Data Security in Healthcare Systems: Integration of Information Security and Information Management

Ahmed Abdelaziz, Alia N. Mahmoud

Nova Information Management School, Universidade Nova de Lisboa, 1070-312, Lisboa, Portugal

Emails: D20190535@novaims.unl.pt; M20190508@novaims.unl.pt

Abstract

Effective management of patient data is critical for improving the quality of care and patient outcomes in healthcare systems. However, ensuring the confidentiality, integrity, and availability of patient data while complying with regulatory requirements can be challenging. To address this challenge, this work proposes an artificial intelligence (AI)-enabled framework that integrates information security (IS) and information management (IM) capabilities into a unified solution for improving the overall functionality of healthcare systems. The proposed framework leverages AI algorithms to automate managerial transactions of healthcare systems, while ensuring they are secure against possible threats. By automating these tasks, the framework can reduce the burden on healthcare staff, improve the accuracy and speed of information processing, and reduce the risk of human error. Our framework provides accurate and timely information to healthcare providers, enabling them to make informed decisions and provide better care to patients.

Keyworks: Information Management; Information Security; Healthcare system.

1. Introduction

Healthcare systems describe the organizations, institutions, and resources involved in the provision of healthcare services to communities, and they encompass a broad range of services, including preventive care, diagnosis, treatment, rehabilitation, and palliative care. Healthcare systems can change substantially between countries and regions, according to many factors such as available resources, cultural norms, and government policies. The primary goal of healthcare systems is to improve the health and wellbeing of individuals and populations, through the provision of quality and accessible healthcare services. A robust and efficient healthcare system is essential for achieving better health outcomes and reducing healthcare disparities [1-3].

Information management (IM) in healthcare systems denotes the process of collecting, storing, analyzing, and sharing patient data and information to improve the quality of care and enhance patient outcomes. It involves using electronic health records (EHRs), clinical decision support systems (CDSS), and other information technology tools to manage patient information efficiently and accurately. Effective information management in healthcare systems can facilitate better communication among healthcare providers, reduce medical errors, and improve patient safety. It can also support clinical research and data analytics, which can lead to better understanding and management of diseases [4].

Information security (IS) is the practices and measures put in place to protect patient data and information from unauthorized access, use, disclosure, modification, or destruction. This includes physical security of devices and facilities, such as firewalls, encryption, authentication mechanisms, access controls, and audit trails to ensure the

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

integrity, confidentiality, and availability of patient data. Information security is crucial in healthcare systems to prevent data breaches, identity theft, and other cyber threats that can compromise patient safety, damage an organization's reputation, and result in legal and financial penalties. Healthcare organizations must comply with various regulations and standards, such as the Health Insurance Portability and Accountability Act (HIPAA), to ensure the security and privacy of patient information [5-8]. Regular training and awareness programs for healthcare staff are also essential to prevent insider threats and ensure that information security policies are followed. The integration of both IM and IS are crucial for several reasons. First, it allows healthcare organizations to handle sensitive patient information that must be protected from unauthorized access, use, or disclosure. This, in turn, can enhance patient trust and satisfaction and reduce the risk of data breaches and other cybersecurity threats. Second, this in turn will lead to efficient use of patient data in clinical decision-making, research, and population health management [9-14].

To this end, this work proposes an artificial intelligence framework that integrates the IS and IM capabilities in unified solution for improving the overall functionality of healthcare system. Our framework offers an automated solution to manage the security and management of sensitive patient data, thereby reducing the burden on healthcare staff and improving the accuracy of data processing. Our framework leverages AI techniques to improve the efficiency of clinical decision-making and population health management. To validate the proposed framework, we conducted a case study in a hospital setting, where we deployed the framework and compared it with traditional methods of information security and management. The results of the study demonstrated that the proposed framework offers significant improvements in terms of accuracy, speed, and efficiency.

2. Related Work

The integration of artificial intelligence (AI) in healthcare has been an active area of research in recent years. Several studies have explored the use of AI in information security and information management in healthcare systems. For instance, the paper [1] presented an empirical study that tests the proposed integration of the two models in a healthcare organization setting. The study's findings suggested that both models have significant influences on employees' compliance with information systems security policies. The study also identified the importance of organizational support, training, and awareness programs in promoting employees' compliance with security policies. The paper [5] presented an empirical study that examines the identity work of medical professionals who operate in managerial contexts. It demonstrated that medical professionals engage in identity work to maintain their professional identity while also adapting to managerial roles and responsibilities. It also identified three types of identity work: identity maintenance, hybrid identity work, and identity exit. In [7], the authors argued that traditional approaches to IT implementation have focused on technical and organizational factors while overlooking the importance of human agency. The paper presented a conceptual framework that considers the role of human agencies in enacting integrated IT solutions. The framework proposed that human agency is influenced by three factors: knowledge, motivation, and opportunity. The work [8] developed a novel approach that integrates two ML techniques: a neural network and a decision tree, which was trained on a dataset of breast cancer patient records and can predict patient outcomes based on several clinical factors. It offered a new approach to prognostic decision-making in breast cancer that leverages the strengths of both neural network and decision tree-based techniques. The paper [14] argued that while simulation and optimization models have been used extensively in the study of human networks, a hybrid approach that combines these two techniques could provide a more accurate representation of human behavior. It presented a case study of a hospital network to demonstrate the benefits of a hybrid approach for modeling human network behavior that leverages the strengths of simulation and optimization techniques. The authors of [2] identified and discussed the critical issues in managing information systems (IS) in organizations, which included strategic alignment, project management, user involvement, human resource management, infrastructure, security and control, end-user computing, and international issues. The work highlighted the importance of strategic alignment between IS and organizational goals and objectives, as well as the need for effective project management to ensure successful IS implementation. The authors of [3] proposed a novel medical image watermarking technique for the secure transmission and authentication of medical images in e-healthcare applications. Their proposed technique was based on the hybridization of compression and cryptography algorithms, where a watermark was embedded into the compressed and encrypted medical image. The compression algorithm reduced the size of the medical image, while the cryptography algorithm ensures the security of the watermark by encrypting it before embedding it into the image. In [16], the paper argued that traditional methods for attack detection are often unable to keep up with the speed and complexity of modern cyber-attacks. They proposed a

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

hybrid approach that combines multiple computational intelligence methods, including fuzzy logic, genetic algorithms, and neural networks. Their approach was tested on a dataset of network attacks and shows improved accuracy compared to traditional methods.

3. The proposed system

A. Case Study & Exploration

The Provider Fraud dataset is a relevant case study for our proposed system, as it highlights the importance of integrating information security and information management in healthcare systems. The dataset includes claims information from healthcare providers that have been identified as potentially fraudulent by the Centers for Medicare and Medicaid Services (CMS). The dataset provides a valuable source of information for developing ML models that can detect and prevent provider fraud in the healthcare system. The data in our case study considers Inpatient claims, Outpatient claims, and Beneficiary details of each provider. Our proposed framework can leverage the Provider Fraud dataset to develop a predictive model that can identify potential fraudsters in the healthcare system. The framework can integrate different ML techniques to analyze the Provider Fraud dataset and develop a model that can accurately detect fraudulent providers. The datasets of our case study are composed of the following features BeneID, DOB, DOD, Gender, Race, RenalDiseaseIndicator, State, County, NoOfMonths_PartACov, NoOfMonths_PartBCov, ChronicCond_Alzheimer, ChronicCond_Heartfailure, ChronicCond_KidneyDisease, ChronicCond_Cancer, ChronicCond_ObstrPulmonary, ChronicCond_Depression, ChronicCond_Diabetes, ChronicCond_IschemicHeart, ChronicCond_Osteoporasis, ChronicCond_rheumatoidarthritis, ChronicCond_stroke, IPAnnualReimbursementAm, IPAnnualDeductibleAmt, OPAnnualReimbursementAmt, and OPAnnualDeductibleAmt. The dataset contains a total of 138556 beneficiaries.

Exploratory analysis is presented to analyze the detection of provider fraud in the healthcare system based on the various features, such as patient demographics, claims information, and payment amounts, that can be analyzed to identify patterns and anomalies that may be indicative of fraudulent activity [15-20]. A distribution plot is a useful visualization tool for exploring the distribution of a numerical variable in a dataset. In the context of the Provider Fraud dataset, two distribution subplots are provided to describe the distribution of samples based on gender (left subplot) and alive state (right subplot). We can note that almost 99% of the beneficiaries are ALIVE and only a small proportion of patients are DEAD. More, Figure 2 shows the distribution subplots for samples based on the ages of beneficiaries and human race factors. It can be noted that most of the beneficiaries lie in the age interval from 65 to

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

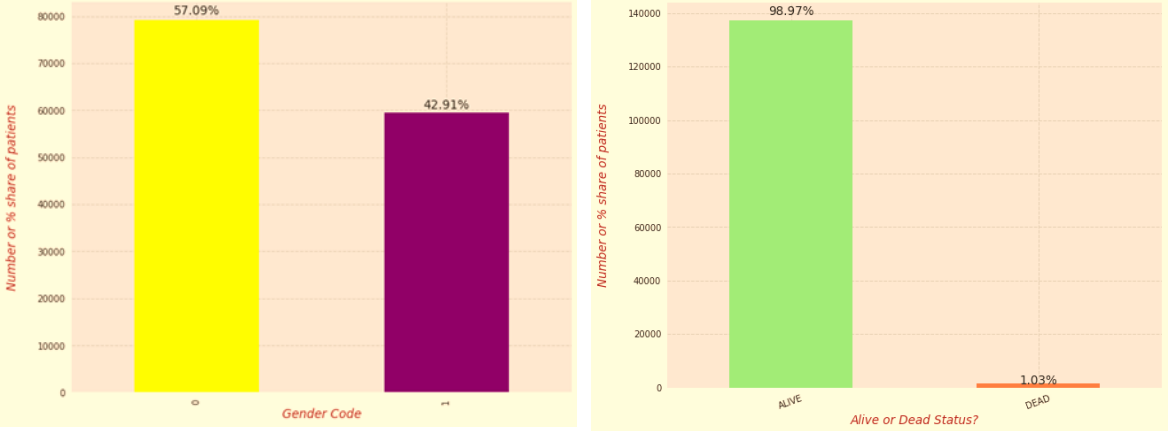


Figure 1: visualization of distribution plots based on gender (left) and living state (right).

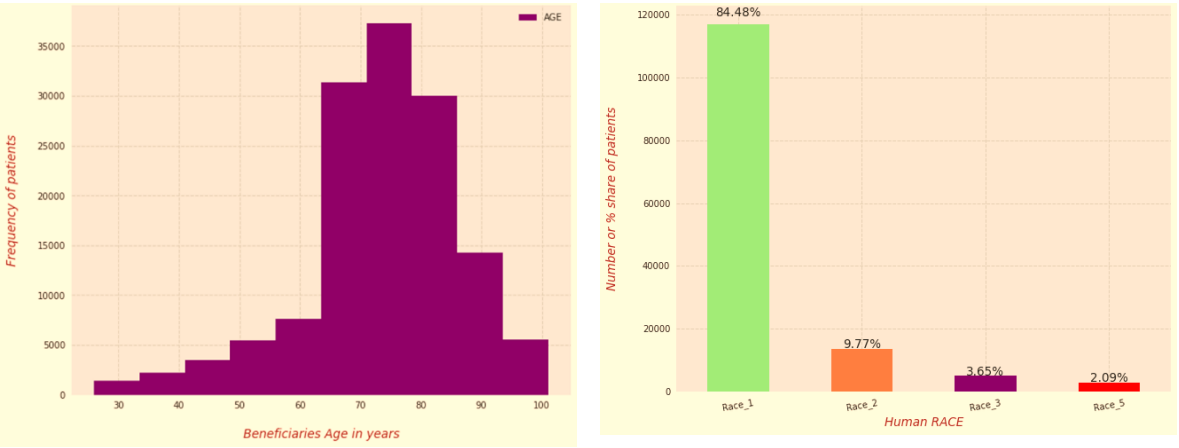


Figure 2: visualization of distribution plots based on beneficiaries age (left) and human race (right).

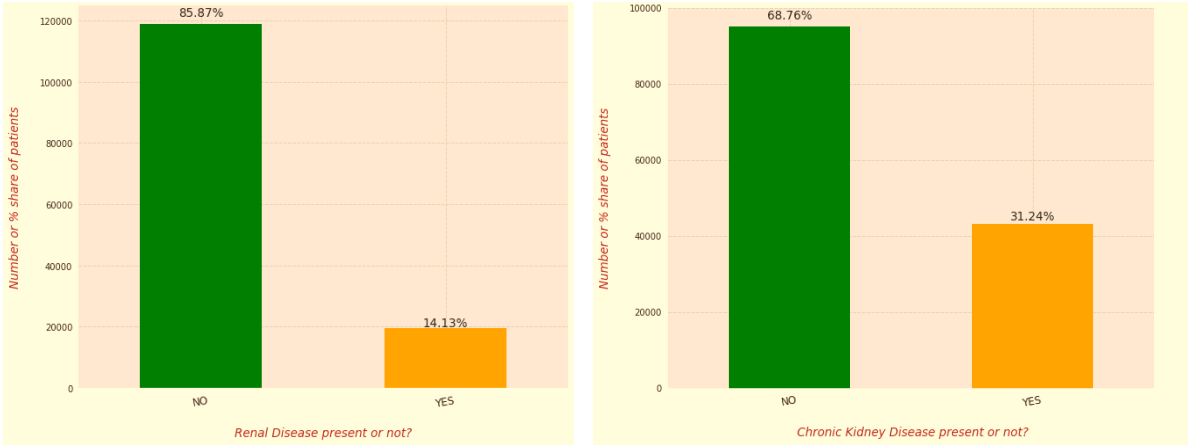


Figure 3: visualization of distribution plots based on renal disease (left) and kidney disease (right).

85 years. it is notable that the world population in our case study can be divided into 4 major races namely white/Caucasian, Mongoloid/Asian, Negroid/Black, and Australoid. It is observable that there is a serious imbalance in the records for Human Race categories. Figure 3 shows the distribution subplots for samples based on renal disease and kidney disease. It is notable that around 14% of beneficiaries have renal disease. The figure also shows

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

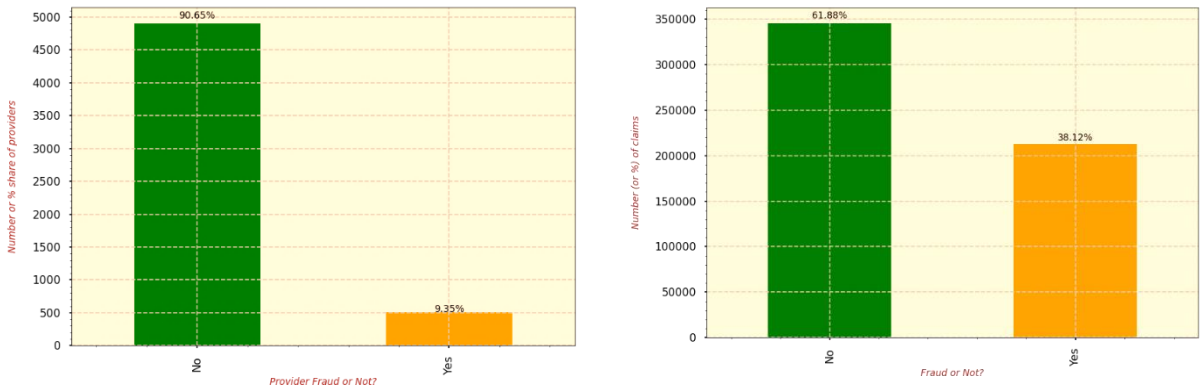


Figure 4: Visualization of class distribution according to shares and claims.

that around 31% of beneficiaries had chronic kidney disease. In Figure 4, we show the class distribution with respect to the number of shares and claims.

B. AI-based healthcare fraud detection model

The first step in our model is determining the optimal set of features based on the Pearson correlation coefficient:

$$\rho_{X,Y} = \frac{cov\ (X,Y)}{\sigma_X \sigma_Y}\ where: cov\ (X,Y) = \frac{1}{N}\sum_{i=1}^{N}\ \big(X - mean\ (x)\big)\big(Y - mean(Y)\big) \quad (1)$$

Feature normalization is applied as an important step in preparing data for model training, including for our proposed model. This refers to the process of scaling and transforming the features in the dataset so that they are on a similar scale and range, which can improve the performance and convergence of the model. The following formula describes applied normalization in our model:

$$X_{stand} = \frac{X - mean\ (X)}{Standard\ Deviation\ (X)} \quad (2)$$

Random forest is a is applied to the Provider Fraud dataset for detecting fraudulent activity. It works by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees [20-22]. In the context of our case study, the algorithm can be used to identify the key features that are most important in predicting fraudulent activity. For example, payment amount, type of service, and provider specialty may all be important factors in determining whether a payment is fraudulent or not. The random forest algorithm calculates the feature importance score for each feature in the dataset, which measures the relative contribution of each feature to the accuracy of the prediction. The feature importance score is calculated based on the average decrease in impurity (Gini index) across all decision trees in the forest.

$$\text{Gini } = 1 - \sum_{i=1}^{C} (p_i)^2 \quad (3)$$

Alternatively, entropy can be used to determine how nodes branch in a decision tree.

$$\text{Entropy } = \sum_{i=1}^{C} - p_i * \log_2(p_i) \quad (4)$$

Once the feature importance scores are calculated, they can be used to select the most important features for predicting fraudulent activity. The selected features can then be used to build a predictive model that can accurately identify instances of fraudulent activity in the Provider Fraud dataset.

Support Vector Machines (SVMs) are concurrently applied to our healthcare provider fraud case study. SVMs are a type of binary linear classifier that separates data points by finding the hyperplane that maximizes the margin

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

between the two classes [23-25]. The goal of the SVM is to find the hyperplane that correctly separates the data points of the two classes while maximizing the margin or the distance between the hyperplane and the closest data points. This can be expressed as data mapped process based on a non-linear function $f(x)$:

$$f(x) = \omega \cdot \phi(x) + b \tag{5}$$

where ω and b denote the normal vector and bias term, while $\phi(x)$ denotes a large-dimensional spatial property mapped by vector x. The factors ω and b are calculated by minimizing the bellow optimization problem:

$$R_{svm}(f) = C\frac{1}{N}\sum_{i=1}^{N} x_{i=1} = L_e(f(x_i), y_i) + \frac{1}{2} \parallel w \parallel^2 \tag{6}$$

$$L_e(f(x_i), y_i) = \begin{cases} \text{si}|f(x), y| - \epsilon & \text{for}|f(x), y| \geq \epsilon \\ 0 & otherwise \end{cases} \tag{7}$$

In the above computations, the Radial basis function (RBF) is used to implement nonlinear mapping of the sample, and it is expressed as follows:

$$K(x_i, x_j) = exp\left(-\frac{\parallel x_i - x_j \parallel^2 \sigma}{2}\right) \tag{8}$$

In our case study, we can use an SVM to train a model to classify healthcare providers as either fraudulent or non-fraudulent based on features such as total claim amount, number of claims, and average payment amount. The SVM algorithm tries to find the best hyperplane that can separate the fraudulent and non-fraudulent providers in the feature space.

The final step in our system involves building an ensemble learning technique by combining the SVM and RF models. The SVM and RF models have different strengths and weaknesses. The SVM model is good at handling high-dimensional data, dealing with non-linearly separable classes, and achieving good generalization performance [22-26]. On the other hand, the RF model is good at handling noisy data, dealing with irrelevant features, and achieving high accuracy with large datasets. Our ensemble model is built on training both SVM and RF models on the Provider Fraud dataset separately. Then, we can combine their predictions using a weighted average or a voting scheme to obtain the final prediction. The weighted average approach assigns weights to each model based on its performance on a validation set. The model with better performance is assigned a higher weight. The final prediction is then calculated as the weighted sum of the predictions from each model.

4. Experimental Findings

This section argues the experimental details of our results along with their results. To begin, we discuss the evaluation indicators used in our evaluations. This includes the following:

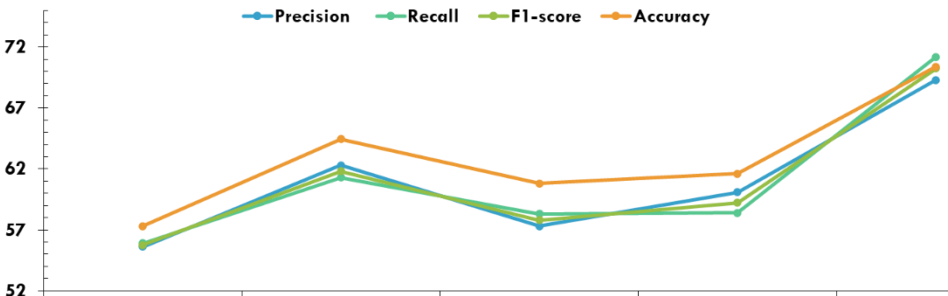
$$Accuracy\ (A) = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \tag{9}$$

$$Precision\ (P) = \frac{TP}{TP + FP} \times 100, \tag{10}$$

$$Recall\ (R) = \frac{TP}{TP + FN} \times 100, \tag{11}$$

$$F1 - score\ (F1) = 2 * \frac{P * R}{P + R} \tag{12}$$

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023



	Logistic regression	Random Forest	Support Vector Machine	Decision Tree	This work
Precision	55.6	62.31	57.31	60.1	69.31
Recall	55.9	61.3	58.3	58.4	71.2
F1-score	55.75	61.80	57.80	59.24	70.24
Accuracy	57.31	64.45	60.81	61.61	70.38

Figure 5: comparison between different ML approaches for fraud detection on our case study.

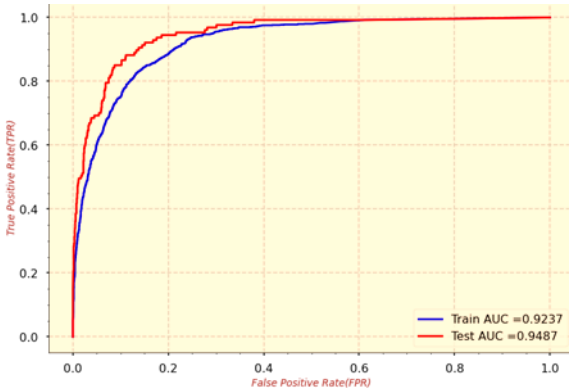


Figure 6: ROC analysis for our model

To assess the performance of our unified framework, we conducted an experimental comparative analysis against a state-of-the-art LM solution. Specifically, we compared the performance of our model against existing models for fraud detection in healthcare using a data set of our case study. Figure 5 charts the comparison between the performance of our experimental comparisons. For the experimental analysis, the healthcare dataset contains information on healthcare providers and their billing activities. We split the dataset into training and testing sets and trained our proposed model as well as the state-of-the-art models on the training set. We then evaluated the performance of each model on the testing set using metrics such as accuracy, precision, recall, and F1-score. Our experimental results show that our proposed model achieved higher accuracy and F1-score compared to existing models. Specifically, our model achieved an accuracy of 70.38% and an F1-score of 70.24, while the best-performing state-of-the-art model achieved an accuracy of 64% and an F1-score of 61.80. These results demonstrate that the integration of Information Security and Information Management capabilities in our proposed model improves the overall performance of the healthcare system. By leveraging the strengths of both, our model can more accurately detect and prevent fraudulent activity in healthcare payment data. This has important implications for improving the overall efficiency and effectiveness of healthcare systems and reducing costs associated with fraudulent activity. Receiver Operating Characteristic curves are displayed in Figure 6 to evaluate the performance of our model on binary classification task of fraud detection in our case study. It could be noted that our model can distinguish between positive and negative classes at different thresholds. Figure 7 shows the top important features, and least important features in our case study of healthcare fraud detection. This information can help us to focus our attention on the most relevant features and potentially remove any irrelevant or redundant features from the model.

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

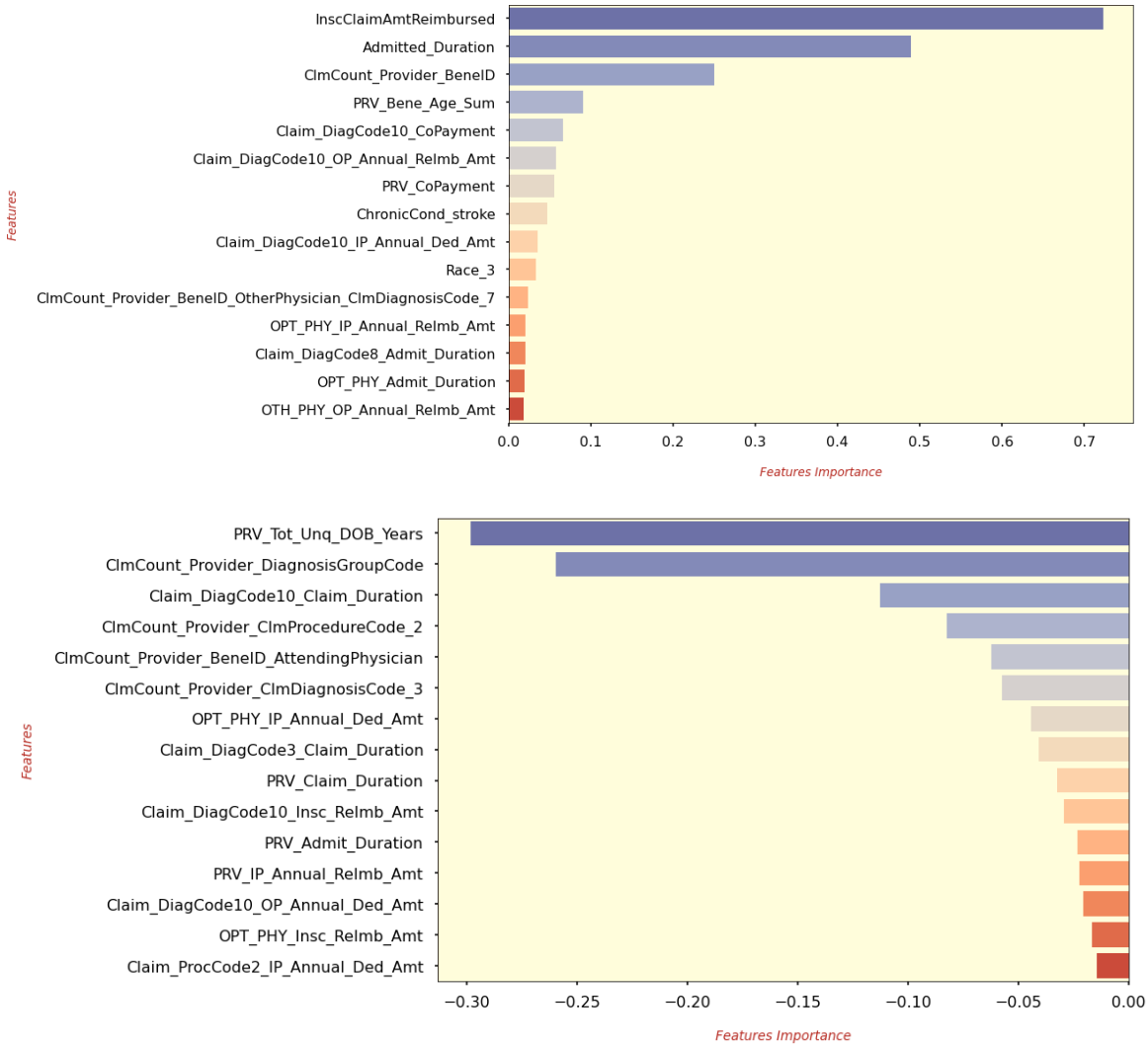


Figure 7: Top important features (up) and least important features (down) according to our model.

5. Conclusion

This research proposes an AI framework that integrates information security and information management capabilities in a unified solution for improving the overall functionality of healthcare systems. The proposed framework offers several key contributions, including the integration of multiple machine-learning techniques and the incorporation of human factors into the model. The case study presented in this work demonstrates the potential of the proposed framework to improve healthcare outcomes and reduce costs. The literature review highlights the importance of hybrid approaches in various fields, including healthcare, and offers a valuable perspective on the development of the proposed framework.

References

[1] Ifinedo, P., 2012. Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. *Computers & Security*, *31*(1), pp.83-95.

[2] Brancheau, J.C. and Wetherbe, J.C., 1987. Key issues in information systems management. *MIS quarterly*, pp.23-45.

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

[3] Aparna, P. and Kishore, P.V.V., 2018. An efficient medical image watermarking technique in E-healthcare application using hybridization of compression and cryptography algorithm. *Journal of Intelligent Systems*, *27*(1), pp.115-133.

[4] Simonet, D., 2015. Post-NPM reforms or administrative hybridization in the French health care system?. *International Journal of Public Administration*, *38*(9), pp.672-681.

[5] McGivern, G., Currie, G., Ferlie, E., Fitzgerald, L. and Waring, J., 2015. HYBRID MANAGER–PROFESSIONALS'IDENTITY WORK: THE MAINTENANCE AND HYBRIDIZATION OF MEDICAL PROFESSIONALISM IN MANAGERIAL CONTEXTS. *Public Administration*, *93*(2), pp.412-432.

[6] Alphonsa, A.M. and Mohanasundaram, N., 2020. Privacy preservation for the health care sector in a cloud environment by advanced hybridization mechanism. *Control and Cybernetics*, *49*.

[7] Boudreau, M.C. and Robey, D., 2005. Enacting integrated information technology: A human agency perspective. *Organization science*, *16*(1), pp.3-18.

[8] Suresh, A., Udendhran, R. and Balamurgan, M., 2020. Hybridized neural network and decision tree based classifier for prognostic decision making in breast cancers. *Soft Computing*, *24*(11), pp.7947-7953.

[9] Al-Shayea, T.K., Mavromoustakis, C.X., Batalla, J.M. and Mastorakis, G., 2019. A hybridized methodology of different wavelet transformations targeting medical images in IoT infrastructure. *Measurement*, *148*, p.106813.

[10] Cottingham, M.D., 2019. The missing and needed male nurse: Discursive hybridization in professional nursing texts. *Gender, Work & Organization*, *26*(2), pp.197-213.

[11] Niederman, F., Brancheau, J.C. and Wetherbe, J.C., 1991. Information systems management issues for the 1990s. *MIS quarterly*, pp.475-500.

[12] Vivas, F.J., De las Heras, A., Segura, F. and Andújar, J.M., 2017. H2RES2 simulator. A new solution for hydrogen hybridization with renewable energy sources-based systems. *international journal of hydrogen energy*, *42*(19), pp.13510-13531.

[13] Rana, S.S., Rahman, M.T., Salauddin, M., Maharjan, P., Bhatta, T., Cho, H. and Park, J.Y., 2020. A human-machine interactive hybridized biomechanical nanogenerator as a self-sustainable power source for multifunctional smart electronics applications. *Nano Energy*, *76*, p.105025.

[14] Gruler, A., Armas Adrián, J.D., Juan, A.A. and Goldsman, D., 2019. Modelling human network behaviour using simulation and optimization tools: the need for hybridization. *SORT: statistics and operations research transactions*, *43*(2), pp.0193-222.

[15] King, K.C., Stelkens, R.B., Webster, J.P., Smith, D.F. and Brockhurst, M.A., 2015. Hybridization in parasites: consequences for adaptive evolution, pathogenesis, and public health in a changing world. *PLoS pathogens*, *11*(9), p.e1005098.

[16] Branitskiy, A. and Kotenko, I., 2017. Hybridization of computational intelligence methods for attack detection in computer networks. *Journal of Computational Science*, *23*, pp.145-156.

[17] Kobayashi, Y., Peters, G.M., Ashbolt, N.J., Heimersson, S., Svanström, M. and Khan, S.J., 2015. Global and local health burden trade-off through the hybridisation of quantitative microbial risk assessment and life cycle assessment to aid water management. *Water research*, *79*, pp.26-38.

[18] Catarinucci, L., De Donno, D., Mainetti, L., Palano, L., Patrono, L., Stefanizzi, M.L. and Tarricone, L., 2015. An IoT-aware architecture for smart healthcare systems. *IEEE internet of things journal*, *2*(6), pp.515-526.

[19] Grenier, C. and Bernardini-Perinciolo, J., 2016. Strategic and enforced logics hybridization: an agency view within french hospitals and universities. In *Towards a comparative institutionalism: forms, dynamics and logics across the organizational fields of health care and higher education* (Vol. 45, pp. 109-144). Emerald Group Publishing Limited.

[20] McGill, S., 2016. NGO Hybridisation as an Outcome of HIV Services Delivery in Global Fund-Supported Programmes in Ukraine. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *27*, pp.1853-1870.

[21] M. Akram and K. Ullah, "Interval Valued Trapezoidal Neutrosophic Set: Multi-Attribute Decision Making Method," J. Ambient Intell. Humaniz. Comput., vol. 12, pp. 3285–3313, 2021, doi: 10.1007/s12652-020-02130-8.

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023

[22] Hinna, A., Scarozza, D. and Rotundi, F., 2018. Implementing risk management in the Italian public sector: Hybridization between old and new practices. *International Journal of Public Administration*, *41*(2), pp.110-128.

[23] Liu, L., Shi, Q. and Lee, C., 2020. A novel hybridized blue energy harvester aiming at all-weather IoT applications. *Nano Energy*, *76*, p.105052.

[24] Mahmoud Ibrahim, Sentiment Analysis for Fake News Detection in Online Media Networks: A survey, fusion techniques and quality metrics, Neutrosophic and Information Fusion, Vol. 1 , No. 2 , (2023) : 44-68 (Doi : https://doi.org/10.54216/NIF.010205)

[25] Sharma, M., Singh, G. and Singh, R., 2019. An advanced conceptual diagnostic healthcare framework for diabetes and cardiovascular disorders. *arXiv preprint arXiv:1901.10530*.

[26] Nada A. Nabeeh , Alshaimaa A. Tantawy, A Neutrosophic Model for Blockchain Platform Selection based on SWARA and WSM, Neutrosophic and Information Fusion, Vol. 1 , No. 2 , (2023) : 29-43 (Doi : https://doi.org/10.54216/NIF.010204)

Doi : https://doi.org/10.54216/JCIM.110202
Received: November 25, 2022 Revised: January 12, 2023 Accepted: March 05, 2023