b^{(l)} \ldots \ldots \ldots . (2)$$

Where Xrepresents the input, $W^{(l)}$denotes the filter weights at layer l, $b^{(l)}$is the bias term, and $Z_{i,j}^{(l)}$is the output feature map. After convolution, a non-linear activation function such as ReLU is applied to introduce non-linearity into the model. The ReLU activation function is defined in eq. (3) as:

$$f(x) = \max(0, x) \ldots \ldots . . (3)$$

This function allows the network to learn complex patterns by transforming negative values to zero while retaining positive values. To reduce dimensionality and computational complexity, a pooling layer is applied, commonly using max pooling. The max pooling operation can be represented in (4) as:

$$P_{i,j} = \underset{(m,n)\in R}{max} Z_{m,n} \dots\dots\dots..(4)$$

Where R denotes the pooling region. Pooling helps in reducing overfitting and preserving dominant features. Thus, CNN combines convolution, activation, pooling, and fully connected layers to automatically learn discriminative features and perform accurate classification.

3.8. Proposed Random Forest Model

This work proposes a supervised machine learning–based model, Random Forest, for Air Cooling Control in Thermal HVAC Systems. Random Forest is an ensemble machine learning algorithm that builds and combines multiple decision trees to improve classification or regression performance. Each tree is trained on a random subset of the training data (bagging) and uses a random subset of features at each split, which increases diversity among the trees. The final prediction is made by majority voting (for classification) or averaging (for regression) across all trees. The data is recursively split into partitions. At a particular node, the split is done by asking a question on an attribute. The choice for the splitting criterion is based on some impurity measures such as Shannon Entropy or Gini impurity. Gini impurity is used as the function to measure the quality of split in each node. Gini impurity at node N is given by in eq (5):

$$g(N) = \sum_{i\neq j} P(w_i)P(w_j) \dots\dots\dots(5)$$

Where P(w_i) is the proportion of the population with class label i. Another function which can be used to judge the quality of split is Shannon Entropy. It measures the disorder in the information content. In Decision trees, Shannon entropy is used to measure the unpredictability in the information contained in a particular node of a tree (In this context, it measures how mixed the population in a node is). The entropy in a node N can be calculated in eq. (6) as follows:

$$H(N) = \sum_{i=1}^{i=d} P(w_i)log_2(Pw_i)) \dots\dots\dots.(6)$$

Where d is the number of classes considered and P(w_i) is the proportion of the population labeled as i. Entropy is the highest when all the classes are contained in equal proportion in the node. It is the lowest when there is only one class present in a node (when the node is pure). The obvious heuristic approach to choose the best splitting decision at a node is the one that reduces the impurity as much as possible. In other words, the best split is characterized by the highest gain in information or the highest reduction in impurity. The information gain due to a split can be calculated (7) as follows:

$$\Delta\mathrm{I(N)} = \mathrm{I(N)} - P_L * \mathrm{I}(N_L) - P_R * \mathrm{I}(N_L) \dots\dots.(7)$$

Where I(N) is the impurity measure (Gini or Shannon Entropy) of node N, P_L is the proportion of the population in node N that goes to the left child of N after the split and similarly, P_R is the proportion of the population in node N that goes to the right child after the split. N_L and N_R are the left and right child of N, respectively.

3.9. Evaluation metrics

To assess the performance of the proposed model in terms of classifications, such essential metrics as validation accuracy, precision, recall, and F1-score were chosen because they are suitable in the case of balanced datasets. After the training and testing of machine learning models with the help of baseline algorithms, the performance of the models was evaluated using standard evaluation measures. True Positive (TP) refers to the model in predicting and classifying malicious activities correctly, and True Negative (TN) refers to the model in predicting and classifying legitimate, non-malicious activities correctly. These fundamental elements are the foundation of obtaining the overall performance measures which are practically described in Equations (8) to (11).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+FP+FN+TN}} \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \quad (9)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \quad (10)$$

$$\text{F1} - \text{score} = 2 * \frac{(\text{precision}*\text{recall}}{(\text{precision+recall})} \quad (11)$$

Accuracy is used to measure the percentage of accurately classified instances in the total number of samples and is best used when the samples are balanced. Precision is the percentage of accurate positive predictions among all the positive predictions, whereas Recall is the capability of the model to distinguish the actual positive cases. The Precision and Recall values have a harmonic mean that is called the F1-score, which gives a fair assessment of the overall effectiveness of the model.

4. Results and Discussion

The experiments have been conducted on a high-performance computer system which has Intel Core i9-13900K-processor (3.0 GHz), 64 GB of DDR5 RAM, and NVIDIA RTX 4090-GPU (24 GB VRAM) running on Windows 11 Pro. All the implementation and analysis were done using Python libraries, which are Pandas, NumPy, Matplotlib, Seaborn, and Scikit-learn. In the case of Regulatory (RegTech) Compliance Monitoring in financial environment. As Table II demonstrates, the suggested CNN model is slightly better with 99.8% accuracy and 99.9% precision, recall, and F1-score, in contrast to 99% in each of the metrics with RF. The result shows that the CNN model slightly outperformed the RF model, which points to the high capability of deep learning methods in simulating complex transaction patterns to ensure successful regulatory compliance monitoring.

Table 2: Classification Results of Proposed ML and DL Models On CCFD Dataset

Matrix	Proposed CNN	Proposed RF
Accuracy	99.8	99
Precision	99.9	99
Recall	99.9	99
F1-score	99.9	99

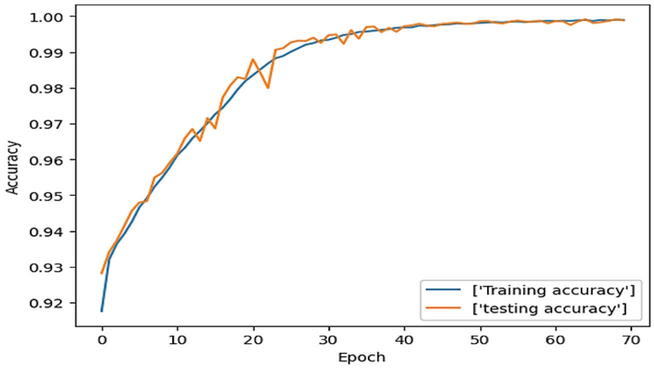

Fig 5: Accuracy Curve for the Proposed CNN Model

Figure 6 shows how the values of training and testing loss changes with 70 epochs. The similarity between the two curves is that starting with a consistent downward trend, which shows that learning and the model convergence are taking place effectively within the training process. The training loss initially begins with a high value and decreases rapidly in the early epochs and thereafter gradually levels off with increase in the number of epochs. Likewise, the testing loss is almost at the same level as the training loss with a small divergence, which indicates that there is a good generalization performance, and there is no substantial overfitting. The intersecting nature of the two curves with the close values of loss showing near zero loss values reflects the strength and stability of the model optimization process.

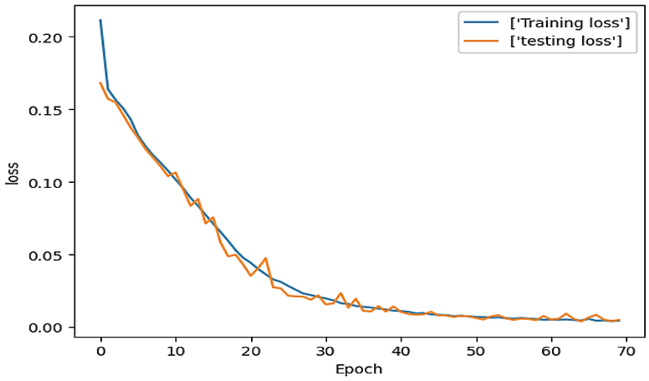

Fig 6: Loss Curve for the Proposed CNN Model

Figure 5 shows the training and testing accuracy as a function of 70 epochs. The two curves show a smooth increase at the end, which means that the model is learning, and the performance of the model in terms of classification is enhanced. There is rapid improvement in the training accuracy in the early epochs with gradual progression towards almost perfect accuracy. The accuracy of the testing is very precise to the training curve with a few deviations indicating a high level of generalization and low-level overfitting. The fact that both curves converge at an accuracy value near to 1.0, ascertains the strength, stability and good predictive capabilities of the obtained model.

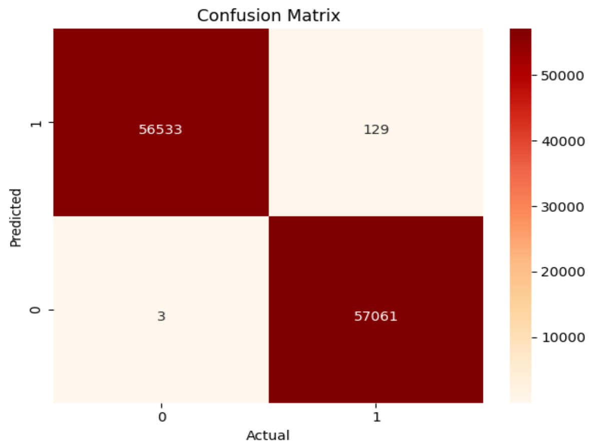

Fig 7: Confusion Matrix for the Proposed CNN Model

Figure 7 illustrates The confusion matrix shows how the binary classification model performs in the classification of instances of class 0 and class 1. It was able to correctly classify 56, 534 instances of class 1 and 57, 061 class 0, which stated that there were high numbers of true positive and true negative predictions. A few misclassifications were noted and this was 129 false positive and 3 false negative. The findings indicate a high predictive power with low classification error indicating the effectiveness and strength of the proposed model in binary classification assignments.

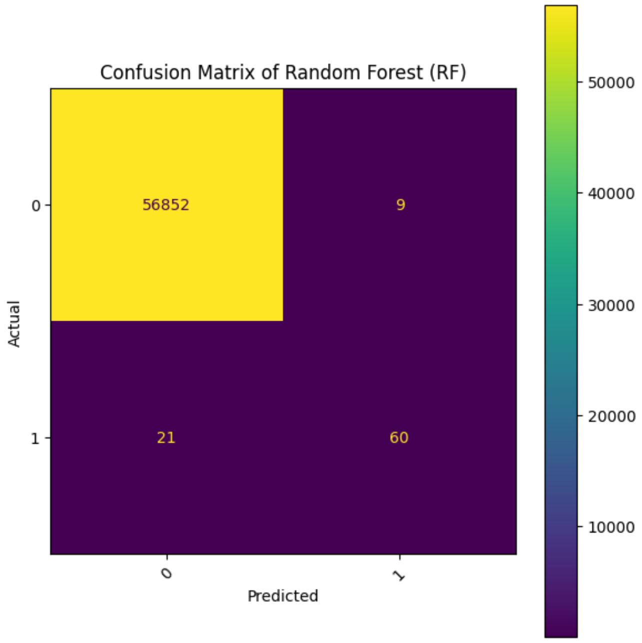

Fig 8: Confusion Matrix for the Proposed RF Model

Figure 8 presents the confusion matrix shows the classification rate of the Random Forest (RF) model in binary classification. The model had correctly identified 56,852 cases of class 0 (true negatives) and 60 cases of class

1 (true positives). There were 9 false positives and 21 false negatives. The fact that the correct number of samples was much higher than that of misclassifications shows that there is high prediction power and class discrimination ability. The low false positive and false negativity rates also prove the strength and high reliability of the RF model in imbalanced classification cases.

4.1. Comparative analysis

To assess the effectiveness of the proposed machine learning and deep learning models, a comparative accuracy test against existing models is presented in Table III. Table III shows a comparative perspective of various Machine Learning and Deep Learning models of Regulatory (RegTech) Compliance Monitoring. Decision Tree (DT) model had a precision of 95.05%, recall of 93.01%, F1-score of 94.02% and an accuracy of 94.02%, which represented moderate performance. The MLP model enhanced the scores to 95.9 % accuracy, 98.5 % precision, 93.2 % recall and 95.7 % F1-score with high precision but a weak recall. The KNN model additionally achieved a higher performance of 96.09% accuracy, 94.7% precision, 97.5 % recall and 96.1% F1-score which indicates balanced recognition. Nevertheless, the Proposed Random Forest (RF) model performed much better than these conventional models with 99% on all measures of evaluation and the Proposed CNN had the best results with 99.8 % accuracy and 99.9 % precision, recalls, and F1-score. Comprehensively, the findings indicate that the suggested deep learning CNN reduction model provides excellent predictive accuracy in effective monitoring of regulatory compliance than the traditional machine learning models.

Table 3: Comparison of Different Ml and Dl Models for Regulatory (Regtech) Compliance Monitoring.

Model	Accuracy	Precision	Recall	F1-score
DT[26]	94.02	95.05	93.01	94.02
MLP[27]	95.9	98.5	93.2	95.7
KNN[28]	96.09	94.7	97.5	96.1
Proposed RF	99	99	99	99
Proposed CNN	99.8	99.9	99.9	99.9

The presented model combining random forest (RF) and convolutional neural network (CNN) is shown to be more efficient in Regulatory (RegTech) compliance monitoring as the model has high accuracy of 99% and 99.8 with the Kaggle transaction dataset. Hybrid machine learning and deep learning increases the ability to detect fraud, generalization, as well as decreases the false positives when monitoring financial transactions. This would enhance automated compliance systems because it will allow quicker, more dependable, and scalable detection of suspicious operations in the monetary environment.

5. Conclusion and future study

Regulatory (RegTech) compliance surveillance system based on machine learning and deep learning algorithms on Credit Card Fraud Detection (CCFD) dataset which is highly imbalanced. Data consistency and better reliability of the models were provided by the use of extensive preprocessing such as data cleaning, normalization, feature selection, and SMOTE-related class balancing. The accuracy, precision, recall, and F1-score were used to measure the proposed Convolutional Neural Network (CNN) and Random Forest (RF) models. The experimental findings proved that CNN model performed better with 99.8% percent accuracy, 99.9% percent precision, recall, and F1-score compared to the RF model. The confusion matrices and learning curves proved that the ability to generalize is quite high with a small overfitting. The results indicate that deep learning architectures are very useful in modeling complex transaction patterns and identifying fraud activities, which enhances automated regulation compliance system in financial settings by precise, scalable and dependable fraud detection systems.

Future studies can consider hybrid ensemble designs that combine CNN and LSTM or transformer-based models, to be better at analyzing temporal patterns. Scalability, robustness, and regulatory interpretability can also be improved by real-time deployment over continuously changing financial data, explainable AI integration to improve regulatory disclosure, and cross-domain financial datasets evaluation.

References

[1] S. Agarwal, V. Kandoria, Y. Kankriya, A. Kuckian, and V. Wadhe, “Innovations in Financial Intelligence Applications using Artificial Intelligence,” in *5th IEEE International Conference on Advances in Science and Technology, ICAST 2022*, 2022. doi: 10.1109/ICAST55766.2022.10039547.

[2] V. Jain, A. Balakrishnan, D. Beeram, M. Najana, and P. Chintale, “Leveraging Artificial Intelligence for Enhancing Regulatory Compliance in the Financial Sector,” *Int. J. Comput. Trends Technol.*, vol. 72, no. 5, pp. 124–140, 2024, doi: 10.14445/22312803/ijctt-v72i5p116.

[3] S. B. Shah, “Advancing Financial Security with Scalable AI: Explainable Machine Learning Models for Transaction Fraud Detection,” in *2025 4th International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, IEEE, Apr. 2025, pp. 1–7. doi: 10.1109/ICDCECE65353.2025.11034838.

[4] S. Ahluwalia, R. V. Mahto, and M. Guerrero, “Blockchain technology and startup financing: A transaction cost economics perspective,” *Technol. Forecast. Soc. Change*, 2020, doi: 10.1016/j.techfore.2019.119854.

[5] V. Verma, “Deep Learning-Based Fraud Detection in Financial Transactions : A Case Study Using Real-Time Data Streams,” *ESP J. Eng. Technol. Adv.*, vol. 3, no. 4, pp. 149–157, 2023, doi: 10.56472/25832646/JETA-V3I8P117.

[6] Naga Krishna Mahesh Pulikonda, “Real-Time Regulatory Intelligence Framework: LLM-powered compliance automation for financial services,” *World J. Adv. Eng. Technol. Sci.*, 2025, doi:

10.30574/wjaets.2025.15.2.0784.

[7] S. B. Shah, "Evaluating the Effectiveness of Machine Learning in Forecasting Financial Market Trends: A Fintech Perspective," in *2025 3rd International Conference on Integrated Circuits and Communication Systems (ICICACS)*, IEEE, Feb. 2025, pp. 1–6. doi: 10.1109/ICICACS65178.2025.10968297.

[8] M. R. Peddamallu, "AI-Powered Compliance Automation in Financial Services," *Int. J. Sci. Technol.*, vol. 16, no. 4, Oct. 2025, doi: 10.71097/IJSAT.v16.i4.9557.

[9] D. W. Arner, J. Barberis, and R. P. Buckley, "FinTech, regTech, and the reconceptualization of financial regulation," *Northwest. J. Int. Law Bus.*, vol. 37, no. 3, pp. 373–415, 2017.

[10] D. W. Arner, J. N. Barberis, and R. P. Buckley, "FinTech and RegTech in a Nutshell, and the Future in a Sandbox," *SSRN Electron. J.*, 2018, doi: 10.2139/ssrn.3088303.

[11] A. Nerella, P. Badri, K. Sundravadivelu, and R. Murugesan, "Navigating Regulatory Hurdles in AI-Driven Credit Card Approvals: Balancing Innovation and Compliance," *J. Inf. Syst. Eng. Manag.*, vol. 8, no. 4, pp. 1–9, 2023.

[12] S. Thangavel, S. Srinivasan, S. B. V. Naga, and K. Narukulla, "Distributed Machine Learning for Big Data Analytics: Challenges, Architectures, and Optimizations," *Int. J. Artif. Intell. Data Sci. Mach. Learn.*, vol. 4, no. 3, pp. 18–30, 2023, doi: 10.63282/3050-9262.IJAIDSML-V4I3P103.

[13] A. N. Naveen Kolli, John Wesly Sajja, "Building Secure AI Agents for Autonomous Data Access in Compliance/Regulatory-Critical Environments," *Comput. Fraud Secur.*, vol. 2024, no. 9, pp. 363–373, Sep. 2024, doi: 10.52710/cfs.746.

[14] P. Liang, "Leveraging artificial intelligence in Regulatory Technology (RegTech) for financial compliance," *Appl. Comput. Eng.*, vol. 93, no. 1, pp. 166–171, Nov. 2024, doi: 10.54254/2755-2721/93/20240964.

[15] G. Modalavalasa, "Leveraging Machine Learning Techniques for Proactive Regulatory (RegTech) Compliance in Financial Landscape," in *2025 3rd World Conference on Communication &amp; Computing (WCONF)*, IEEE, Jul. 2025, pp. 1–6. doi: 10.1109/WCONF64849.2025.11233296.

[16] N. Bagherifam, S. Naghdi, V. Ahmadian, A. Fazlzadeh, and M. Baghalzadeh Shishehgarkhaneh, "Digital Regulatory Governance: The Role of RegTech and SupTech in Transforming Financial Oversight and Administrative Capacity," *Int. J. Financ. Stud.*, vol. 13, no. 4, p. 217, Nov. 2025, doi: 10.3390/ijfs13040217.

[17] R. P. Buckley, D. W. Arner, R. Veidt, and D. A. Zetzsche, "Building FinTech Ecosystems: Regulatory Sandboxes, Innovation Hubs and Beyond," *SSRN Electron. J.*, 2019, doi: 10.2139/ssrn.3455872.

[18] Hariharan Pappil Kothandapani, "Automating financial compliance with AI: A New Era in regulatory technology (RegTech)," *Int. J. Sci. Res. Arch.*, 2024, doi: 10.30574/ijsra.2024.11.1.0040.

[19] H. P. Kapadia, "Reducing Cognitive Load in Online Financial Transactions," *Int. J. Curr. Sci.*, vol. 12, no. 2, 2022.

[20] N. Malali and S. R. P. Madugula, "Predictive Analytics and Artificial Intelligence for Regulatory (RegTech) Compliance in the Financial Industry," in *4th IEEE International Conference on Distributed Computing and Electrical Circuits and Electronics, ICDCECE 2025*, 2025. doi: 10.1109/ICDCECE65353.2025.11035220.

[21] G. Modalavalasa, "Leveraging Machine Learning Techniques for Proactive Regulatory (RegTech) Compliance in Financial Landscape," 2025. doi: 10.1109/wconf64849.2025.11233296.

[22] M. Al Jameel *et al.*, "Early Detection of Bankruptcy Risk in Corporate Law Using Ensemble Machine Learning Techniques," 2025. doi: 10.1109/icbats66542.2025.11258208.

[23] N. K. Bhasin, S. Kadyan, K. Santosh, R. Hp, R. Changala, and B. K. Bala, "Enhancing Quantum Machine Learning Algorithms for Optimized Financial Portfolio Management," in *2024 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, INCOS 2024 - Proceedings*, 2024. doi: 10.1109/INCOS59338.2024.10527612.

[24] R. Agrawal, S. Desai, D. Dholwani, N. Kedari, and A. Banerjee, "Artificial Intelligence/Machine Learning Driven Decision making in Business Analytics for Financial Sector using Ensemble Machine Learning Techniques," in *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, 2024, pp. 371–376. doi: 10.1109/AIC61668.2024.10731028.

[25] Vibhor Pal and Satyadhar Kumar Chintagunta, "Transformer-Based Graph Neural Networks for RealTime Fraud Detection in Blockchain Networks," *Int. J. Adv. Res. Sci. Commun. Technol.*, vol. 3, no. 3, pp. 1401–1411, Jul. 2023, doi: 10.48175/IJARSCT-11978Y.

[26] O. Israel, A. Oluseye, S. Ojo, and O. T. Deborah, "Financial Fraud Detection using Machine Learning : Credit Card Fraud," *Int. J. Recent Eng. Sci.*, 2023, doi: 10.14445/23497157/ijres-v10i3p104.

[27] K. Hayat and B. Magnier, "Data Leakage and Deceptive Performance: A Critical Examination of Credit Card Fraud Detection Methodologies," *Mathematics*, 2025, doi: 10.3390/math13162563.

[28] A. M. Idrees, N. S. Elhusseny, and S. Ouf, "Credit Card Fraud Detection Model-based Machine Learning Algorithms," *IAENG Int. J. Comput. Sci.*, vol. 51, no. 10, pp. 1649–1662, 2024.