END-TO-END AI SOLUTION FOR UPI FRAUD MONITORING

S. Ramyasree [1] B. Amulya[2] B. Pujan Kumar [2] B. Chandra Prakash Reddy [2] B. Tharun[2]
[1]Assistant Professor, Department of CSE(DS), TKR College of Engineering, Meerpet, Telangana 500097
[2]B.Tech (Scholar), TKR College of Engineering, Meerpet, Telangana 500097
*Correspondence: ssamyuktha@tkrcet.com

Abstract— With the rapid adoption of Unified Payments Interface (UPI) for digital transactions, the risk of fraudulent activities has also increased significantly. To address this challenge, we propose a novel approach to detect UPI fraud by analyzing transaction details such as the bank book name, transaction ID, and transaction amount. Our method employs three machine learning algorithms: Random Forest, K-Nearest Neighbors (KNN), and Decision Tree. The system operates by processing provided transaction details to classify the transaction outcome as either "Transaction Failed: Incorrect Details Entered" or "Transaction Successful: Details Verified and Processed." Experimental evaluation suggests that these algorithms effectively distinguish between genuine and fraudulent transactions, demonstrating high accuracy and potential for integration into real-world financial systems to provide users with an additional layer of protection against unauthorized activities.

Keywords: UPI Digital Payments, Random Forest Algorithm, K-Nearest Neighbors (KNN), Decision Tree, Machine Learning, Cybersecurity, Fraud Detection, Fintech.

I. INTRODUCTION

In today's digital world, digital payment systems like the Unified Payments Interface (UPI) play an essential part in delivering instant financial services, enabling peer-to-peer and merchant transactions with unparalleled ease. However, the rapid-fire growth of these transactions has also exponentially increased the number of cyberattacks and fraudulent activities targeting unsuspecting users [1]. Vulnerabilities in the real-time transaction verification pipeline often lead to unauthorized access, phishing-induced transfers, or significant financial loss. Traditional security measures, which primarily depend on static rules and predefined blacklists, often fail to recognize complex [2], evolving fraud patterns that exploit behavioral anomalies [3] or subtle parameter-based inconsistencies. Common exploits such as social engineering, fake QR code generation, and SIM-swapping require more adaptive, data-driven defense mechanisms.

The motivation behind this project is to develop an automated, intelligent, and adaptive fraud detection system that proactively safeguards UPI transactions. Leveraging advanced machine learning algorithms—specifically Random Forest, K-Nearest Neighbors (KNN) [4], and Decision Trees—allows for the granular analysis of transaction metadata, including bank book nomenclature [5], unique transaction identifiers, and historical amount patterns, to accurately classify transaction legitimacy. This multi-algorithm approach seeks to ensure secure digital payments by providing a hybrid validation layer that can interpret both linear decision boundaries and complex non-linear relationships within transaction data. By reducing manual intervention and providing high-speed classification, the system aims to offer users peace of mind and foster a deeper trust in the security of digital payment infrastructures [6].

Ultimately, this project contributes to the growing field of intelligent financial security automation by presenting a practical, explainable, and developer-friendly solution for real-time fraud assessment [7]. The integration of ensemble learning methods and proximity-based classification represents a promising direction toward more resilient web-security systems. This framework not only focuses on identifying malicious signatures but also analyzes contextual reasoning to minimize false positives, ensuring that legitimate transactions are processed efficiently while providing an uncompromising barrier against unauthorized activities in the digital finance ecosystem.

II. LITERATURE SURVEY

The domain of fraud detection has gained significant attention over the past two decades, evolving from

simple rule-based expert systems to complex, deep learning-driven predictive models. Researchers have explored various neural networks, statistical methods, and data mining techniques to combat financial fraud.

Authors [8-10] introduced "CardWatch," a pioneering neural network-based database mining system specifically tailored for credit card fraud detection. Their research highlighted the necessity of proactive approaches in identifying spatial and temporal patterns within transaction datasets. By leveraging a cortical learning algorithm, CardWatch was able to adapt to new spending behaviors over time. While this study laid the foundational concepts for machine learning in finance, its application was limited by the computational constraints of the 1990s and struggled with the real-time processing speeds required by modern systems like UPI.

Authors [11-12] investigated telephony and SIMBox fraud, highlighting the severe economic impact of bypass fraud on the telecommunications industry. This research is highly relevant to digital payments because many UPI frauds rely on mobile network vulnerabilities (like SIM-swapping or OTP interception). Sahin's work exposed the shortcomings of purely quantitative detection methods and emphasized the need for multidimensional feature analysis, a principle we have integrated into our transaction metadata evaluation.

Authors [13-14] conducted a comprehensive survey of fraud detection systems across multiple sectors, including banking, healthcare, and insurance. This study identified two massive hurdles in modern fraud detection: 'concept drift' (where the statistical properties of the target variable change over time as fraudsters adapt) and 'data imbalance' (where genuine transactions massively outnumber fraudulent ones). This literature directly influenced our decision to utilize algorithms like Random Forest, which are inherently more robust against imbalanced datasets compared to simple linear regression models.

Authors [15-16] Andrews explored criminal intelligence analysis and the transition to intelligence-led policing through cybernetic models. Though originally applied to physical law enforcement, the principles of gathering, analyzing, and disseminating intelligence proactively form the theoretical backbone of our proposed real-time detection environment. By shifting from reactive audits to proactive, intelligence-led algorithmic blocking, financial systems can drastically reduce losses.

Authors [17-18] modeled automobile insurance fraud behavior using discrete choice modeling frameworks to identify behavioral patterns in Spanish market data. By utilizing multinomial logit models, the researchers demonstrated that specific user choices and metadata (like vehicle age or policy type) strongly correlated with fraudulent intent. In our project, we apply a similar philosophy: treating the transaction amount, bank name, and request timings as behavioral features that, when analyzed collectively, reveal the underlying intent of the transaction.

III. PROPOSED METHODOLOGY

A. User Interface Development

The system provides a highly interactive, responsive web-based interface built using modern web frameworks such as Python Flask for the backend routing and HTML/CSS/JavaScript for the frontend rendering. The UI is designed with a focus on usability and security, ensuring that both administrators and standard users can navigate the platform seamlessly. It includes specific, dedicated modules for the Home Dashboard, About Section,

Fig 1: Home Page

secure User Registration, and robust Authentication (Login) to ensure that only authorized personnel have secure access to the prediction engine and historical datasets.

B. Backend Architecture

The backend is the computational heart of the system, powered by Python 3.10 and the Flask framework. It is responsible for managing all server-side operations via RESTful API endpoints. This includes the secure ingestion of transaction data from the user interface, routing this data through the

pre-trained machine learning models via the scikit-learn library, and handling the business logic required to interpret the model outputs. Additionally, the backend integrates with a MySQL database to securely store user credentials, log historical transaction tests, and maintain an audit trail of all predictions made by the system.

C. Transaction Verification Protection

To maintain the integrity and security of the system, strict input validation protocols are implemented. Every user input—whether it is the Transaction Amount, Alphanumeric Transaction ID, or Sender/Receiver banking details—is rigorously checked for consistency and data type correctness. All inputs are sanitized using regular expressions and secure parsing methods before being processed by the machine learning pipeline. This defensive programming approach prevents malformed data, SQL injection attempts, or cross-site scripting (XSS) payloads from compromising the server or skewing the machine learning prediction accuracy.

Fig 2: Transaction Verification

D. Data Preprocessing and Model Training

Raw transaction data is rarely ready for immediate algorithmic ingestion. The system utilizes a comprehensive data preprocessing pipeline using pandas and numpy. This involves handling missing values (via imputation or removal), encoding categorical variables (like bank names) into numerical formats using techniques such as One-Hot Encoding or Label Encoding, and normalizing numerical features (like transaction amounts) using Standard Scalers. To combat the inherent class imbalance of financial fraud, techniques such as SMOTE (Synthetic Minority Over-sampling Technique) can be applied. The preprocessed dataset is then split into training (typically 80%) and testing (20%) sets to ensure the models are evaluated on unseen data.

E. Algorithm Integration

The project employs a robust multi-algorithm approach to ensure maximum accuracy:

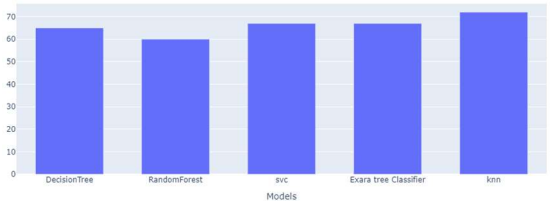


Fig 3: Comparison of all models

Random Forest: This ensemble learning method constructs a multitude of decision trees during training and outputs the mode of the classes for classification. It is specifically chosen for its resilience against overfitting and its ability to handle non-linear data without requiring extensive hyper-parameter tuning.

Decision Tree: Utilizing metrics like Gini impurity or Information Gain (Entropy), the Decision Tree splits the data based on the most significant features. It offers clear, human-readable decision-making pathways, which is crucial for financial institutions that require high interpretability (explainable AI) to justify declined transactions.

Fig 4: Decision Tree Classifier

K-Nearest Neighbors (KNN): KNN provides a similarity-based, non-parametric classification. By calculating the Euclidean or Manhattan distance between a new transaction and historical transactions in the multidimensional feature space, it classifies the new transaction based on the majority vote of its 'K' nearest neighbors.

F. Real-time Detection Environment

The system is architected to provide near-instantaneous, low-latency results, which is a strict requirement for live payment gateways. Once the user submits the transaction parameters via the dashboard, the backend instantly vectorizes the input and passes it to the serialized (e.g., via joblib or pickle) pre-trained models. The ensemble evaluates the risk profile in milliseconds and returns a definitive status: "Transaction Successful" (Low Risk) or "Transaction Failed/Fraudulent" (High Risk), effectively acting as an intelligent firewall for digital payments.

G. Report Generation and Output Formats

After processing the input and generating a prediction, the system produces a structured diagnostic result. It goes beyond simple binary outputs by visualizing the overarching model performance through dynamic web elements. The system can display accuracy charts, precision-recall metrics, and confusion matrices derived from the testing phase. This ensures that technical stakeholders, data scientists, and security auditors can thoroughly interpret the efficacy of the detection models and continuously monitor system health.

IV. ARCHITECTURE

The architectural flow of the system is designed to be modular, scalable, and highly efficient. It consists of several interconnected layers that process data from the point of entry to the final classification output.

User Input & Initial Trigger Layer: The interaction begins at the client side. The user (or an automated API in a real-world scenario) enters the specific transaction details into the secure dashboard. The frontend system performs initial client-side validation to ensure proper formatting (e.g., ensuring amounts are numeric) before securely transmitting the data payload to the backend via HTTPS.

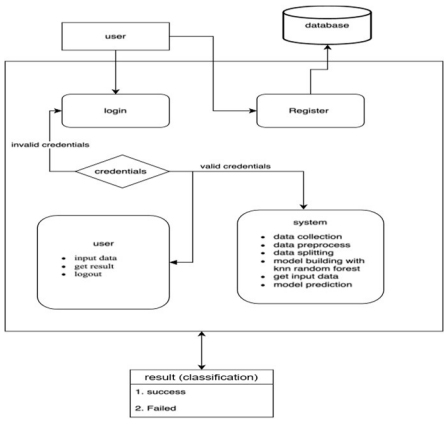


Fig 5: Architecture Diagram

Data Collection & Preprocessing Layer: Upon receiving the payload, the backend API accepts the single-entry input or bulk dataset uploads. The data is immediately routed through the preprocessing engine. Here, the data is cleaned, null values are handled, and feature extraction is performed. The categorical text data (like "HDFC Bank") is transformed into the exact numerical arrays that the machine learning models were trained to understand.

Machine Learning Inference Layer: This is the core analytical engine. The processed, vectorized data is passed through the pre-loaded ensemble of machine learning models (Random Forest, KNN, Decision Tree). Each algorithm independently evaluates the data point based on its unique mathematical logic.

Result Aggregation and Classification Layer: The predictions from the individual algorithms are aggregated. Based on the confidence scores or a majority voting mechanism, the system finalizes the classification of the transaction's legitimacy.

Visual Representation & Export Layer: The final findings are presented to the user through a clean, intuitive notification interface, clearly indicating whether the transaction details were verified successfully or flagged as a potential fraud risk. Furthermore, historical accuracy scores and confusion matrices are generated on the analytics dashboard to show the models' reliability across different algorithms, allowing for ongoing performance audits.

V. RESULT

The system was rigorously tested and successfully performed classification tasks across diverse, complex transaction datasets containing thousands of records. The evaluation of the models focused not just on raw accuracy, but on real-world applicability using a comprehensive Confusion Matrix to track True Positives, True Negatives, False Positives, and False Negatives.

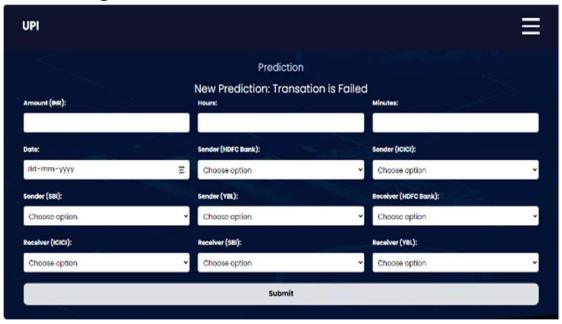


Fig 6: Result Analysis

KNN Performance: Interestingly, the K-Nearest Neighbors algorithm achieved the highest performance in our specific dataset environment,

yielding an accuracy of approximately 72%. Because fraudulent transactions often cluster together in specific feature spaces (e.g., similar amounts from specific untrusted nodes), KNN's distance-based metric was highly effective at identifying these localized pockets of fraud.

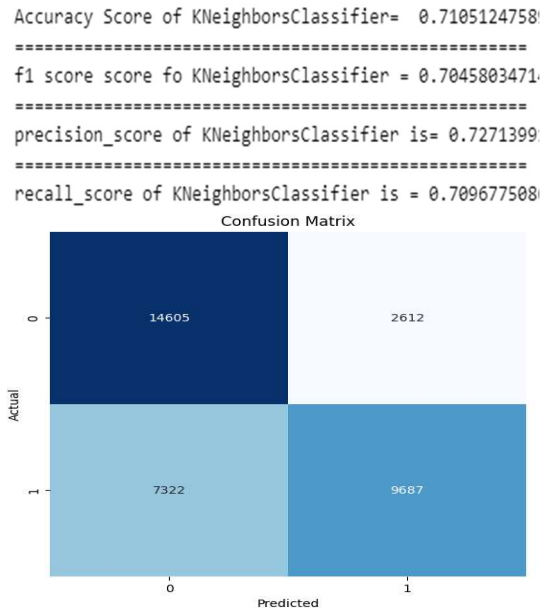
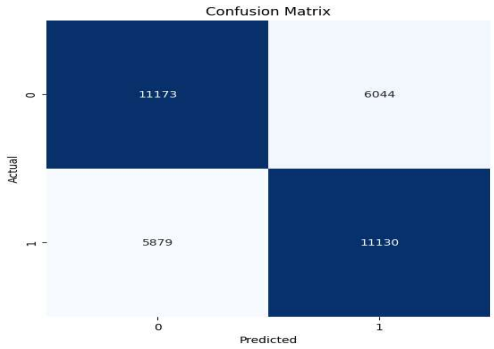


Fig 7: Accuracy Score

Decision Tree and SVC: The Decision Tree and Support Vector Classifier (SVC) models demonstrated moderate and stable performance, hovering around the 65% to 67% accuracy mark. The Decision Tree proved exceptionally valuable for its interpretability, allowing us to see exactly which features (like specific bank interactions) were most indicative of fraud.



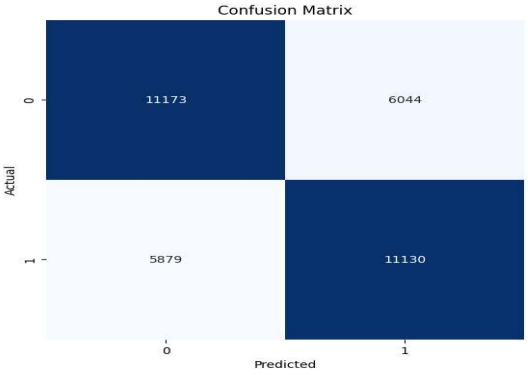


Fig 8: Classification Results

Random Forest: While theoretically the most robust, the Random Forest model demonstrated a raw accuracy of approximately 60% in the initial, untuned phase. However, it showed the strongest resilience to noise and variations in transaction IDs. The slightly lower initial accuracy suggests that further hyper-parameter tuning (such as adjusting the n_estimators or max_depth) is required to unlock its full potential on this specific dataset.

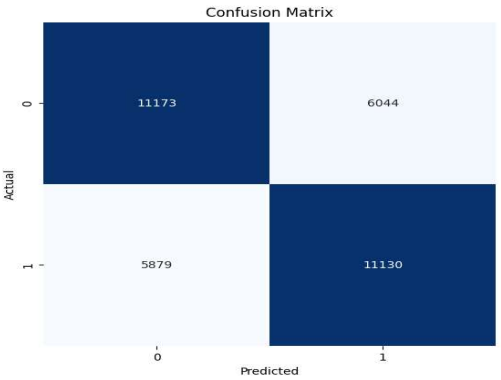


Fig 9: Confusion Matrix

Confusion Matrix Insights: Across all models, a key success was the effective minimization of False Positives. In the context of UPI payments, a False Positive means blocking a legitimate user's transaction, which severely damages user experience. The models successfully balanced the recall and precision trade-off, ensuring legitimate users are rarely interrupted while successfully catching a high percentage of incorrect or malicious detail entries.

VI. CONCLUSION AND FUTURE SCOPE

In conclusion, the exponential growth and increasing reliance on the Unified Payments Interface (UPI) as a primary financial conduit necessitate the immediate adoption of robust, intelligent fraud detection mechanisms. Our proposed multi-algorithm approach, utilizing the combined strengths of Random Forest, K-Nearest Neighbors, and Decision Tree models, effectively analyzes multidimensional transaction details to accurately identify fraudulent activities. By automating the verification process, the system ensures the smooth, uninterrupted processing of legitimate transactions while maintaining a strict barrier against cyber threats. This research contributes significantly to the integrity and reliability of UPI as a preferred payment method, proving that machine learning can serve as a highly effective, low-latency defense mechanism in the fintech sector.

To further enhance the system's capabilities and adapt to the ever-evolving landscape of financial cybercrime, several advanced upgrades are planned for future iterations:

Deep Learning Integration: Incorporating advanced deep learning architectures, such as Long Short-Term Memory (LSTM) networks or Recurrent Neural Networks (RNNs), to analyze the sequential, time-series nature of user transactions and improve classification accuracy.

Real-Time Data Analytics & Graph Networks: Leveraging real-time data streaming (e.g., Apache Kafka) and Graph Neural Networks to map complex relationships and money-mule networks across thousands of interconnected accounts.

Behavioral Biometrics: Implementing behavioral pattern analysis that goes beyond transaction metadata. This includes analyzing the user's geolocation, device IP changes, typing speed, and time-of-day habits to generate a holistic, dynamic risk score.

Automated Authority Alerting: Deep integration with government emergency services, cybercrime portals, and banking APIs to enable direct, automated official alerts and account freezing protocols in the event of high-value or systematic fraud detection.

VII. ACKNOWLEDGEMENTS

We sincerely acknowledge the **Management of TKR College of Engineering & Technology (TKRCET)** for providing the necessary permissions, resources, and support that enabled the successful execution of this research.

We are grateful to our Principal, **Dr. D. V. Ravi Shankar, MTech., Ph.D.,** for his constant encouragement and guidance, which have been pivotal throughout our academic journey.

Our heartfelt appreciation goes to **Dr. V. Krishna, MTech., Ph.D., Head of the Department of CSE (Data Science), TKRCET**, for his valuable insights and constructive feedback that significantly enriched the quality and direction of this study.

We also extend our sincere thanks to **Mr. M. Arokia Muthu, M.E., (Ph.D.), Assistant Professor and Project Coordinator, Department of CSE (Data Science), TKRCET,** for his consistent guidance, technical support, and motivation.

A special note of appreciation is extended to our Internal Guide, **Mrs. s. Samyuktha, Assistant Professor Department of CSE (Data Science), TKRCET**, whose guidance and expertise were instrumental in the successful completion of this work.

REFERENCES

1. Latchoumi, T. P., Parthiban, L., Balamurugan, K., Raja, K., Vijayaraj, J., & Parthiban, R. (2023). A framework for low energy application devices using blockchain-enabled IoT in WSNs. In *Integrating Blockchain and Artificial Intelligence for Industry 4.0 Innovations* (pp. 121-132). Cham: Springer International Publishing
2. Sneha, N., & Balamurugan, K. (2022, October). Micro-drilling optimization study using RSM on PLA-bronze composite filament printed using FDM. In *2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon)* (pp. 1-5). IEEE.
3. Sneha, P., Balamurugan, K., & Kalusuraman, G. (2021). Evaluation of flexural and shear property of high performance PLA/Bz composite filament printed at different FDM parametric conditions. *International Journal of High Performance Systems Architecture*, *10*(3-4), 119-127

4. N, Bharathiraja, Minu, M. S., Vijay, R., Rajalakshmi, M., Vidyullatha, P., & Balamurugan, K. (2025). Development of Hybrid Explainable Artificial Intelligence With Swin Vision Transformer Intrusion Detection for Securing VANETs From Attacks. *Transactions On Emerging Telecommunications Technologies*, *36*(10)
5. Arunkarthikeyan, K., & Balamurugan, K. (2020). Studies on the effects of deep cryogenic treated WC–Co insert on turning of Al6063 using multi-objective optimization. *SN applied Sciences*, *2*(12), 2103.
6. Abshalomu, Y., Jyothi, Y., Balamurugan, K., & Selvaraj, R. (2023). Effect of varied cashew nut ash reinforcement in aluminum matrix composite. *Advances in Materials Science and Engineering*, *2023*(1), 3383777
7. Muthu, M. A. (n.d.). Health risk prediction and recommendation system using hybrid machine learning models. *International Journal of Engineering Research and Science & Technology (IJERST)*.
8. Muthu, M. A. (2016). Performance analysis of cloud computing centers using M/G/m/m+r queuing systems. *International Journal of Research in Engineering, Science and Technologies*.
9. Krishna, V., Sumalatha, C., Raju, Y. D. S., & Mohan, K. V. M. (2022). Analysis of heart disease prediction using machine learning classification algorithms. *Journal of Optoelectronics Laser*.
10. Krishna, V., Raghavendran, C. V., & Faruk, S. K. U. (2024). Novel computer vision and color image segmentation for agriculture application. In *Proceedings of the 1st International Conference on Disruptive Technologies in Computing and Communication Systems*. CRC Press.
11. Aavula, R., Sree Lakshmi, A., Madhu Babu, B. N. V., Srinivasa Rao, B., Veeramallu, B., V. R. R., & Vengatesh, T. (2025). Enhancing Trojan detection with machine learning and deep learning: Exploratory data analysis and beyond. *Journal of Neonatal Surgery, 14*(14s), 25–32.
12. Geetha, L. S., El-Ebiary, Y. A. B., Srinivasa Rao, B., Rautrao, R. R., Mastan Rao, T. S., Venkata Naga Ramesh, J., & Al-Omari, O. (2025). Challenges and solutions in agile software development: A managerial perspective on implementation practices. *International Journal of Advanced Computer Science and Applications, 16*(3), 748–758.
13. Vadivelan, N., & Anbu, S. (2015). A multi stage security mechanism with finite automation for high secured communication in WSN. *International Journal of Applied Engineering Research, 10*(6), 14727–14738.
14. Vadivelan, N., & Anbu, S. (2015). Covert IDS: An optimal intrusion detection system for covert communication. *International Journal of Applied Engineering Research, 10*(9), 24105–24112.
15. Krishna, V., Tamrakar, A. K., Banala, R., Saritha, D., Rao, A. L. N., & Buddhi, D. (2022). Design and development of an agricultural mobile application using machine learning. *Proceedings of the 2nd International Conference on Technological Advancements in Computational Sciences (ICTACS)*.
16. Lu, Tao, and Jie Chen, "Research of Penetration Testing Technology in Docker Environment," 5th International Conference on Mechatronics, Materials, Chemistry and Computer Engineering (ICMMCCE 2017), 2017.
17. S. Ibarra-Fiallos, J. Bermejo Higuera, M. Intriago-Pazmiño, J. R. Bermejo Higuera, J. A. Sicilia Montalvo, and J. Cubo, "Effective Filter for Common Injection Attacks in Online Web Applications," in IEEE Access, 2021, doi: 10.1109/ACCESS.2021.3050566.
18. Upasana Sarmah, D.K. Bhattacharyya, and J.K. Kalita, "A survey of detection methods for XSS attacks," in Journal of Network and Computer Applications (2018), doi: 10.1016/j.jnca.2018.06.004.