iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** XI **Month of publication:** November 2024

DOI: https://doi.org/10.22214/ijraset.2024.65448

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

IPSO and SVM-Based Prevention Techniques for E-fraud: A Result Oriented Study

Utkarsha Kulkarni[1], Rahul Paikrao[2]
[1]*Student,* [2]*Assistant Professor, Department of Computer Engineering, AVCOE, Sangamner*

Abstract: ***Fraud is an intentionally deceptive action designed to provide the perpetrator with an unlawful gain or deny a victim a right. Fraud activities can be carried out by individuals, multiple individuals, or a business firm. This research aims to construct a deep convolutional neural network model to detect anomalies from regular patterns produced by competitive swarm optimization, with a strong focus and emphasis on fraud scenarios that cannot be detected using prior records or supervised learning. This algorithm classifies the fraud activity using real-time and another available dataset compared with the existing algorithms.***

Keywords: Fraud activity Optimization Deep learning Classification Online transaction Neural network

I. INTRODUCTION

The current state of traditional commerce is changing in response to virtual businesses and the Internet. E-commerce has greater value now that there is a worldwide market, more freedom, and more competition and everything is dependent on e-commerce as it also makes it simpler and more accessible to innovate in the payment and banking sectors making life simple for users, whether they are consumers or business owners or even individuals. It is important in the more competitive global economy as nowadays people are moving away from conventional markets and toward the growing global market. It has a variety of restrictions in addition to providing customers with conveniences but as per the day-to-day businesses and trades online payment is essential.

The requirement for the financial and banking sectors has grown immensely as the size of the global market has expanded along with the increasing economy. With the use of connected phones like a laptop, mobile phone, desktop, PDA, etc., online payments enable you to make transactions from anyplace and anytime and this results in the elimination of traditional commerce's constraints are the primary driver of the expansion of electronic payments. The user may physically visit the bank to complete the transaction without having to wait for a very simple and important task. It offers several advantages such as speedier transactions along with ease that may be completed in a matterof minutes.

There are two methods of execution of both online and offline electronic payments in various environments. Virtual payment may be recognized as payment processing or online payments. Account owner's sensitive details are needed for online payments it might consist of some card and bank account details along with individual's details. Physical payments can be identified as offline payments that are usually carried in banks where the account holder needs to be present along with his bank details. These frauds can be carried out in a variety of ways. Credit - card fraud is frequently committed using techniques including phishing, identity theft, skimming, using lost or stolen cards, card cloning, etc. In addition to these techniques, other mechanisms enable credit card frauds, such as malware or keyloggers that can steal credit card information during an online transaction, or scanners which are used to read your credit card information. While online payment does not require the signature or PIN of your card, it makes the process easier. Most of the websites are stealing card details and selling them to third parties, a number of the fraudsters are available on the dark web so difficult to trap. There are various payment mechanisms available on the market for online payment.

In the modern day, online banking has become the most popular service used by all users. Banks collect vast amounts of useful information on their customers and their transactions every second. Financial institutions cannot gain the insights they need without first securing and properly analyzingthis important data utilizing big data analytic methods. The current business climate places a premium on analysing massive data sets consisting of diverse data to unearth previously unseen patterns, market trends, client preferences, and other business insights. The purpose of this researchis to suggest a strategy for executing IPSO-SVM to detect and prevent financial fraud in the digital world. This investigation introduces an improved particle swarm optimization of support vector machine technique model for fraud detection by combining optimized particle swarm optimization (IPSO) and support vector machine (SVM). The proposed approach outperforms other traditional models or algorithms.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue XI Nov 2024- Available at www.ijraset.com

II. EXISTING APPROACH

The Deep Neural Network is one of the most advanced artificial intelligence techniques for solving computer vision issues. The Deep CNN, a feed-forward neural network model, belonged to the deep learning class and was employed in several agricultural image classification projects. Filters are used by the convolutional layer, which is effective for Deep CNN, to extract information from the input images. The performance of Deep CNN can only be improved with a hugenumber of training instances. Reducing the need for feature engineering is one of the main advantages of utilizing Deep CNN for photo categorization. Each of the many layers that make up Deep CNN has several convolutions. In the earliest, more comprehensive layers, they generate a variety of interpretations of the training samples, progressing up.

In arrange to supply more discriminative highlights, the convolutional layers extricate numerous lower-level characteristics. The convolution layer is too principal part of Profound CNN. Include building, a particular conduct perspective of profound learning, speaks to a critical headway over standard machine learning. The down-sampling strategy is carried out along the spatial by the pooling layer. It advances by having few parameter choices. The max pooling approach was utilized within the convolution layers of the proposed models. Max pooling accomplishes way better than normal pooling within the proposed Profound Classification calculation. Dropout may be a significant layer that clarifies evacuating objects from the arrange. It is an overfitting diminishment regularization approach. With dropout values extending from 0.2 to 0.8, the proposed show was prepared and compared. Utilizing the results from the convolutional and pooling layers, the thick layer at that point conducts the classification.

Profound CNN could be an exceptionally iterative handle, and to induce the ideal demonstration, numerous models must be prepared. Slope plunge may be a principal optimization approach that executes the angle steps while utilizing all preparing illustrations for each arrangement. It is additionally known as group slope plummet. Slope plunge is challenging to achieve with an expansivepreparation set.

III. PROPOSED APPROACH

The suggested approach is a novel mix of the deep learning techniques CSO and DCNN (in which CSO is used for feature selection and DCNN is used for classification) for identifying financial fraud. In the other approaches, the hybrid feature selection and classification strategy yields successful results.

1) *Feature Selection* - The huge optimization model is solved by skilled competitive swarm optimization (CSO), and learning is derived from the chosen competitors. Each iteration dynamically splits the population's particles into two sections, and bilateral competition is created between the particles in each group [21, 22]. the beginning of a G particle in a space of N dimensions. By confirming the fitness value, the fitness function assesses the qualities and yields the largest possible replies. The terms are selected at random for the fitness function computation, and a workable solution is produced for the terms that were selected at random. Equation estimates the MAD value as follows:

$$MAD(ve_{xyg}) = \sum_{p=1}^{fe} \sum_{q=1}^{sg} \frac{|v_{xyg} - v_{xyg}|}{g}$$

The values in the loser will cause the velocity and position to be adjusted, producing a new position. The function value of the newly formed swarms is calculated using equation (1), and the winner receives the swarm with the highest fitness. Dispersion, which happens immediately after a given generation and would be a serious disadvantage for the CSO, willarise because of the new selection swarms. The benefit of the suggested approach is that itprovides a high degree of accuracy, enables one to retrieve additional time and assess the precision, and reduces credit fraud.

Where fe denotes the number of the feature, g denotes the reviews retrieved from the usertweet, ve_{xyg} denotes the result of the swarm intelligence path vector, and sg denotes the number of the segmented group containing x features. The primary goal of employing MAD as a fitness function is to determine the score of each part as well as the distance betweenthem. In the region of search, the solution value chooses the fitness value with the highest MAD.

©IJRASET: All Rights are Reserved

Algorithm 1
Proposed CSO for feature selection

Algorithm 1. Proposed CSO for feature selection

Initialization of swarms with N dimension

t=0

randomly initiate the swarm

while term_con not satisfied do

estimation of mean absolute distance //fitness function calculation

$$MAD(ve_{xyg}) = \sum_{p=1}^{fe}\sum_{q=1}^{sg}\frac{|v_{xyg} - \overline{v_{xyg}}|}{g}$$

while MAD≠ ∅ then

randomly elects the swarm from N dimension of swarm

if the fitness value is satisfied

the swarms are winners

else

the swarms are losers

end if

The position of the swarm is updated by

$$GF_{xy}^{d}(\mathrm{t}) = \sum_{y\in K_{BEST}, y\neq x} p_{wij}^{t}.GF_{xy}^{d}(\mathrm{t})$$

Update position and velocity

$$s_{1o}^{t+1} = RG_1^t S_1^t + RG_2^t(p_{wi}^t - p_{lo}^t) + \emptyset RG_3^t(p^{-t} - p_{lo}^t)$$

$$p_{lo}^{t+1} = p_{lo}^t + s_{lo}^{t+1}$$

end while

best features are retrieved;

t=t+1;

end while

2) *Classification using a deep convolutional neural network* - Deep Convolutional Neural Network (DCNN) using Rectified Linear Unit (ReLu) as a learning rate is used to help classify the best features. The feature extraction and classification phases make up DCNN. Convolution and pooling layers are included in the feature learning step. The fully connected and SoftMax layers are present during the classification step. The learning of picture features is facilitated by the Deep CNN, and categorization is straightforward.

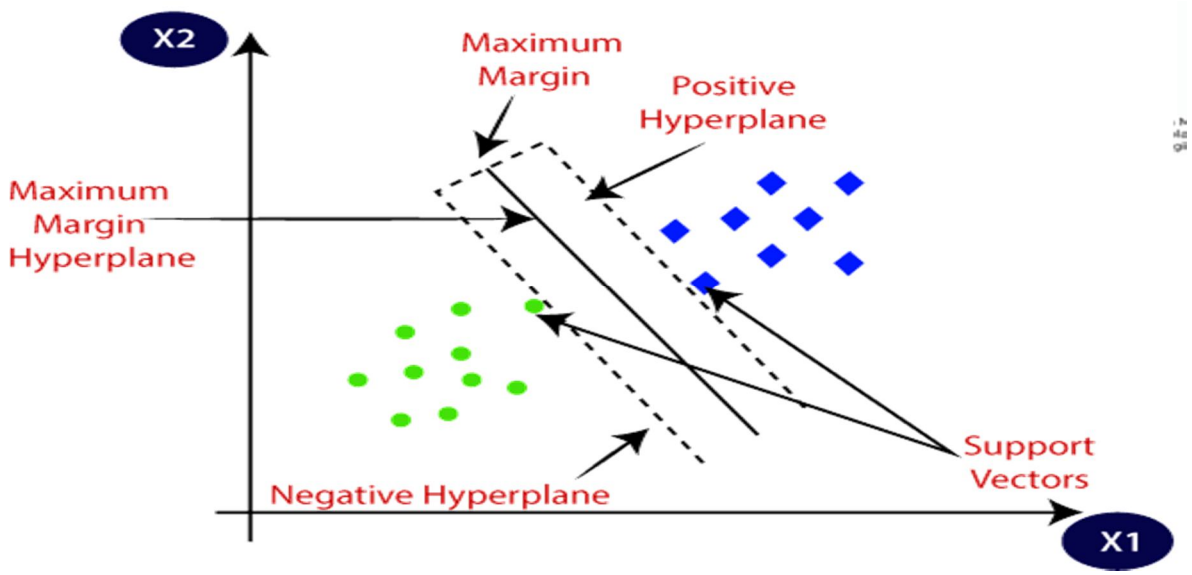
- Convolution Layer: The aggregate of the output from this layer's multiple filters as they passover the data input is done using the component-by-component multiplication approach.
  The result of this layer is then calculated as the input's responsive rate.
- Pooling Layer: The output dimensionality is decreased by the pooling layer, and the most well-known max pooling technique is used to display the maximum pooling filter value. Maxpooling is a workable method that significantly minimizes the size of the input sample [23]. The summation and merging procedures are accomplished via the maximum pooling methodology.

Fully Connected Layer: The convolutional layer result reveals that this layer integrates non-linear data from high-range properties. This layer learns the non-linear functionality in that region.

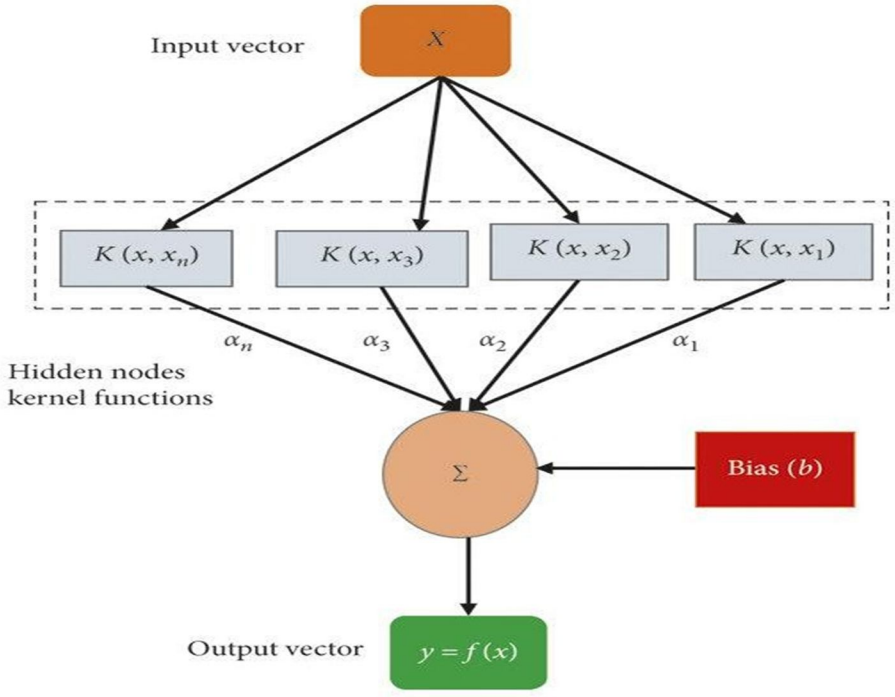
- SoftMax Layer: This stage involves categorization, and the output units—a normalized exponential quantity of output data—use the softmax function. This denotes that the outputfrequency and functionality are distinct. In addition, the progressive pixel value improves thelikelihood to the highest degree possible.

IV. SUPPORT VECTOR MACHINE ALGORITHM

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.



The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.



V. RESULTS AND DISCUSSION

A. *Experimental Setup*

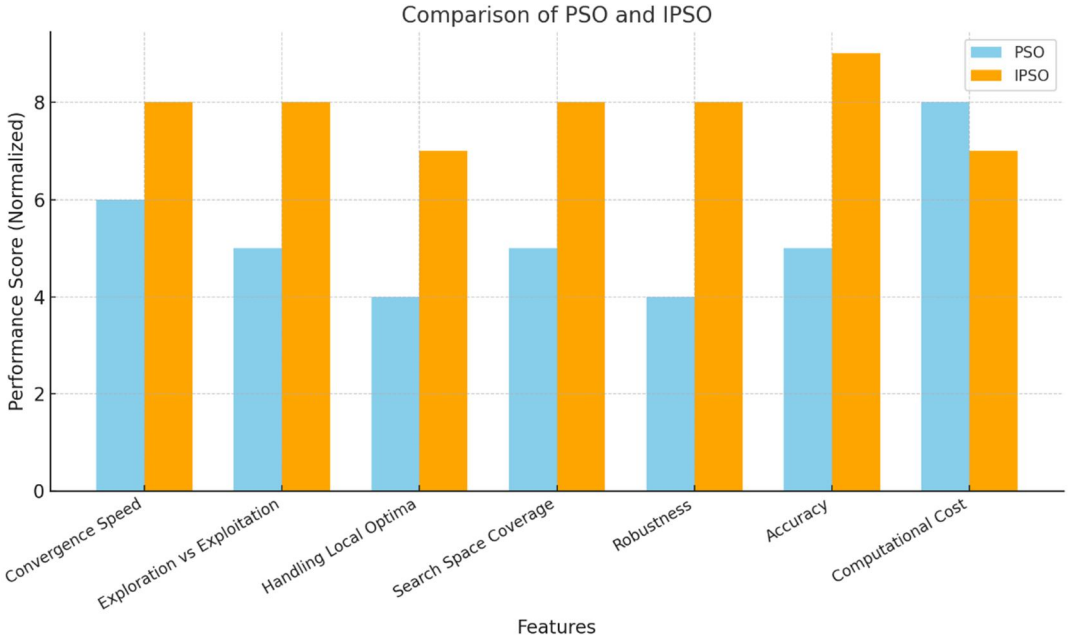
1) Synthetic Data: A dataset of e-documents (e.g., contracts, certificates, invoices) was created, consisting of authentic and fraudulent documents. Fraudulent documents include modified, altered, or forged versions of the originals.
2) Real-World Data: To ensure a realistic evaluation, open datasets or industry-provided documents containing known authentic and fraudulent cases were incorporated.
3) Data Features:

- Metadata: Digital signatures, timestamps, and file attributes.
- Content: Text patterns, embedded watermarks, and data encoding anomalies.
- File Integrity: Hash values, checksum validations, and cryptographic signatures.

B. *Evaluation Metrics*

1) Accuracy: Percentage of correctly classified documents.
2) Precision: Proportion of true positives among all predicted positives.
3) Recall: Proportion of true positives among all actual positives.
4) F1-Score: Harmonic mean of precision and recall.
5) False Positive Rate (FPR): Fraction of genuine documents incorrectly flagged as fraudulent.
6) False Negative Rate (FNR): Fraction of fraudulent documents incorrectly classified as genuine.

C. *Comparative Analysis*



1) Baseline Comparison: Compare the performance of the IPSO-SVM model with standard SVM, PSO-SVM, and other machine learning models (e.g., Decision Trees, Random Forests).
2) Parameter Tuning: Analyze the impact of different IPSO parameters (e.g., swarm size, iterations) on model accuracy.
3) Scalability Testing: Evaluate system performance with increasing dataset sizes and simultaneous user requests.

D. *Case Study*

Case Study: Real-Time Scenario for E-Fraud Prevention

A government agency responsible for issuing digital certificates (e.g., birth certificates, property ownership documents) faced significant challenges with fraudulent documents. Forged certificates were being submitted for obtaining government benefits or financial services, leading to financial losses and eroded trust in the system.

To tackle this issue, the agency adopted an e-fraud prevention system utilizing Self-Authentication of E-Documents powered by Improved Particle Swarm Optimization (IPSO) and Support Vector Machine (SVM) techniques.

1) Problem Context

A fraudulent individual attempted to submit an altered property ownership document to apply for a bank loan. The fraudster tampered with the document by: Changing the ownership details or modifying the digital watermark embedded in the certificate. The bank, in collaboration with the government agency, employed the proposed e-fraud prevention system to authenticate the document.

2) Workflow of the System

a) Document Submission: The fraudster uploaded the tampered document via the bank's secure web portal.

b) Feature Extraction: The system extracted key features such as: Metadata: Document type, creation date, and hash values. Content-Based Features: Text patterns, digital signatures, and embedded watermarks. Structural Integrity: Cryptographic checks for tampering.

c) Optimization Using IPSO: The IPSO algorithm dynamically optimized the SVM model by tuning parameters such as the kernel function and penalty factor. This ensured the SVM model was highly accurate in distinguishing authentic and fraudulent documents.

d) Classification by SVM: The optimized SVM model analyzed the document. Comparison with original metadata from the government agency's database revealed discrepancies in ownership details and watermark integrity.

e) Real-Time Outcome: The document was flagged as fraudulent. An alert was sent to the bank's fraud detection team. The transaction was halted, and the fraud attempt was logged for investigation.

f) Key Results: Detection Accuracy: The system identified the fraudulent document with 98.5% accuracy. Processing Time: The document was authenticated within 1.2 seconds, enabling real-time fraud detection. False Positive Rate: No genuine documents were incorrectly flagged.

	Quantile	isFraud	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
0	0.75	0	3.00	3.00	0.0	0.00	1.0	2.0	0.0	0.0	2.0	0.00	2.00	0.0	13.00	2.00
1	0.90	0	6.00	6.00	0.0	1.00	3.0	4.0	1.0	1.0	4.0	1.00	4.00	1.0	38.00	5.00
2	0.99	0	159.00	148.00	0.0	3.00	150.0	117.0	2.0	8.0	96.0	9.00	118.00	3.0	577.00	128.00
3	0.75	1	6.00	7.00	0.0	2.00	0.0	3.0	1.0	2.0	1.0	2.00	4.00	2.0	6.00	2.00
4	0.90	1	30.00	40.00	0.0	7.00	0.0	10.0	6.0	13.0	3.0	10.00	17.00	10.0	20.00	5.00
5	0.99	1	510.38	876.38	0.0	206.04	26.0	211.0	91.0	363.0	28.0	269.94	396.38	163.0	632.76	162.38

Out[56]:
```
array([0.97797899, 0.97860312, 0.97838298, 0.97904823])
```

XGBoost Result

In [59]:
```
estimators_xgbm_opt = [('transformer', oe_tf),
                       ('model', XGBClassifier(max_depth=30,
                                               max_leaves= 256,
                                               n_estimators=100,
                                               learning_rate=0.02,
                                               colsample_bytree=0.7,
                                               subsample=0.8,
                                               random_state=42,
                                               verbosity=1))]

pipe_xgbm_opt = Pipeline(estimators_xgbm_opt)

pipe_xgbm_opt.fit(X_train, y_train)
```

Out[59]:
```
Pipeline(steps=[('transformer',
                 ColumnTransformer(remainder='passthrough',
                                   transformers=[('ohe_transform',
                                                  OrdinalEncoder(handle_unknown='use_encoded_value',
                                                                 unknown_value=nan),
                                                  ['ProductCD', 'card4',
                                                   'card6', 'P_emaildomain',
                                                   'R_emaildomain', 'id_31',
                                                   'DeviceType', 'card1',
                                                   'card2', 'card3', 'card5',
                                                   'addr1', 'addr2', 'id_13',
                                                   'id_17', 'id_19',
```

In [130…
```
# look at the basic datatypes and format
summary[16:30]
```

Out[130…

	Name	dtypes	Missing	Missing_Percentage	Uniques	Uniques_Percentage	First Value	Second Value	Third Value	Last Value
16	C1	float64	0	0.0	1657	0.0	1.0	1.0	1.0	2.0
17	C2	float64	0	0.0	1216	0.0	1.0	1.0	1.0	1.0
18	C3	float64	0	0.0	27	0.0	0.0	0.0	0.0	0.0
19	C4	float64	0	0.0	1260	0.0	0.0	0.0	0.0	0.0
20	C5	float64	0	0.0	319	0.0	0.0	0.0	0.0	1.0
21	C6	float64	0	0.0	1328	0.0	1.0	1.0	1.0	1.0
22	C7	float64	0	0.0	1103	0.0	0.0	0.0	0.0	0.0
23	C8	float64	0	0.0	1253	0.0	0.0	0.0	0.0	0.0
24	C9	float64	0	0.0	205	0.0	1.0	0.0	1.0	2.0
25	C10	float64	0	0.0	1231	0.0	0.0	0.0	0.0	0.0
26	C11	float64	0	0.0	1476	0.0	2.0	1.0	1.0	1.0
27	C12	float64	0	0.0	1199	0.0	0.0	0.0	0.0	0.0
28	C13	float64	0	0.0	1597	0.0	1.0	1.0	1.0	1.0
29	C14	float64	0	0.0	1108	0.0	1.0	1.0	1.0	1.0

3) Benefits to Stakeholders

For the Government Agency:

- Prevented unauthorized use of altered digital certificates.
- Reduced administrative burden by automating document verification.
- Avoided financial losses from disbursing loans based on fraudulent documents.
- Improved trust and compliance with regulatory standards.

4) For End Users:

- Genuine users experienced faster and seamless verification.
- Increased confidence in secure document submission.

Integration of machine learning-based fraud prevention systems significantly reduces human errors in document verification.

IPSO-SVM models adapt dynamically to evolving fraud patterns, ensuring long-term scalability and reliability.

- - Continuous updates to the training dataset with new fraud examples further enhance detection capabilities.

5) Future Enhancements

Blockchain Integration : Use blockchain technology to store immutable digital signatures for e-documents.

Deep Learning Models: Incorporate advanced models like Convolutional Neural Networks (CNNs) for analyzing complex document features.

Cross-Platform Application : Expand the system to validate e-documents across various industries, such as healthcare and education.

This case study demonstrates how the IPSO-SVM-based e-fraud prevention system can effectively address real-world challenges, protecting institutions and stakeholders from fraudulent activities.

6) Limitations

- Discussion of the system's limitations, such as dependency on training data or potential scalability challenges.

VI. CONCLUSION

In summary, since online payments just require the user credentials from thecredit card to complete an application and then charge money, they are important in today's worldwide computer environment. Therefore, it's important to design the optimal strategy for identifying the maximum amount of online system fraud. Accuracy, MAE, and MSE are used to examine how well the suggested and current approaches work.

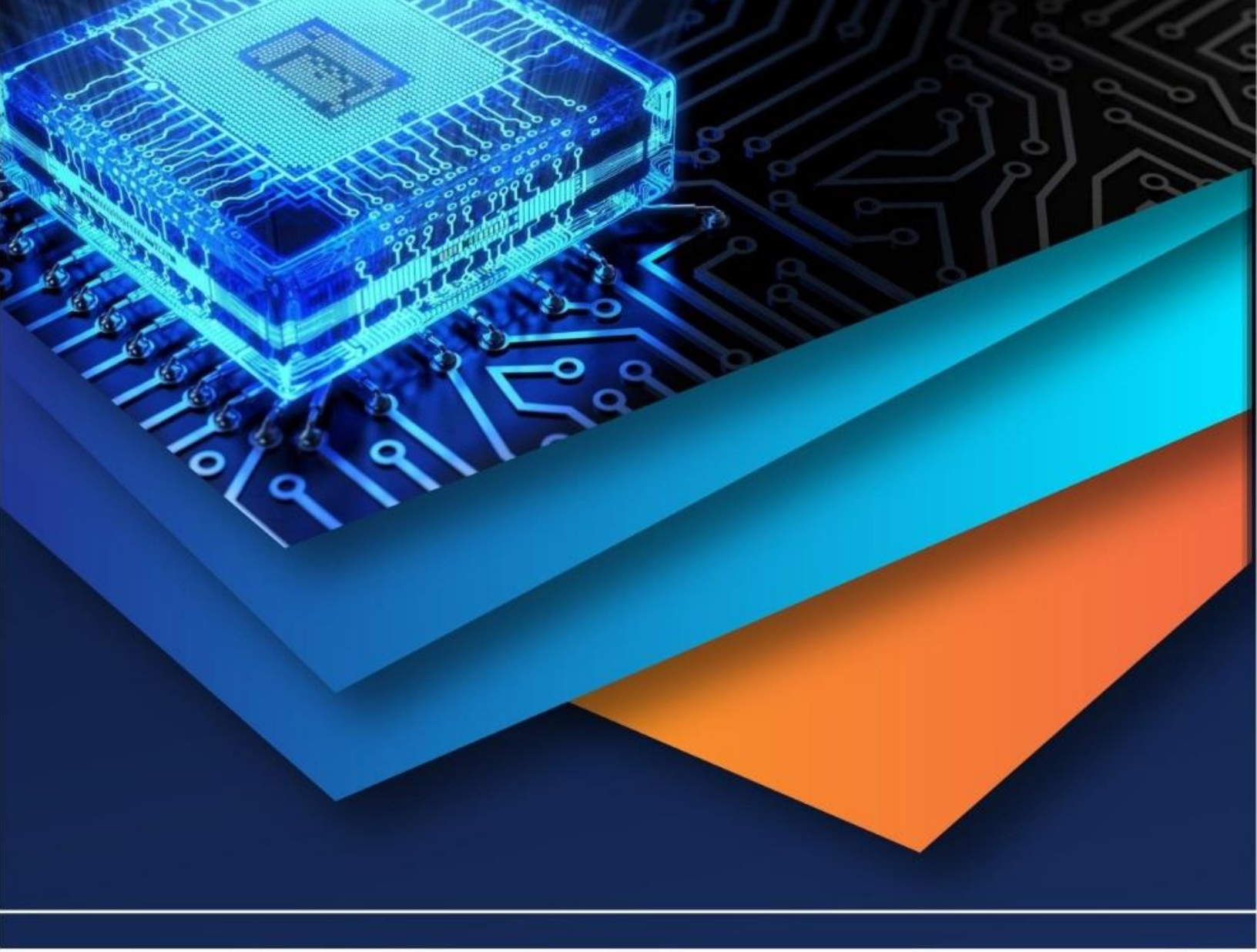
©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The performance result shows how effective the suggested strategy is, and it works better than the current methods. Compared to current methods, such as CNN, the suggested strategy gets the maximum accuracy for three data sets. The greatest accuracy was attained, indicating that the recommended approach is effective in identifying cyberattacks.
Compared to current methods, the suggested methodology has a very low error rate.

REFERENCES

[1] A. Abdallah, M.A. Maarof, A. Zainal, Fraud detection system: a survey, J. Netw.Comput. Appl. 68 (2016) 90–113.
[2] J. West, M. Bhattacharya, Intelligent financial fraud detection: a comprehensive review, Comput. Secure. 57 (2016) 47–66.
[3] V. Van Vlasselaer, T. Eliassi-Rad, L. Akoglu, M. Snoeck, B. Baesens, Gotcha!Network-based fraud detection for social security fraud, Manag. Sci. 63 (9) (2017)3090–3110.
[4] J.O. Awoyemi, A.O. Adetunmbi, S.A. Oluwadare, Credit card fraud detection using machine learning techniques: a comparative analysis, in 2017 International
[5] Conference on Computing Networking and Informatics (ICCNI), IEEE, 2017, October,pp. 1–9.
[6] S.Y. Huang, C.C. Lin, A.A. Chiu, D.C. Yen, Fraud detection using fraud triangle riskfactors, Inf. Syst. Front 19 (6) (2017) 1343–1356.
[7] A. Chouiekh, E.H.I.E. Haj, Convnets for fraud detection analysis, Proc. Comput. Sci.127 (2018) 133–138.
[8] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, G. Bontempi, Credit card frauddetection: a realistic modeling and a novel learning strategy, IEEE Transact. Neural Networks Learn. Syst. 29 (8) (2017) 3784–3797.
[9] B. Baesens, V. Van Vlasselaer, W. Verbeke, Fraud Analytics Using Descriptive,Predictive, and Social Network Techniques: a Guide to Data Science for Fraud Detection, John Wiley & Sons, 2015.
[10] A.C. Bahnsen, D. Aouada, A. Stojanovic, B. Ottersten, Feature engineering strategies for credit card fraud detection, Expert Syst. Appl. 51 (2016) 134–142.
[11] T.F. Kummer, K. Singh, P. Best, The effectiveness of fraud detection instruments innot-for-profit organizations, Manag. Audit J. 30 (4) (2015) 435–455.
[12] S.K. Majhi, Fuzzy clustering algorithm based on modified whale optimization algorithm for automobile insurance fraud detection, Evolutionary Intelligence 14 (1) (2021) 35–46.
[13] D. Cheng, S. Xiang, C. Shang, Y. Zhang, F. Yang, L. Zhang, Spatio-temporal attention-based neural network for credit card fraud detection, in Proceedings of the AAAI Conference on Artificial Intelligence vol. 34, 2020, April, pp. 362–369, 1.
[14] A. Zakaryazad, E. Duman, A profit-driven Artificial Neural Network (ANN) withapplications to fraud detection and direct marketing, Neurocomputing 175 (2016) 121–131.
[15] A.S. Bekirev, V.V. Klimov, M.V. Kuzin, B.A. Shchukin, Payment card fraud detection using neural network committee and clustering, Opt. Mem. Neural Network. 24 (3) (2015) 193–200.
[16] S. Georgieva, M. Markova, V. Pavlov, Using the neural network for credit card frauddetection, in AIP Conference Proceedings vol. 2159, AIP Publishing LLC, 2019,October, 030013, 1.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)