Unveiling Insights: Exploring Healthcare Data through Data Analysis

Amar Deep Gupta
Department of Computer Science
Amity University Greater Noida
adgupta@gn.amity.edu

Karam Singh
Department of Computer Science
Amity University Greater Noida
karamsinghrainu@gmail.com

Upendra Pratap Pandey
Department of Computer Science
Delhi Technical Campus, Greater Noida
uppandey998@gmail.com

Pradeep Kumar Kushwaha
Department of Computer Science
Amity University Greater Noida
pkkushwaha@gn.amity.edu

Bhanu Prakash Lohani
Department of Computer Science
Amity University Greater Noida
bplohani@gn.amity.edu

Shitij Kumar
Department of Computer Science
Amity University Greater Noida
shitij.kumar@s.amity.edu

Abstract - This paper explores how one can use Python for Exploratory Data Analysis (EDA) in various data sets in healthcare; we deep-dive into different Python libraries like Pandas, NumPy, and Matplotlib to scrutinize different healthcare data [1]. Our main goals include deriving patterns, trends, and anomalies in the data and unearthing patient details, diseases, and treatments that may not already be known to the public or medical field. Our methodology encompasses data cleaning, managing of missing values, and variable transformation to ensure data quality, not forgetting the use of descriptive stats, visualizations, and statistical tests, all of which will empower us to spot significant healthcare data characteristics [2]. We will see first-hand how Python can serve as a very potent tool for working efficiently with complex healthcare data sets. Also, we will also underscore how EDA allows us to find relationships between variables and may enable predictive modelling and evidence-based decision-making [4]. Findings from this talk will contribute towards informing healthcare trends which will benefit healthcare professionals, policy-makers, and a general audience [5].

Keywords: Exploratory Data Analysis (EDA), Machine Learning, Graphs, Mean, Median, Mode, Heatmap, Histogram, Pie Chart, Analysis, Skewness.

I. INTRODUCTION

Exploratory Data Analysis (EDA) plays a critical role in understanding and making sense of complex healthcare data. It helps provide leverage in understanding the complex patterns hidden in the data and thus aids policy making keeping in mind the full spectrum of insights [6]. The primary goal of EDA is to maximize the analyst's insight into a data set and the Unites States medical patent dataset is no exception [7]. Identifying hidden insights, complex patterns and making healthcare data as informative as possible.

Python and its robust libraries such as Pandas, NumPy, Matplotlib, Seaborn provide an excellent environment to explore and make sense of extensive and diverse healthcare datasets [8]. That data scientists and healthcare professionals can use to dive into the ascertain details and unlock the full spectrum of observations in healthcare data [9].

In the end, the results of EDA go far beyond traditional data analysis to the point where it creates understanding that "transforms how organizations view their healthcare data" into "meaningful and actionable information for use in their patient care strategies, decisions on implementation of healthcare policy, use for strategic initiatives within their organizations and the creation of the medical knowledge of the future."

The main research contribution of EDA is, in order:

1. Get to know the database.
2. Create a visual such as a chart or graph.
3. Seek out unusual or unexpected values.
4. Create relevant models.
5. Create specific data points for healthcare analysis.

There are many ways we can categorize the EDA techniques [10-11]. For your convenience, I have included the table that categorizes some of the EDA techniques as below shown.

Univariate analysis on numerical data is used to help determine each individual data point's distribution measurements. Histograms and box plots are good univariate analysis techniques. Once you have 2D arrays of numerical data, you can perform bivariate analysis [12]. Scatter plots are a great way to do this. Multivariate analysis for numerical data may involve techniques such as PCA and correlation matrices.

Table 1: Suggested EDA Techniques for Data

Type of Data	Suggested EDA Techniques
1. Categorical	Descriptive Statistics
2. Univariate continuous	Line plot, Histograms
3. Bivariate continuous	2D scatter plots
4. 2D arrays	Heatmap
5. Multiple groups	Side-by-side boxplot

Table 2: Useful EDA Techniques depending on the objectives

Objective	Suggested EDA Techniques
1. Getting an idea of the distribution of a variable	Histogram
2. Finding outliers	Histogram, scatterplots
3. Quantify the relationship between 2 variables	2D scatter plot+/curve fitting Covariance and correlation.
4. Visualize the relationship between 2 exposure variables and 1 outcome variable	Heatmap
5. Visualization of high-dimensional data	PCA + 2D/3D scatterplot

II. LITERATURE REVIEW

In the dynamic landscape of healthcare, the utilization of data analysis techniques has become indispensable for unravelling critical insights. This literature review endeavours to comprehensively explore existing research on the application of data analysis methodologies in healthcare contexts. By synthesizing findings from a diverse array of sources, this review aims to shed light on the methodologies employed, challenges encountered, and significant discoveries made in this burgeoning field.

Key Themes and Findings:

Methodological Diversity: The review underscores the diverse array of methodologies utilized in healthcare data analysis, including descriptive statistics, visualization techniques, machine learning algorithms, and predictive modelling [13-17]. Each methodology offers unique strengths and limitations, influencing the interpretation of findings and decision-making processes within healthcare contexts.

Applications Across Healthcare Domains: Numerous studies demonstrate the wide-ranging applications of data analysis techniques in various facets of healthcare, including clinical decision support, disease surveillance, patient outcomes prediction, and resource optimization. These applications highlight the transformative potential of data-driven insights in enhancing healthcare delivery, optimizing resource allocation, and improving patient outcomes.

Challenges and Limitations: Despite the promise of data analysis in healthcare, several challenges persist, including issues related to data quality, privacy concerns, interpretability of complex models, and ethical considerations. Addressing these challenges is imperative for maximizing the utility and impact of data-driven approaches in healthcare settings.

Future Directions: The literature review identifies emerging trends and future research directions in healthcare data analysis, such as the integration of disparate data sources, advancements in explainable AI techniques, and enhancements in data privacy and security measures. These developments hold the promise of further augmenting the efficacy and relevance of data analysis in healthcare contexts.

1. Data Sources in Healthcare

Healthcare data is coming from diverse data sources which include Electronic Health Records (EHR), claims data, medical imaging, and from medical devices such as wearable devices. EHR contains patient's history and diagnosis details and details of treatment which creates a challenge for data security and standardization but also creates an opportunity for personalized medicine. Claims data contains details of service and cost and has a challenge to ensure accuracy of coding and also to prevent fraud on the other hand it creates opportunities for health policy research and for temporal trend analysis. Medical imaging data contains insights to understand the condition but also has challenges to deal with large file size and also to ensure the privacy and also it creates opportunity for early disease detection [18]. Wearable devices continuously generate health data and creates a challenge for privacy and to ensure data accuracy, and they also create an opportunity for continuous real-time monitoring of health.

2. Python libraries for EDA

• In healthcare data analysis, several Python libraries take centre stage for Exploratory Data Analysis (EDA). Principal libraries include Pandas for data manipulation, NumPy for numerical operations, and Matplotlib for basic plotting. Seaborn builds on top of Matplotlib for aesthetic visualizations, while Plotly and Bokeh crop up for interactive plotting.

• Pandas are potent for structured data operations and are designed to make data analysis and manipulation straightforward. NumPy is critical for numerical operations. It provides support for arrays and matrices, as well as a collection of mathematical functions to operate on these arrays. Matplotlib is a basic plotting library with embedded charts. Seaborn works particularly well with Pandas *data frames* and arrays and is a high-level interface to Matplotlib.

• Plotly is interesting because unlike Matplotlib and Seaborn, Plotly is able to create interactive plots that allow the user to hover the data and each time Python converts that into JSON and transmits that to Plotly. Bokeh has its own, useful data manipulation library that allows you to filter, transform and aggregate your data before feeding it to a plot. Bokeh is particularly interesting if you're interested in or active on the Bokeh server.

In some situations, it pays off to invest time learning and practicing advanced plotting libraries. The payoff is in the ability to dynamically explore data in charts without needing to code one's way through multiple iterations of the same chart.

3. Handling Missing Data & Data Pre-processing

In healthcare datasets, dealing with missing data is critical during the data pre-processing stage, before carrying out any analyses on the data to ensure the analyses execute on accurate and reliable information [19-20]. Python, being the versatile programming language, it is, provides a multitude of strategies and techniques you can use to handle missing data, with scores of dedicated functions and methods in several libraries. PANDAS and scikit-learn, for example, are two libraries that offer the Simple Imputer function, which allows for mean or median imputation by rule-of-thumb imputation essentially simple and fast.

4. Statistical Analysis & Visualization Techniques

In healthcare Exploratory Data Analysis (EDA), statistical methods are important. Descriptive statistics provide a "summary" that may capture much of the important features of the data in a concise statement. For example, the sample mean, median, sample standard deviation, etc. can provide as a sample summary. Hypothesis testing formally juxtaposes observed differences or relationships with those expected due to random chance. Inferential statistics encompasses the process of making predictions or inferences about a larger population based on a sample of data.

The statistical analyses described above are implemented using Python's SciPy library. This is a powerful tool, as its stats module allows us to compute descriptive statistics using a function like describe [21-22]. Functions for hypothesis testing are also available, such as testing for computing the p-value of the means of two independent samples This in itself might not substantially benefit their understanding of complex health patterns, as very large datasets must be communicated through visualization methods that are more useful in enhancing our knowledge of complex information.

Healthcare professionals can use these visualization techniques to (a) make sense of the data, (b) see trends, and (c) communicate effectively complex data to a lay audience.

III. RESEARH METHODOLOGY

1. Understanding the Data

Handling missing data is one of the pivots in pre-processing healthcare datasets, an essential step that dictates the correctness and reliability of downstream analyses. The domain's go-to language, Python, is equipped with a plethora of techniques thanks to its science-friendly libraries, each offering an array of functions and methods tailored to address different missing data intricacies.

Among these techniques is imputation, a fundamental method that entails replacing or estimating missing values. Python's Pandas and scikit-learn libraries have a compact and user-friendly function, Simple Imputer, that provides a mechanism to encapsulate simple methods like mean or median imputation [23]. For more complex scenarios, however, we often must resort to machine learning-based imputation, which leverages algorithms—such as those implemented in scikit-learn and other libraries—to predict missing values based on their relationships with other, non-missing values.

Normalization and feature scaling are a crucial pre-processing step and it ensures that ML models perform optimally. Scikit-learn library in Python has several functions that let you carry out the normalization and feature scaling process. MinMaxScaler and Standard Scaler are two commonly used functions for this purpose. Normalization makes sure the numerical features have a value range between 0 and 1. This helps make features more uniform [24]. Feature scaling ensures all feature have the same scale. The technique is used so that one significant feature doesn't dominate the other feature during the machine learning process.

2. Descriptive Statistics

In the realm of Exploratory Data Analysis (EDA) for healthcare datasets, a pivotal step involves the calculation of fundamental statistics to unveil essential insights into the numerical variables. This encompasses measures such as the mean, median, mode, standard deviation, and range, which collectively offer a comprehensive overview of the central tendencies and variability inherent in these quantitative attributes [25].

1. Mean: - The mean, or average, provides a measure of the dataset's central tendency, representing the arithmetic average of all numerical values. It serves as a benchmark for understanding the typical value within a given variable.

2. Median: - The median is the middle value in a dataset when arranged in ascending or descending order. It is particularly insightful as it is less sensitive to extreme values.

3. Mode: – Mode represent the most repetitive data within the dataset and it is very important to understand that at which point this graph or distribution is high.

1. Frequency Tables: - They display the counts or the percentages of the unique categories within a variable. It enables a more detailed inspection of the variable's distribution and also provides the actual value of the count.

Two-way frequency table

maint class	high	low	med	vhigh	All
acc	105	92	115	72	384
good	0	46	23	0	69
unacc	314	268	268	360	1210
vgood	13	26	26	0	65
All	432	432	432	432	1728

Two-way row relative frequency table

maint class	high	low	med	vhigh	All
acc	0.27	0.24	0.30	0.19	1.0
good	0.00	0.67	0.33	0.00	1.0
unacc	0.26	0.22	0.22	0.30	1.0
vgood	0.20	0.40	0.40	0.00	1.0
All	0.25	0.25	0.25	0.25	1.0

Figure 1: Two-Way Frequency Table

2. **Bar Charts: -** These are visual displays of the frequencies or the proportions of the categories within a categorical variable. These extend the interpretability as you can quickly identify any imbalances in the dataset and also the categories that are distinctly more prevalent than the others.

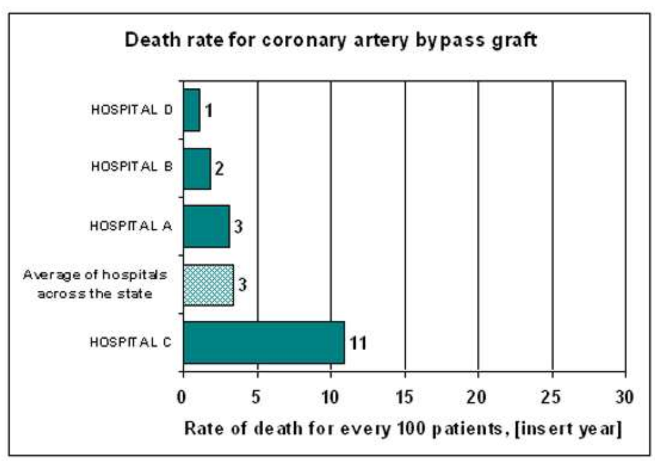

Figure 2: Bar Chart for Death rate for coronary artery bypass graft

The EDA process, through these statistical calculations and visualizations, unravels the rudimentary characteristics of numerical and categorical variables and paves the way for a glimpse at the nuances of the healthcare dataset to come in later insights. This meticulous examination informs decision-making and comprehension of data patterns enriching the health research and practice process with a much more refined strategy.

3. Data Visualization & Handling Outliers

Harness the capabilities of prominent visualization libraries, such as Matplotlib, Seaborn, and Plotly, to craft engaging and informative visualizations. Employ diverse visualization techniques, including histograms, box plots, and violin plots, to effectively illustrate the distribution of numerical variables [26-27]. For categorical variables, leverage bar charts, pie charts, and count plots to provide a clear and concise representation.

4. Feature Engineering

One example of how domain knowledge can be leveraged is the incorporation of subject matter expertise within the industry or field of study into data science. This specialized knowledge can point data analysts to the development of new features that capture information about the dataset that enables distinction based on pertinent categories or pertinent distinctions. These features and new categories capture information in a dataset that goes beyond the raw data and can give context that reveal patterns or identify relationships that are not visible within the raw data. We'll look at practical technique used for the handling of categorical variables in machine learning models, One-Hot Encoding, which creates a binary column for each category and returns a sparse matrix [28]. By transforming categorical data into the numerical format our model prefers, we can include categorical data in our models, and the One-Hot Encoder ensures that those categories are appropriately interpreted. It was a crucial step in prepping the dataset for the exploration and modelling steps.

5. Group Analysis & Correlation Analysis

Clustering numerical data into categories based on relevant categorical variables is a process that requires creating subsets of data with some common attributes or properties. Each category is then analysed in more details and a deeper understanding can be gained from the data, i.e. to see how different factors influence the data [29]. With the data organized into these groups, the next step is to compare distributions and/or summary statistics across these groups. This comparative approach often exposes subtle trends, patterns, and variations that are difficult to uncover when *analysing* the entire data set. This granularity can lead to rich insights into the unique characteristics of each category.

6. Documentation & Communication

Your EDA process must be thoroughly documented: this should include your discoveries, insights, and decisions made. You should explain any data transformations or feature engineering that is done so that you are as clear as possible about the decisions you make throughout your process, as want to enable someone to follow your analysis (and value deriving process!).

Your communication of your insights and final recommended actions (for the question(s) from the brief) as a result of your analysis must be effective. Provide actionable insights so your technological stakeholders can have a plan on how to action, the result of your findings. Having a technical work colleague, communicate these results effectively, a good approach to breaking down how to communicate each step of the said action plan and finally providing them with that plan.

IV. OBSERVED IMPROVEMENT BEFORE EDA PROCESS

Improving results before EDA involves optimizing your data and refining your features to increase the power of any later analyses or modelling efforts. This preliminary step is meant to remove egregious errors or anomalies—such as missing data, outliers, or feature engineering—to create a more solid foundation on which to move into the exploration.

Here are a few specific tasks to consider:

1. Handling Missing Data: Identify and implement a strategy to handle missing data within the dataset. Techniques such as imputation or deletion of incomplete records are common to boost the completeness of the data.

2. Outlier Detection and Treatment: Identify outliers within numerical variables and make a judgment concerning how to deal with them. This may consist of deleting extreme values, conducting a transformation on the data, or applying imputation techniques to help mitigate their impact on an analysis.

3. Feature Engineering: Engineer new features or transform existing ones to capture relevant information and relationships within the dataset. This can involve incorporating domain knowledge, creating interaction terms, or converting categorical variables into a numerical format for machine learning compatibility.

4. Normalization and Scaling: Ensure that numerical features are appropriately normalized and scaled. This step is vital for preventing certain features from dominating the analysis due to differing scales, promoting a balanced and fair assessment of each variable's contribution.

5. Data Cleansing: Identify and rectify any inconsistencies, errors, or anomalies present in the data. This ensures that the dataset is reliable and accurate for subsequent analyses.

By first addressing these aspects before conducting any Exploratory Data Analysis, the analyst sets the stage for a more effective and insightful exploration. By doing so, she creates a cleaner and more refined dataset that results in more accurate and insightful findings during Exploratory Data Analysis, and thus better results in any subsequent analyses or modelling endeavours.

V. OBSERVED IMPROVEMENT AFTER EDA PROCESS

Improving after performing Exploratory Data Analysis (EDA) includes refining insights, improving feature selection, and a deeper understanding of the dataset. EDA provides insights such as patterns, trends, and relationships; these insights allow for targeted improvements that contribute to better overall modelling and analysis. Key aspects of improvement post-EDA include (but are not limited to) the following:

1. **Feature Selection and Engineering:** EDA often sheds light on how important certain features are. Analysts can use these hints to refine their lists of features, discarding irrelevant variables or creating new features based on what EDA shows them about subtle patterns. Targeted feature engineering helps build a more focused and predictive model.

2. **Model Tuning:** Insights from EDA guide the adjustments of model parameters and hyperparameters. With these characteristics of the data at their disposal, analysts can tweak models to optimize their performance and predictive capabilities.

3. **Addressing Data Skewness or Distribution Issues:** EDA may reveal skewed distributions or non-normality in the data. By addressing these through transforming or normalizing the data, the data is better suited for various modelling algorithms.

4. **Handling Correlations:** Understanding correlations between variables can help to manage multicollinearity. By refining the feature set to include just the useful information, you reduce redundancy and help to improve model interpretability.

5. **Optimizing Data Cleaning Strategies:** EDA can reveal nuances in data quality that can be addressed and further refined through other data cleaning strategies that may not have been initially considered.

6. **Incorporating Domain Knowledge:** Often the best outcome of EDA is to realize how little you know about the data. Insights gained during EDA can often bring up particular domain-related considerations which can then be built into the modelling process for a model that more closely aligns with real-world scenarios.

7. Validation and Cross-Validation Techniques: EDA helps in identifying the most suitable validation and cross-validation strategies by the analyst. These techniques can be improved upon based on the specific characteristic identified from the data, thus enabling robust model evaluation.

In summary, the improvement after EDA involves taking the new understanding of the dataset to enhance models and feature engineering, as well as optimize various elements of the analytics process. This iterative process of exploration and refinement will eventually produce more accurate and reliable results in subsequent analyses and modelling efforts.

VI. ETHICAL AND PRIVACY CONSIDERATIONS

In healthcare data analysis, the stakes are heighted in terms ethical considerations, especially when it comes to respecting patient privacy, securing data, and following regulations like HIPAA. Protecting private and sensitive information isn't just about maintaining trust in your analysis; it's about being a responsible data practitioner. As you explore your data through your EDA, you can use the tools that Python have to help you act ethically with your patient data along every step of analysis.

CONCLUSION

In summary, the literature review confirms that assessing a variety of healthcare datasets through EDA is extremely significant. The literature reflects that EHR, claims data, medical imaging, wearable devices and genomic data are among the primary sources of healthcare data, and can be used for a wide variety of analytical endeavors. Included in this literature review are applicable Python libraries, which are very important, such as Pandas, NumPy, Matplotlib, Seaborn, and Plotly. Also necessary are Python libraries, such as Bokeh, which allow for extremely versatile applications in EDA with this dataset. These libraries allow for effective data manipulation, a significant number of visualization variants, and interactive exploration of the dataset. As the dataset is truncated due to incomplete entries, this dataset pre-processing step is very important. Use of SciPy and other Python libraries also allows for confidence that statistical analyses leading to profiling on the potential findings from these truncated healthcare datasets, will be of the highest quality. The literature reflects that ongoing limitations to data interoperability and privacy issues must be given significant attention, so should the lack of standardized formats in healthcare datasets. Integration of machine learning with EDA and the near-future expectation that healthcare datasets originate from analytical activities, will be handled by in-memory platforms, leading to truly real-time analytics are two areas of the greatest promise for future research.

REFERENCE

[1] Natrella M (2010) NIST/SEMATECH e-Handbook of Statistical Methods. NIST/SEMATECH

[2] Tukey J (1977) Exploratory data analysis. Pearson, London.

[3] Kaski, Samuel (1997) "Data exploration using self-organizing maps.

[4] Joanes DN, Gill CA (1998) Comparing measures of sample skewness and kurtosis.

[5] Seltman HJ (2012) Experimental design and analysis.

[6] Andres, C., Fukunaga, A., Biesiadecki, J., Lynn, N., Whelley, P., Greeley, R., et al. (2008). Automatic detection of dust devils and clouds on Mars. Mach. Vis. Appl. 19, 467–482. doi:10.1007/s00138-007-0081-3

[7] Brinckerhoff, W., Willis, P., Ricco, A. O., Kaplan, D. A., Danell, R. M., Grubisic, A., et al. (2022). Europan molecular indicators of life investigation (EMILI) for a future Europa lander mission. Front. Space Technol. 2. doi:10.3389/frspt.2021.760927

[8] Brockwell, T., Meech, K., Keith, P. S., Waite, J. H., Miller, G., Roberts, J., et al. (2016). "The mass spectrometer for planetary exploration (MASPEX)," in IEEE aerospace conference 2016. doi:10.1109/AERO.2016.7500777

[9] Chien, S., Bue, B., Castillo-Rogez, J., Gharibian, D., Knight, R., Schaffer, S., et al. (2014). Agile science: Using onboard autonomy for primitive bodies and deep space exploration. Proc. Intl. Symposium Artif. Intell. Robotics, Automation Space. doi:10.2514/6.2014-1888

[10] Chou, L., Paul, M., Trainer, M., Eigenbrode, J., Arevalo, R., Brinckerhoff, W., et al. (2021). Planetary mass spectrometry for agnostic life detection in the solar system. Front. Astronomy Space Sci. 8. doi:10.3389/fspas.2021.755100

[11] Cohen-Addad, V., Kanade, V., Mallmann-Trenn, F., and Mathieu, C. (2017). Hierarchical clustering: Objective functions and algorithms. Arxiv. doi:10.48550/arXiv.1704.02147

[12] Da Poian, V., Lyness, E., Ryan, D., Xiang, L., Theiling, B., Trainer, M., et al. (2022). Science autonomy and space science: Application to the ExoMars mission. Front. Astronomy Space Sci. 9. doi:10.3389/fspas.2022.848669

[13] Gandhi, G., and Srivastava, R. (2014). Review paper: A comparative study on partitioning techniques of clustering algorithms. Int. J. Comput. Appl. 87, 10–13. doi:10.5120/15235-3770

[14] Kanungo, T., Mount, D., Nathan, N., Piatko, C., Silverman, R., and Wu, A. (2002). An efficient k-

means clustering algorithm: Analysis and implementation. IEEE Trans. Pattern Analysis Mach. Intell. 24 (7), 881–892. doi:10.1109/TPAMI.2002.1017616

[15] Kutbay, U. (2018). Partitional clustering. Recent applications in data clustering. doi:10.5772/intechopen.75836

[16] Maharana, K., Mondal, S., and Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. Glob. Transitions Proc. 3 (1), 91–99. ISSN 2666-285X. doi: 10.1016/j.gltp.2022.04.020

[17] Gupta A, Kumar Rout R. Energy harvesting-enabled energy-efficient routing using spotted hyena optimization in wireless sensor network. *Int J Commun Syst*. 2024; 37(8): e5747.

[18] S. Suman, P. Kaushik, S. S. N. Challapalli, B. P. Lohani, P. Kushwaha and A. D. Gupta, "Commodity Price Prediction for making informed Decisions while trading using Long Short-Term Memory (LSTM) Algorithm," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 406-411

[19] B. M. Sahoo, A. D. Gupta, S. A. Yadav and S. Gupta, "ESRA: Enhanced Stable Routing Algorithm for Heterogeneous Wireless Sensor Networks," 2019 International Conference on Automation, Computational and Technology Management (ICACTM), London, UK, 2019, pp. 148-152,

[20] B. D. Shivahare, S. Suman, S. S. N. Challapalli, P. Kaushik, A. D. Gupta and V. Bibhu, "Survey Paper: Comparative Study of Machine Learning Techniques and its Recent Applications," 2022 2nd International Conference on Innovative Practices in Technology and Management (ICIPTM), Gautam Buddha Nagar, India, 2022, pp. 449-454

[21] Gupta AD, Kumar Rout R. SMEOR: Sink mobility-based energy-optimized routing in energy harvesting-enabled wireless sensor network. *Int J Commun Syst*. 2024

[22] S. S. N. Challapalli, P. Kaushik, S. Suman, B. D. Shivahare, V. Bibhu and A. D. Gupta, "Web Development and performance comparison of Web Development Technologies in Node.js and Python," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 303-307

[23] B. D. Shivahare, M. Singh, A. Gupta, S. Ranjan, D. Pareta and B. M. Sahu, "Survey Paper: Whale optimization algorithm and its variant applications," 2021 International Conference on Innovative Practices in Technology and Management (ICIPTM), Noida, India, 2021, pp. 77-82

[24] S. Singh, D. Chaudhary, A. D. Gupta, B. Prakash Lohani, P. K. Kushwaha and V. Bibhu, "Artificial Intelligence, Cognitive Robotics and Nature of Consciousness," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), London, United Kingdom, 2022, pp. 447-454

[25] Gupta AD, Rout RK. ROTEE: Remora Optimization and Tunicate swarm algorithm-based Energy-Efficient cluster-based routing for EH-enabled heterogeneous WSNs. *Int J Commun Syst*. 2023; 36(2):e5372

[26] Gupta, A. D., & Rout, R. K. (2022). A WOA Based Energy Efficient Clustering in Energy Harvesting Wireless Sensor Networks. *Amity Journal of Computational Sciences*, *5*(2).

[27] A. D. Gupta and R. K. Rout, "An Effective Optimization Method for Energy Efficient Clustering in EH Wireless Sensor Networks," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 699-702

[28] T. Anand, V. Singh, B. Bali, B. M. Sahoo, B. D. Shivhare and A. D. Gupta, "Survey Paper: Sentiment Analysis for Major Government Decisions," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 104-109

[29] D. Chaudhary, S. Singh, R. Patel, B. M. Sahoo, B. D. Shivhare and A. D. Gupta, "The Relationship Between Neural Networks and Stock Market Predictions," 2020 Research, Innovation, Knowledge Management and Technology Application for Business Sustainability (INBUSH), Greater Noida, India, 2020, pp. 111-118