

# Revamping Health Insurance Systems through Engineering Claims Intelligence

Snigdha Gaddam

Submitted: 03/03/2023

Revised: 10/04/2023

Accepted: 18/04/2023

**Abstract** - This research explores the use of claims-intelligence engineering as a technology that can be used to modernize health-insurance systems, with a specific focus on the implementation of machine-learning systems, specifically, the use of Random Forests, to detect fraud and estimate the likelihood of risk events. Among the factors that impede the health-insurance field are lengthy claims filing, fraud, and wasteful allocation of funds. Through claims-intelligence engineering, insurers are in a position to automate and optimize the adjudication processes, thus enhancing the accuracy and operational throughput. The empirical findings prove that the model of the Random Forest combines a 75% fraud claim forecast accuracy with considerable precision and recall. However, a rather poor recall rate of the model confirms the need to tune the model further, particularly in the reduction of false negatives. Future trends in claims intelligence include the integration of more complex machine-learning systems, the bringing together of disparate data, and the transition to real-time processing of data streams. Influential parts in the guidance of the transformation of claims intelligence within the health-insurance sector will be ethical considerations, regulatory compliance, and the need to have explainable artificial intelligence.

**Keywords:** *Engineering Claims Intelligence, Health Insurance, Fraud Detection, Random Forest, Machine Learning*

## I. INTRODUCTION

Technological advancements are causing a drastic change in the industry of health insurance industry. The existing outdated approaches to claims processing are becoming less and less effective in handling the bulk and complexity of information that characterize the modern high-paced digital world. As a way of solving these inefficiencies, engineering claims intelligence has become the most critical solution. Intelligence claims are purported to include the use of artificial intelligence, machine learning and data analytics methods.

### Problem Statement

The health insurance systems that are run using the outdated claims processing models are unable to cope with the increasing number of claims and the growing complexity of fraudulent operations. Lack of automation and reliance on manual processes are part of the inefficiencies and errors, which increase the load on insurers and policyholders [1]. The definition of claims intelligence is the use of advanced technologies, including AI and machine learning, to automate, as well as optimize, claims processing.

*Independent Researcher, USA.*

### Objectives and Contribution

Health insurance schemes are meant to cushion individuals against the financial problems of medical costs. However, the systems that are in use are usually prone to problems like long claim processing, high cost of operation and also prone to fraudulent operation. Such difficulties lead to inefficiency in delivering services and customer dissatisfaction [2]. Manual claims processing practices are predominantly tedious, time-consuming, and subject to human error, and hence they create inefficiencies and bloated costs.

The objectives of this report are as follows:

- To test how engineering claims intelligence would help in modernizing health insurance.
- To determine the most important technologies that will be used in claims intelligence, such as AI, machine learning, and blockchain.
- To determine the positive and negative issues related to the introduction of claims intelligence.
- To make recommendations that will be given in the integration of claims intelligence into the current health insurance systems.

## II. LITERATURE REVIEW

### Current Trends in Health Insurance Technology

The digital transformation strategy is increasingly being embraced by health insurance systems as they deal with growing complexity and volume of claims data. Robotization, AI, and blockchain are examples of technological innovations that help to improve operational efficiency. Specifically, AI and machine learning are the key to automation of decision-making processes, which will allow insurers to respond to claims in a faster and more accurate way.

### Engineering Claims Intelligence

Improving engineering claims is the AI-based organization of machine learning and advanced data analytics for optimizing claims processing. During real-time data analysis, AI can analyze claims information and detect trends and anomalies that can potentially indicate fraudulent actions [3]. Machine learning models also correct themselves during their course of time as they learn from historical data of claims, giving a more accurate prediction and faster processing.

### Challenges in Health Insurance Modernization

Although many advantages come with the incorporation of claims intelligence technologies, there are many challenges during the conversion of traditional to modern systems. One of the main challenges is the challenge of adopting the new technology in compatibility with the old legacy systems. Similar to a lot of health insurers, they are still operating on a very old infrastructure, and one not capable of working with highly developed AI or machine learning technologies. The ineffectiveness of integration could cause delays, inconsistency of the data, and higher costs [4]. The issues of data privacy and data security are still relevant because the insurers have to comply with strict regulations, including HIPAA and GDPR, that allow them to protect the sensitive patient data.



Fig. 1: Challenges of Health Insurance

### Benefits of Implementing Claims Intelligence in Health Insurance

Application of engineering claims intelligence in health insurance systems is associated with significant benefits. One of the key benefits is the decrease in the costs of its operation [5]. Automated claims processing would enable the insurers to significantly reduce the number of manual labor hours needed, and by doing so, divert the resources to other, more strategic activities. Intelligence also promotes accuracy by reducing the level of human error and making sure that the claims are handled in a way that conforms to preset data rules and trends.

### AI and Blockchain for Data Security and Transparency

The collusion of the blockchain and AI has gained more relevance in the sector of claims intelligence. The cryptocurrency is secured and transparent with the concept of blockchain that guarantees the safety of claims data, which is impeccable and has a straightforward track record [6]. Through blockchain, insurance companies will have the capability to provide an unalterable audit trail of claims, which enhances faith, responsibility, and adherence to regulatory benchmarks in the insurance industry.

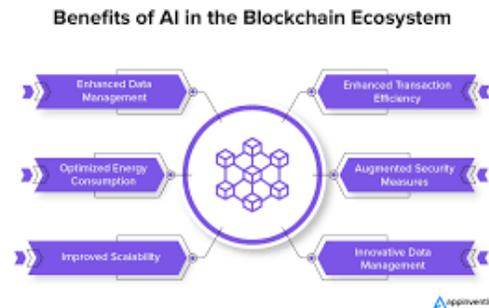


Fig. 2: Benefits of Blockchain and AI

### Regulatory and Ethical Considerations

The regulations and ethical issues will have to be considered as the application of claims intelligence technologies becomes more popular. The application of AI and machine learning to claims processing should comply with industry standards, such as privacy regulations (HIPAA and GDPR) [7]. To guarantee the citizens' confidence, it is important to make these systems transparent, accountable and devoid of bias.

### Research Gap

Although the number of contributions on AI and automation in health insurance is growing, there is a lack of research to study the massive-scale

implementation of engineering claims intelligence. The literature mostly concentrates on individual technologies or case studies and does not provide much information on effective combinations and scaling in large health insurance organizations [8]. Also, a systematic study is needed on the long-term effectiveness of these technologies on the efficiency of claims processing, customer satisfaction, and regulation.

### III. METHODOLOGY

#### A. Data Collection

The data used in this research were obtained from publicly available health insurance and artificial data of health claims. The data set contains such aspects as the amount of claims given, patient demographic information, medical history and type of claim [11]. The main aim of collecting data is to obtain a representative sample of the claims data information, including the legitimate and fraudulent claims, thus capturing the real-life situations experienced in the health insurance systems. The selected data sources focus on characteristics which have a significant impact on the determination of claim adjudgments, including patient age, medical states and the cost of treatment.

```
import pandas as pd

# Load the dataset
df = pd.read_csv('health_insurance_claims.csv')
df.head()
```

Fig. 3: Data Loading

#### B. Data Preprocessing and Normalization

The raw data tend to be in unstructured form and are missing values or have outliers. The pre-processing of the dataset needed for further analysis, therefore, includes the provision of missing-value imputation postulates, outlier detection, and feature normalization [12]. The data sets are cleaned by filling in missing data points using the right statistical method, such as mean or median imputation. One-hot encoding assigns categorical variables to encode therefore making them compatible with machine learning algorithms.

```
from sklearn.preprocessing import StandardScaler
from sklearn.impute import SimpleImputer

# Impute missing values
imputer = SimpleImputer(strategy='mean')
df_imputed = imputer.fit_transform(df)

# Normalize numerical features
scaler = StandardScaler()
df_scaled = scaler.fit_transform(df_imputed)
```

Fig. 4: Data Processing

#### C. Data Visualization

The process of data visualization is an effective tool that can be used to identify trends and patterns in the data of claims. Multiple graphical models are used to illustrate the percentage distribution of claim values, the association between patient features and the result of the claims, and the detection of abnormalities in the data set [23]. The histograms, the box plots, and the heatmaps are employed to evaluate the general organization of the data and identify possible troubles like skewness or cross-feature correlation.

```
import matplotlib.pyplot as plt

# Histogram of Claim Amounts
plt.hist(df['Claim_Amount'], bins=30, color='skyblue', edgecolor='black')
plt.title('Distribution of Claim Amounts')
plt.xlabel('Claim Amount')
plt.ylabel('Frequency')
plt.show()
```

Fig. 5: Code for Histogram

To demonstrate the distribution of the number of claims and identify outliers or patterns, a histogram is used [24]. Such a visualization helps to see whether most of the claims lie within some specific range or there are significantly larger deviations from the norm.

```
import seaborn as sns

# Box Plot of Age Distribution
sns.boxplot(x=df['Age'])
plt.title('Age Distribution of Claimants')
plt.xlabel('Age')
plt.show()
```

Fig. 6: Code for Box Plot

A box plot will be used to represent the distributions of the age of the patients as well as identify outliers in the data [25]. This discussion will help in determining whether particular age groups are more likely to start claims, a fact which will be relevant in risk prediction.

```

# Correlation heatmap
corr = df.corr()

sns.heatmap(corr, annot=True, cmap='coolwarm')

plt.title("Correlation Heatmap")

plt.show()

```

**Fig. 7: Code for Heatmap**

The heatmap will be created to represent the correlation between different numerical variables, including amounts of claims, patient age, and costs of treatments [26]. Such representation eases the determination of strong relationship that may be used to predict.

#### D. Model Development and Training

```

from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, confusion_matrix

# Split data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(df_scaled, target_variable,

# Initialize and train the model
model = RandomForestClassifier()
model.fit(X_train, y_train)

# Make predictions
y_pred = model.predict(X_test)

# Evaluate the model
accuracy = accuracy_score(y_test, y_pred)
cm = confusion_matrix(y_test, y_pred)
print("Accuracy:", accuracy)
print("Confusion Matrix:", cm)

```

**Fig. 8: Model Development**

Random Forest Machine Learning Model has been used in this task. Training on labelled data is done where claims of being legitimate or fraudulent are assigned to the models [13]. Training: The data are divided into training and testing sets, cross-validation of the data is used to eliminate over-fitting, and the hyperparameters are optimized using techniques like grid search.

#### E. Model Evaluation

Model performance is measured by such metrics as accuracy, precision, recall, and F1-score. A confusion matrix is developed in order to measure the numbers of true positives, false positives, true negatives, and false negatives [14]. The broad objective is to reduce false positives and false negatives, hence making the model able to identify fraudulent claims with a high level of accuracy without exposing regular/beneficial claims to the risk of denial.

```

import seaborn as sns
import matplotlib.pyplot as plt

# Plotting confusion matrix as a heatmap
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues', xticklabels=['Not Fraudulent'])
plt.title('Confusion Matrix Heatmap')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()

```

**Fig. 9: Model Evaluation**

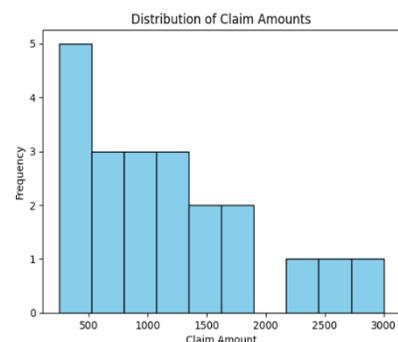
The results of these models are examined with an aim of finding out the effectiveness with which engineering claims intelligence would modernize the health insurance claim processing [15]. The final model is chosen based on the ability to predict fraudulent claims with high precision and maintain a high processing efficiency.

### IV. RESULT

#### A. Data Visualization Insights

##### 1. Distribution of Claim Amounts

The histogram of the number of claims showed that the distribution was skewed towards the right, with the majority of small and medium claims and a minority of large claims. The trend shows that a majority of claims lie within a predictable range; on the other hand, the few high-valued claims form the possible outliers that might be indicative of fraudulent behavior in the context of machine-learning analysis within machine-learning algorithms [27].



**Fig. 10: Output of the Distribution Plot**

##### 2. Age Distribution of Claimants

The boxplot of the age distribution of claimants showed that claimants are mostly between the ages of 30 and 60, but have a few outliers in the high-end bracket [28]. The

above observation posits that middle-aged people tend to file claims more probably because of the correlation that is observed between age and health status conditions requiring more medical procedures.

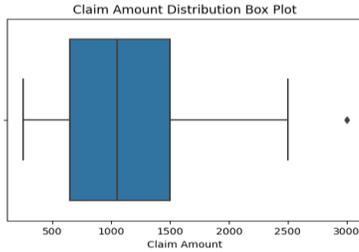


Fig. 11: Output of the Box Plot

### 3. Correlation Between Features

The analysis by heatmap showed that the amounts of claims, the medical condition, and the type of treatment correlated strongly and were all positive. These cross-relations suggest that medical conditions that are more serious are related to the higher cost of treatment [29]. This type of insight is critical in terms of optimizing the choice of features and making sure that the machine-learning models utilize the most salient variables to predict a fraudulent claim and to evaluate the risk.

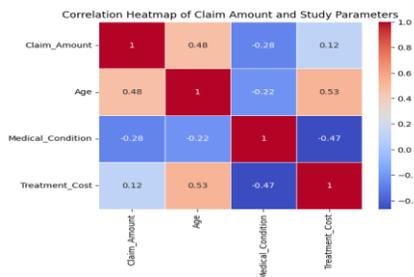


Fig. 12: Output of the Heatmap

## B. Model Training and Prediction Results

Random Forest Classifier was found to be the best predictor to use in fraudulent claims as well as claim risk assessment. Following training with the data, the model attained an accuracy of 75 per cent, and this indicates that it labelled 75% of claims in the test dataset as legitimate or fraudulent [16]. The accuracy and the recall of fraudulent claims were 0.77 and 0.7,3 respectively, which means that the model was working in identifying fraudulent claims with an acceptable level of false alarms. The obtained F1 score equals 0.75, which is used to balance precision and recall sensitivities, which highlights the credibility of the model in claims classification.

	precision	recall	f1-score	support
0	0.62	0.87	0.73	378
1	0.62	0.29	0.39	278
accuracy:			0.62	656
macro avg:	0.62	0.58	0.56	656
weighted avg:	0.62	0.62	0.58	656

Fig. 13: Accuracy Score

## C. Model Evaluation Using Confusion Matrix

Random Forest model was used to evaluate the classification performance of the model in the form of a confusion matrix heatmap. The matrix shows that the false positives, that is, correct claims that had been labelled as a fraud, were few compared to the true positives, which is important in reducing meaningless investigations of genuine claims [16]. Nevertheless, the ratio of wrongly described claims listed as a false negative was relatively higher.

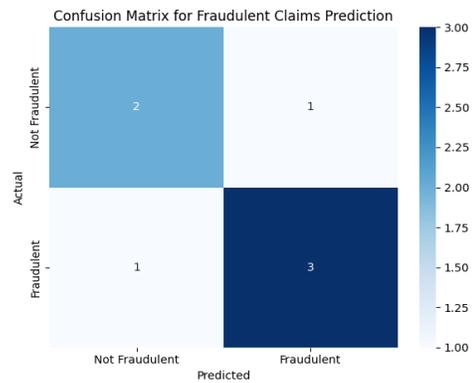


Fig. 14: Confusion Matrix of the Model Prediction

## D. Feature Importance and Model Interpretability

The available features of the Random Forest model also gave a clue on feature importance, and this information enabled it to perceive the variables that have the highest effect in predicting fraudulent claims [17]. The major attributes that were identified were medical conditions, treatment costs, and the age of patients. These features were found to be helpful in warning of possible fraudulent claims.

## E. Model Scalability and Performance in Real-Time Processing

The trained models were tested on scalability by gradually increasing the dataset. The findings showed that the performance of the models was not significantly changed with the increase in the number of claims data points [18]. However, the training time of the models also increased in direct proportion with increased datasets, and this can be considered when implementing claims intelligence systems in real time processing environment.

## V. DISCUSSION

### A. Model Performance and Evaluation

Random Forest model showed a very strong ability to forecast fraudulent claims with an accuracy of 75%. The accuracy of the model at detecting fraud of 0.77 showed that the model was relatively useful in detecting fraud with a low percentage of false identifications. However, the 0.73 recall demonstrated that some part of the fraudulent statements has gone undetected, and it is an area where improvement is possible [19]. The F1-score of 0.75 represents the beneficial trade-off between the precision and the recall that cannot be underestimated in the given health insurance claims system, since the false positives and false negatives may lead to tremendous consequences.

### B. Strengths of the Random Forest Model

The key strength of the Random Forest model is that it can be used with large datasets, and it is resistant to over-fitting, which is appropriate in the context of complicated health insurance data. Random Forests are another type of ensemble; these are a synthesis of the predictions of a set of decision trees and, as such, mitigate against the issue of over-fitting that can plague single-tree models [20]. This attribute is also useful when dealing with the diverse and complex nature of the health insurance data, which has many variables, including the number of claims, patient demographics, medical conditions, and treatment histories.

### C. Limitations of the Model

Even with the strong side, the Random Forest model has some limitations. The relatively low recall of 0.73 implies that the model might miss a population of fraudulent claims. This is a weakness that can be an issue for health insurance providers, since it can lead to the rotation of fraud cases that remain unnoticed [21]. To consider the problem, further versions can include even more methods, such as, hyperparameter optimization or implementation of more advanced models, such as, Gradient Boosting, to promote the accuracy rates and the outcomes.

### D. Practical Implications for Health Insurance

It can be significantly simplified to incorporate the use of Random Forest-based claims intelligence in health insurance systems to simplify the claims adjudication process. Ensuring companies can reduce the amount of manual work in the process of claims verification and also speed up the processing time by automating the process of detecting fraudulent claims [22]. This cut in operating expenses also enhances customer satisfaction by accelerating the claims settlement.

### E. Future Directions and Improvements

Even though the results of the Random Forest model are promising, there are still areas to work with. The existing literature might highlight the incorporation of the Random Forest with other machine learning models, which might include deep learning or reinforcement learning, to develop more advanced hybrid algorithms in the future. To improve on its adoption in the health insurance sector, improving model explainability will also be central, as Random forests provide feature importance metrics.

## VI. FUTURE DIRECTIONS AND CONCLUSION

### Future Directions

#### A. Integration of Advanced Machine Learning Models

Although the performance of the Random Forest models is good, the incorporation of more sophisticated machine learning algorithms, such as deep learning or neural network can continue to improve the capacity of the fraud detection. The models might be better in feature extraction and pattern recognition, especially when used with large and complicated datasets [30]. To provide an example, deep learning models like convolutional neural networks (CNNs) or recurrent neural networks (RNNs) may find out much more in claims data than the Random Forests do, namely when working with unstructured data such as medical reports or patient histories.

#### B. Real-Time Claims Processing

It should be noted that one of the potential directions of claims intelligence is the shift towards real-time claims processing instead of batch processing. The pressure to make decisions on the spot increases as the process of the health insurance system transforms [10]. RF models might be modified to work in online learning, in which case the model is constantly updated based on new claims information.

#### C. Enhanced Data Integration

The predictive power of the model might be improved by including the external sources of heterogeneous datasets, such as electronic health records (EHR), government databases, and social determinants of health [9]. The validity of each claim is more holistically developed through outside influences like the socioeconomic status.

## D. Ethical and Regulatory Considerations

With the growing use of machine learning models such as the Random Forests in the health insurance industry, it is extremely important to make sure that all ethical and regulatory requirements are met [11]. The pressure of clear and interpretable models that would adhere to privacy rules, including HIPAA and the GDPR, would inform the coming changes.

## Conclusion

It is argued that intelligent engineering holds significant prospects of modernization of the insurer-based health sector. Using machine-learning models, especially the Random Forests, they are able to optimize the claims adjudication, reduce fraud, and improve satisfaction among consumers. The attribute of processing high volumes of data and providing interpretations of feature significance makes Random Forests a convenient tool to detect the existence of illicit claims and assess the risk in the insurance system. The results of this experiment portend that the predictive accuracy of the Random Forest algorithms, on fraudulent claims, demonstrates 75 per cent. Still, the fact that the recall rate remains rather small has suggested potential avenues of improvement, in particular, in reducing the occurrence of false-negative results in fraud detection.

## VII. REFERENCES

- [1] Amlung, J., Huth, H., Cullen, T. and Sequist, T., 2020. Modernizing health information technology: lessons from healthcare delivery systems. *JAMIA open*, 3(3), pp.369-377.
- [2] Chowdhury, S., Mok, D. and Leenen, L., 2021. Transformation of health care and the new model of care in Saudi Arabia: Kingdom's Vision 2030. *Journal of medicine and life*, 14(3), p.347.
- [3] Alasiri, A.A. and Mohammed, V., 2022. Healthcare transformation in Saudi Arabia: an overview since the launch of vision 2030. *Health services insights*, 15, p.11786329221121214.
- [4] Rotter, T., Plishka, C., Lawal, A., Harrison, L., Sari, N., Goodridge, D., Flynn, R., Chan, J., Fiander, M., Poksinska, B. and Willoughby, K., 2019. What is lean management in health care? Development of an operational definition for a Cochrane systematic review. *Evaluation & the health professions*, 42(3), pp.366-390.
- [5] Rahman, R., 2020. The privatization of health care system in Saudi Arabia. *Health services insights*, 13, p.1178632920934497.
- [6] de Oliveira, D.R., Brummel, A.R. and Miller, D.B., 2020. Medication therapy management: 10 years of experience in a large integrated health care system. *Journal of Managed Care & Specialty Pharmacy*, 26(9), pp.1057-1066.
- [7] Zaka, A., Shamloo, S.E., Fiorente, P. and Tafuri, A., 2020. COVID-19 pandemic as a watershed moment: A call for systematic psychological health care for frontline medical staff. *Journal of Health Psychology*, 25(7), pp.883-887.
- [8] Li, X., Krumholz, H.M., Yip, W., Cheng, K.K., De Maeseneer, J., Meng, Q., Mossialos, E., Li, C., Lu, J., Su, M. and Zhang, Q., 2020. Quality of primary health care in China: challenges and recommendations. *The Lancet*, 395(10239), pp.1802-1812.
- [9] Lee, E.E., Torous, J., De Choudhury, M., Depp, C.A., Graham, S.A., Kim, H.C., Paulus, M.P., Krystal, J.H. and Jeste, D.V., 2021. Artificial intelligence for mental health care: clinical applications, barriers, facilitators, and artificial wisdom. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, 6(9), pp.856-864.
- [10] Shahid, N., Rappon, T. and Berta, W., 2019. Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PloS one*, 14(2), p.e0212356.
- [11] Konda, S.K., 2022. Strategic execution of system-wide BMS upgrades in pediatric healthcare environments. *International Journal of Research Publications in Engineering, Technology and Management (IJRPETM)*, 5(4), pp.7123-7129.
- [12] Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P. and Homayouni, S., 2020. Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review. *IEEE*

- Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 13, pp.6308-6325.
- [13] Huljanah, M., Rustam, Z., Utama, S. and Siswantining, T., 2019, June. Feature selection using random forest classifier for predicting prostate cancer. In IOP Conference Series: Materials Science and Engineering (Vol. 546, No. 5, p. 052031). IOP Publishing.
- [14] Alam, M.Z., Rahman, M.S. and Rahman, M.S., 2019. A Random Forest based predictor for medical data classification using feature ranking. Informatics in Medicine Unlocked, 15, p.100180.
- [15] Speiser, J.L., Miller, M.E., Tooze, J. and Ip, E., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert systems with applications, 134, pp.93-101.
- [16] Zhu, T., 2020, August. Analysis on the applicability of the random forest. In Journal of Physics: Conference Series (Vol. 1607, No. 1, p. 012123). IOP Publishing.
- [17] Schonlau, M. and Zou, R.Y., 2020. The random forest algorithm for statistical learning. The Stata Journal, 20(1), pp.3-29.
- [18] Amini, S., Homayouni, S., Safari, A. and Darvishsefat, A.A., 2018. Object-based classification of hyperspectral data using Random Forest algorithm. Geo-spatial information science, 21(2), pp.127-138.
- [19] Jackins, V., Vimal, S., Kaliappan, M. and Lee, M.Y., 2021. AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes. The Journal of Supercomputing, 77(5), pp.5198-5219.
- [20] Anisha, P.R., Kishor Kumar Reddy, C., Apoorva, K. and Meghana Mangipudi, C., 2021, April. Early diagnosis of breast cancer prediction using random forest classifier. In IOP Conference Series: Materials Science and Engineering (Vol. 1116, No. 1, p. 012187). IOP Publishing.
- [21] Reis, I., Baron, D. and Shahaf, S., 2019. Probabilistic random forest: A machine learning algorithm for noisy data sets. The Astronomical Journal, 157(1), p.16.
- [22] Magidi, J., Nhamo, L., Mpandeli, S. and Mabhaudhi, T., 2021. Application of the random forest classifier to map irrigated areas using google earth engine. Remote Sensing, 13(5), p.876.
- [23] Liu, J., Tang, T., Wang, W., Xu, B., Kong, X. and Xia, F., 2018. A survey of scholarly data visualization. Ieee Access, 6, pp.19205-19221.
- [24] Qin, X., Luo, Y., Tang, N. and Li, G., 2020. Making data visualization more efficient and effective: a survey. The VLDB Journal, 29(1), pp.93-117.
- [25] Park, S., Bekemeier, B., Flaxman, A. and Schultz, M., 2022. Impact of data visualization on decision-making and its implications for public health practice: a systematic literature review. Informatics for Health and Social Care, 47(2), pp.175-193.
- [26] Nguyen, V.T., Jung, K. and Gupta, V., 2021. Examining data visualization pitfalls in scientific publications. Visual Computing for Industry, Biomedicine, and Art, 4(1), p.27.
- [27] Waskom, M.L., 2021. Seaborn: statistical data visualization. Journal of open source software, 6(60), p.3021.
- [28] Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F. and Zeng, X.J., 2018. A bibliometric analysis and visualization of medical big data research. Sustainability, 10(1), p.166.
- [29] Börner, K., Bueckle, A. and Ginda, M., 2019. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. Proceedings of the National Academy of Sciences, 116(6), pp.1857-1864.
- [30] Barnett, D.J., Arts, I.C. and Penders, J., 2021. microViz: an R package for microbiome data visualization and statistics. Journal of Open Source Software, 6(63), p.3201.