International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 13, Issue 2s, 2025

Medical Insurance Claim Using Face Id

Sheikh Zeeshan Rehman,
B.E Student, Department Of IT,
ISL Engineering College (O.U), India

Syed Sameer Hashmi,
B.E Student, Department Of IT,
ISL Engineering College (O.U), India

Shaik Abdul Gaffar,
B.E Student, Department Of IT,
ISL Engineering College (O.U), India

Dr. Surya Mukhi,
Associate professor, Department Of IT,
ISL Engineering College (O.U), India

sheikhzishan4@gmail.com

ABSTRACT

To protect healthcare providers' finances and stop systematic abuse, it is essential to verify the veracity of medical insurance claims. In order to identify irregularities and fraudulent activity in health insurance claims, this study investigates a hybrid machine learning architecture that combines Support Vector Machines (SVM), Decision Trees, and Random Forest classifiers. To refine the input variables, a carefully selected dataset was subjected to extensive preprocessing, which included feature transformation, data normalization, and sophisticated dimensionality reduction techniques. GridSearchCV was used for hyperparameter tuning, which systematically found the best parameter combinations to maximize the prediction performance of each model. Metrics including accuracy, precision, recall, and F1-score were used to assess the effectiveness of the classifiers. The results showed that the modified ensemble models—Random Forest in particular—had better detection skills than more conventional methods.

1. INTRODUCTION

Recent years have seen a sharp increase in medical insurance claims due to the quick digitization of healthcare services and the growing dependence on insurance coverage. Although this change has made healthcare more accessible, it has also opened up new channels for fraud, which can endanger the integrity of healthcare systems as a whole and cause insurance companies to suffer large financial losses. Therefore, identifying and stopping false claims has become a top priority. The volume and complexity of claim data frequently make traditional manual auditing techniques inadequate. By automating the fraud detection process more accurately and quickly, sophisticated machine learning (ML) algorithms provide a potent alternative to this problem.

In order to improve model accuracy and generalizability, a comprehensive data pretreatment pipeline is put into place, including methods like data cleaning, normalization, feature engineering, and dimensionality reduction. The models are refined using GridSearchCV, a potent hyperparameter optimization tool that thoroughly searches through a variety of parameter combinations to guarantee peak performance. Accuracy, precision, recall, and F1-score are among the complete evaluation metrics used to evaluate the system's efficacy; these measures taken together offer a fair assessment of model performance. Because of its ensemble-based decision-making ability, the Random Forest classifier outperforms the other models under evaluation and is especially good at identifying intricate patterns in the data.

The creation and assessment of a machine learning-based system for identifying fraudulent activity in medical insurance claim databases is included in the study's scope. The goal of the project is to apply a variety of classification methods, including Random Forests, Decision Trees, and Support Vector Machines (SVM), to accurately identify and flag dubious claims. In order to maximize model performance, the study also highlights how crucial it is to use GridSearchCV for hyperparameter tuning. Extensive data preprocessing, including feature selection, engineering, and normalization, is an essential

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

component of this effort in order to improve the model's learning capabilities. The scope goes beyond fraud detection to include secure biometric authentication, particularly Face ID-based registration and login, which gives the system an extra degree of protection.

2. LITERATURE SURVEY

Title: Infrared small object detection using deep interactive U-Net
Author: X. Wu, D. Hong, Z. Huang, and J.Chanussot
Year: 2022.
Description: Infrared objects acquired from a long-distance have small sizes and are easily submerged by a complex and variable background. The existing deep network detection framework suffers greatly from the feature spatial resolution loss caused by the networks' depth and multiple downsampling operations, which is extremely detrimental for small object detection. So, a crucial and urgent goal is, how to trade-off network depth and feature spatial resolution, while learning feature context representation and interaction to distinguish from the background. To this end, we propose a deep interactive U-Net architecture (short for DI-U-Net) with high feature learning and feature interaction ability. First, feature learning is first achieved through a multi-level and high-resolution network structure. This structure ensures feature resolution as the network depth increase, and also focus on the object's global context information. Then, the feature interactive is further achieved by the dense feature encoder (DFI) module to learn object local context information. The proposed method yields strong object context representation and well discriminability, as well as a good fit for infrared small object detection. Extensive experiments are conducted on the SISRT dataset and Synthetic dataset, demonstrating the superiority and effectiveness of the proposed deeper U-Net compared to previous state-of-the-art detection methods.

Title: Sequence Mining and Prediction-Based Healthcare Fraud Detection Methodology.
Author: I. Matloob, S. A. Khan, and H. U. Rahman.
Year: 2020.
Description: This article presents a novel methodology to detect insurance claim related frauds in the healthcare system using concepts of sequence mining and sequence prediction. Fraud detection in healthcare is a non-trivial task due to the heterogeneous nature of healthcare records. Fraudsters behave as normal patients and with the passage of time keep on changing their way of planting frauds; hence, there is a need to develop fraud detection models. The sequence generation is not the part of previous researches which mostly focus on amount based analysis or medication versus diseases sequential analysis. The proposed methodology is able to generate sequences of services availed or prescribed by each specialty and analyse via two cascaded checks for the detection of insurance claim related frauds. The methodology addresses these challenges and self learns from historical medical records. It is based on two modules namely ''Sequence rule engine and Prediction based engine''. The sequence rule engine generates frequent sequences and probabilities of rare sequences for each specialty of the hospital. The comparison of such sequences with the actual patient sequences leads to the identification of anomalies as both sequences are not compliant to the sequences of the rule engine. The system performs further in detail analysis on all non-compliant sequences in the prediction based engine. The proposed methodology is validated by generating patient sequences from last five years transactional data of a local hospital and identifies patterns of service procedures administered to patients using Prefixspan algorithm and Compact prediction tree. Various experiments have been performed to validate the applicability of the developed methodology and the results demonstrate that the methodology is pertinent to detect healthcare frauds and provides on average 85% of accuracy. Thus can help in preventing fraudulent claims and provides better insight into how to improve patient management and treatment procedures

Title: Blockchain-Powered Parallel Healthcare Systems Based on the ACP Approach
Author: S. Wang et al,
Year: 2010.
Description: To improve the accuracy of diagnosis and the effectiveness of treatment, a framework of parallel healthcare systems (PHSs) based on the artificial systems + computational experiments + parallel execution (ACP) approach is proposed in this paper. PHS uses artificial healthcare systems to model and represent patients' conditions, diagnosis, and treatment process, then applies computational experiments to analyze and evaluate various therapeutic regimens, and implements parallel execution for decision-making support and real-time optimization in both actual and artificial healthcare processes. In addition, we combine the emerging blockchain technology with PHS, via constructing a consortium blockchain linking patients, hospitals,

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

health bureaus, and healthcare communities for comprehensive healthcare data sharing, medical records review, and care auditability. Finally, a prototype named parallel gout diagnosis and treatment system is built and deployed to verify and demonstrate the effectiveness and efficiency of the blockchain-powered PHS framework.

Title: A Secure Healthcare System Design Framework using Blockchain Technology.
Author: S. Chakraborty, S. Aich, and H. C. Kim,
Year: 2019
Description: Blockchain, the technology of the future neutrally facilitated the financial transactions in crypt currencies by strictly eliminating the need for a governing authority or a management that was required to authorize the transactions based on trust and transparency. The Blockchain Network also follows the principle of absolute privacy and anonymity on the identification of the users associated in a transaction. Since the time of its inception, the Blockchain Technology has undergone research that has demonstrated some various kinds of methods to sort out the access control system of the conventional system. In recent years Blockchain has also shown optimum reliability in multiple sectors such as Smart Home, Healthcare, Banking, Information Storage Management, Security and etc. This work in terms is further concerned to the sector of Smart Healthcare, which has grown to a much affluence regarding the efficient technique of serving and dictating medical health care to the patients with the point of maintaining privacy of the patients' data and also the process of laying out real time accurate and trusted data to the medical practitioners. But in the scenario of Smart Healthcare, the primary concern arises in the fact of Privacy and Security of the data of the patients due to the interoperability of multiple stakeholders in the process. Also, there has been a fact of determining accurate and proper data to the doctors if the concerned subject is out of reach from the in hand medical service. Therefore, this Concern of privacy and also mitigation of the accurate data has been very much managed in the work by regulating, a monitoring and sensing paradigm with accordance to the IOT and the Blockchain as a transaction and access management system and also an appropriate medium for laying out accurate and trusted data for serving with deliberate medical care and benefits to the patients across.

Title: A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement.
Author: N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud.
Year: 2019**.**
Description: The private insurance sector is recognized as one of the fastest-growing industries. This rapid growth has fueled incredible transformations over the past decade. Nowadays, there exist insurance products for most high-value assets such as vehicles, jewellery, health/life, and homes. Insurance companies are at the forefront in adopting cutting-edge operations, processes, and mathematical models to maximize profit whilst servicing their customers claims. Traditional methods that are exclusively based on human-in-the-loop models are very time-consuming and inaccurate. In this paper, we develop a secure and automated insurance system framework that reduces human interaction, secures the insurance activities, alerts and informs about risky customers, detects fraudulent claims, and reduces monetary loss for the insurance sector. After presenting the blockchain-based framework to enable secure transactions and data sharing among different interacting agents within the insurance network, we propose to employ the extreme gradient boosting (XGBoost) machine learning algorithm for the aforementioned insurance services and compare its performances with those of other state-of-the-art algorithms. The obtained results reveal that, when applied to an auto insurance dataset, the XGboost achieves high performance gains compared to other existing learning algorithms. For instance, it reaches 7% higher accuracy compared to decision tree models when detecting fraudulent claims. The obtained results reveal that, when applied to an auto insurance dataset, the XGboost achieves high performance gains compared to other existing learning algorithms. For instance, it reaches 7% higher accuracy compared to decision tree models when detecting fraudulent claims. Furthermore, we propose an online learning solution to automatically deal with real-time updates of the insurance network and we show that it outperforms another online state-of-the-art algorithm. Finally, we combine the developed machine.

3. RELATED WORK

The healthcare sector has long faced a serious problem with fraudulent medical insurance claims, which can result in monetary losses and erode public confidence in healthcare systems. With numerous research

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

examining different categorization algorithms to find unusual patterns in claims data, the application of machine learning to detect such fraud has attracted more interest in recent years. Although initially successful, traditional rule-based systems frequently lack the agility and flexibility necessary to identify novel and changing fraud tactics. Machine learning algorithms, on the other hand, have the ability to dynamically learn from past data and adjust to intricate and nuanced fraud patterns that human analysts could miss.

[1] Among the most widely studied algorithms in fraud detection are Support Vector Machines (SVM), Decision Trees, and Random Forests. SVMs have demonstrated strong performance in high-dimensional spaces and are particularly effective when there is a clear margin of separation between classes. However, they can struggle with large datasets and noisy data. Decision Trees, known for their interpretability and low computational complexity, have been commonly applied in healthcare fraud detection due to their ability to model non-linear relationships. Nonetheless, they tend to overfit the training data if not carefully pruned. [2] Random Forests, an ensemble technique based on multiple Decision Trees, have emerged as a powerful alternative, offering higher accuracy and better generalization by reducing variance through bagging. Several comparative studies have indicated that Random Forests often outperform individual classifiers in the context of fraud detection, especially when paired with proper feature selection and tuning techniques. Another strategy for utilizing the CNN conspire is proposed in [3] To enhance the performance of these models, recent research has focused on hyperparameter optimization techniques such as GridSearchCV. This method exhaustively searches through a specified subset of hyperparameters, evaluating the model using cross-validation to find the best combination. The effectiveness of GridSearchCV has been validated across various applications, including financial fraud, cyber security, and medical insurance, where finding the right model configuration can significantly affect predictive performance. Studies show that when hyperparameter tuning is employed, even relatively simple models like Decision Trees can achieve competitive accuracy compared to more complex architectures. [4] Building trustworthy fraud detection systems requires not just model tweaking but also feature engineering and thorough data pretreatment. Missing numbers, categorical attributes, class imbalance, and noise are common issues with medical insurance data. [5] To lessen these problems, methods including data imputation, one-hot encoding, normalization, and synthetic sampling (like SMOTE) have been suggested. It has also been demonstrated that efficient feature engineering significantly enhances model performance. This includes the extraction of domain-specific indicators and the conversion of unstructured qualities into useful features. This procedure has been made even more efficient in recent years by the application of automated feature selection techniques including mutual information-based ranking and recursive feature elimination (RFE).[6] The foundation of fraud detection is machine learning, but protecting the system from unwanted access is just as crucial. As digital platforms in healthcare have grown in popularity, it is now crucial to protect sensitive data using strong authentication procedures. The ease of use and high level of security of biometric authentication, especially Face ID-based login systems, have made it a potential strategy. According to studies, facial recognition is more convenient for end users and resistant to spoofing than standard password-based systems. When it comes to handling personally identifiable information (PII) in medical datasets, the incorporation of biometric authentication into fraud detection systems guarantees both data confidentiality and user accountability.

4. METHODOLOGIES

By including a safe Face ID-based verification method, the suggested framework seeks to create an all-encompassing and intelligent system for identifying false medical insurance claims. Data collection and preprocessing, exploratory data analysis, feature engineering, model selection and training, hyperparameter tuning, evaluation and validation, and system integration with biometric authentication are the different yet interconnected processes that make up the technique. Each stage is described in detail in this section:

- Data Acquisition and Preprocessing
- Feature Engineering and Selection
- Machine Learning Model Development
- Hyperparameter Optimization using GridSearchCV
- Model Evaluation and Performance Metrics

Data Acquisition and Preprocessing

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

To guarantee quality and consistency for model training, a large dataset of medical insurance claims is gathered from reputable sources, cleaned, missing values are handled, categorical variables are encoded, and data is normalized.

Feature Engineering and Selection

To improve the predictive performance and lower dimensionality of the model, new, pertinent features are created and the most useful ones are chosen using statistical and domain-specific methods.

Machine Learning Model Development

constructing prediction models with algorithms like Random Forests, Decision Trees, and Support Vector Machines (SVM) that are specifically designed to identify patterns suggestive of false claims

Hyperparameter Optimization using GridSearchCV

Through thorough cross-validation, the ideal parameters for each algorithm are found by fine-tuning the model parameters using GridSearchCV. This maximizes accuracy and minimizes errors.

Model Evaluation and Performance Metrics

To verify the robustness of the trained models and guarantee accurate fraud detection, evaluation metrics such as accuracy, precision, recall, and F1-score are used to gauge the efficacy of the models.

5. OUR PROPOSED MODEL

SVM and GridSearchCV

By utilizing Support Vector Machines (SVM) and GridSearchCV for model optimization, the suggested method intends to improve fraud detection in medical insurance claims. This method makes use of GridSearchCV's effectiveness in fine-tuning model parameters and SVM's ability to handle high-dimensional data. Data collection and preprocessing, model construction, training, and evaluation, as well as interaction with the current insurance claim processing workflow, are all included in the system's essential components. The SVM model learns to distinguish between false and authentic claims through careful feature building and selection.

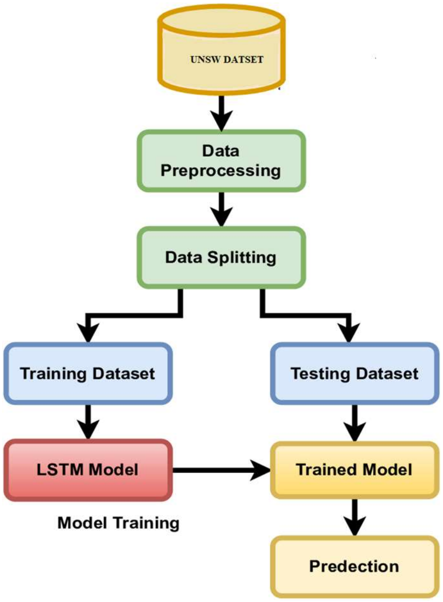


Figure 1. System Architecture.

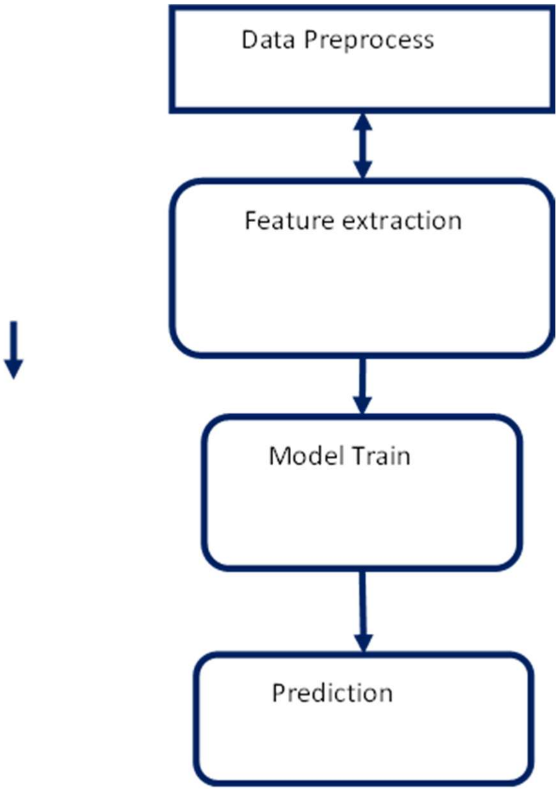


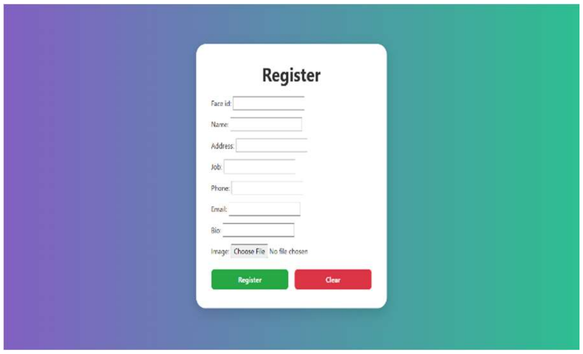
Figure 2.

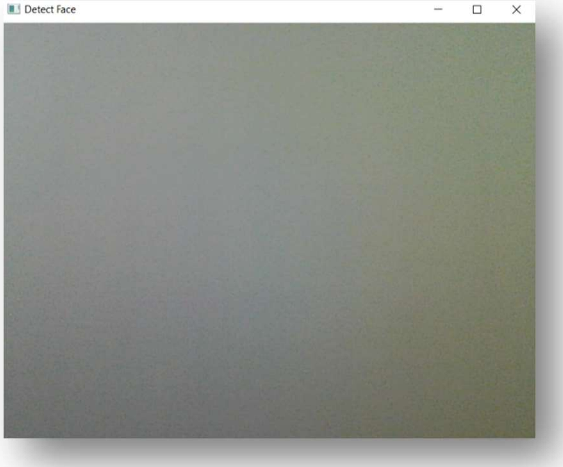
6. RESULT

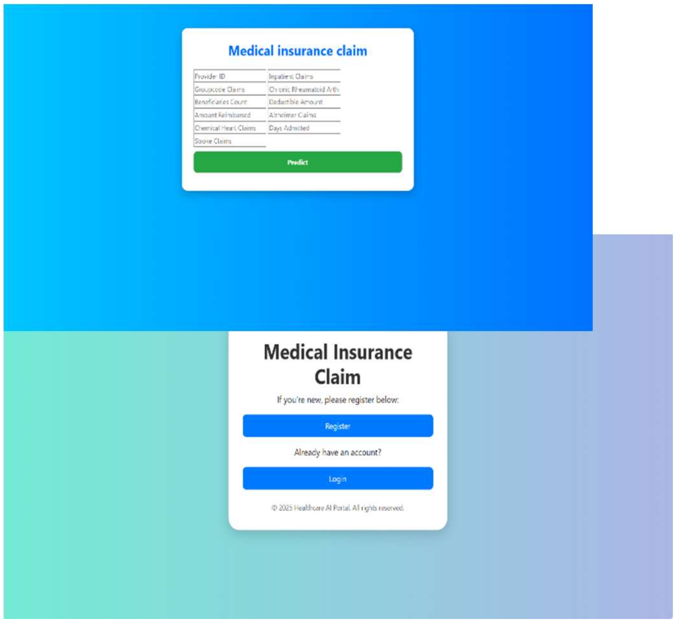
DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

ISSN 2347–3657

Volume 13, Issue 2s, 2025







If you're new, please register below:

Already have an account?

© 2025 Healthcare AI Portal. All rights reserved.

Face Registration: Users upload or scan their face during sign-up, and facial embeddings are stored securely.

Authentication: During login, a real-time camera capture is compared against stored embeddings for identity verification.

Svm model responsible for predicting fraud credit claims

Output Screen

Fig.1

Fig 2

Fig 3

Fig 4

7. CONCLUSION AND FUTURE ENHANCEMENT

In the future, a number of improvements could be made to further bolster the system's capabilities. To capture more intricate fraud patterns and sequential claim behaviors, one significant advancement would be the incorporation of deep learning models like Convolutional Neural Networks (CNNs) or Recurrent Neural Networks (RNNs). Insurance companies may also be able to respond quickly to questionable claims by combining real-time fraud detection with streaming data processing. Multi-factor authentication, which combines Face ID with fingerprint or OTP verification for enhanced security, can also be added to the present biometric authentication system. system that is more scalable and accessible. An ecosystem for fraud detection that is more sophisticated, flexible, and safe would result from these improvements.

8. REFERENCES

1. N. Dhieb, H. Ghazzai, H. Besbes, and Y. Massoud, “A Secure AI-Driven Architecture for Automated Insurance Systems: Fraud Detection and Risk Measurement,” IEEE Access, vol. 8, pp. 58546–58558, 2020, doi: 10.1109/ACCESS.2020.2983300.
2. S. Tanwar, Q. Bhatia, P. Patel, A. Kumari, P. K. Singh, and W. C. Hong, “Machine Learning Adoption in Blockchain-Based Smart Applications: The Challenges, and a Way Forward,” IEEE Access, vol. 8, pp. 474–448, 2020, doi: 10.1109/ACCESS.2019.2961372.

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335

3. [3] M. Bärtl and S. Krummaker, “Prediction of claims in export credit finance: a comparison of four machine learning techniques,” Risks, vol. 8, no. 1, 2020, doi: 10.3390/risks8010022.
4. L. Ismail and S. Zeadally, “Healthcare Insurance Frauds: Taxonomy and Blockchain-Based Detection Framework (Block-HI),” IT Prof., vol. 23, no. 4, pp. 36–43, 2021, doi: 10.1109/MITP.2021.3071534.
5. I. Matloob, S. A. Khan, and H. U. Rahman, “Sequence Mining and Prediction-Based Healthcare Fraud Detection Methodology,” IEEE Access, vol. 8, pp. 143256–143273, 2020, doi: 10.1109/ACCESS.2020.3013962.
6. M.A.Bari & Shahanawaj Ahamad, “Object Identification for Renovation of Legacy Code”, in International Journal of Research and Reviews in Computer Science (IJRRCS),ISSN:2079-2557,Vol:2,No:3,pp:769-773,Hertfordshire,U.K., June 2011
7. G. Kowshalya and M. Nandhini, “Predicting Fraudulent Claims in Automobile Insurance,” Proc. Int. Conf. Inven. Commun. Comput. Technol. ICICCT 2018, no. Icicct, pp. 1338–1343, 2018, doi: 10.1109/ICICCT.2018.8473034.
8. S. Wang et al., “Blockchain-Powered Parallel Healthcare Systems Based on the ACP Approach,” IEEE Trans. Comput. Soc. Syst., vol. 5, no. 4, pp. 942–950, 2018, doi: 10.1109/TCSS.2018.2865526.
9. S. Chakraborty, S. Aich, and H. C. Kim, “A Secure Healthcare System Design Framework using Blockchain Technology,” Int. Conf. Adv. Commun. Technol. ICACT, vol. 2019-February, pp. 260–264, 2019, doi: 10.23919/ICACT.2019.8701983.
10. T. T. A. Dinh, R. Liu, M. Zhang, G. Chen, B. C. Ooi, and J. Wang, “Untangling Blockchain: A Data Processing View of Blockchain Systems,” IEEE Trans. Knowl. Data Eng., vol. 30, no. 7, pp. 1366–1385, 2018, doi: 10.1109/TKDE.2017.2781227.
11. W. Kozlow, M. J. Demeure, L. M. Welniak, and J. L. Shaker, “Acute extracapsular parathyroid hemorrhage: Case report and review of the literature,” Endocr. Pract., vol. 7, no. 1, pp. 32 36, 2001, doi: 10.4158/ep.7.1.32.
12. M. Raikwar, S. Mazumdar, S. Ruj, S. Sen Gupta, A. Chattopadhyay, and K. Lam, “2018 9th IFIP International Conference on New Technologies, Mobility and Security, NTMS 2018 - Proceedings,” 2018 9th IFIP Int. Conf. New Technol. Mobil. Secur. NTMS 2018 - Proc., vol. 2018-January, 2018.
13. R. Roy and K. T. George, “Detecting insurance claims fraud using machine learning techniques,” Proc. IEEE Int. Conf. Circuit, Power Comput. Technol. ICCPCT 2017, 2017, doi: 10.1109/ICCPCT.2017.8074258.
14. X. Liang, J. Zhao, S. Shetty, J. Liu, and D. Li, “Integrating blockchain for data sharing and collaboration in mobile healthcare applications,” IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC, vol. 2017-October, pp. 1–5, 2018, doi: 10.1109/PIMRC.2017.8292361.
15. F. Tang, S. Ma, Y. Xiang, and C. Lin, “An Efficient Authentication Scheme for Blockchain Based Electronic Health Records,” IEEE Access, vol. 7, pp. 41678–41689, 10.1109/ACCESS.2019.2904300.
16. Mr. Pathan Ahmed Khan, Dr. M.A Bari,: Impact Of Emergence With Robotics At Educational Institution And Emerging Challenges”, International Journal of Multidisciplinary Engineering in Current Research(IJMEC), ISSN: 2456-4265, Volume 6, Issue 12, December 2021,Page 43-46
17. Shahanawaj Ahamad, Mohammed Abdul Bari, Big Data Processing Model for Smart City Design: A Systematic Review “, VOL 2021: ISSUE 08 IS SN : 0011-9342 ;Design Engineering (Toronto) Elsevier SCI Oct : 021;Q4 Journal

DOI: https://doi.org/10.62647/IJITCE2025V13I2sPP328-335