Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2024, 11(4):25-31

Research Article ISSN: 2394 - 658X

Optimizing Big Data Management: Integrating SEMMA and CRISP-DM with Formal Concept Analysis for Enhanced Knowledge Discovery

Iqtiar Md Siddique

Department of Industrial, Manufacturing and Systems Engineering, the University of Texas at EL Paso, US.
*iqtiar.siddique@gmail.com

ABSTRACT

This paper proposes a novel approach to handling big data effectively through the fusion of the SEMMA (Sample, Explore, Modify, Model, Assess) and CRISP-DM (Cross-Industry Standard Process for Data Mining) models, augmented by Formal Concept Analysis (FCA) in knowledge discovery. By integrating these established frameworks, the methodology aims to optimize the process of extracting valuable insights from vast and complex datasets. SEMMA and CRISP-DM provide a structured and systematic workflow for data analysis and mining, covering data sampling, exploration, modeling, and evaluation stages. Formal Concept Analysis adds a layer of abstraction, facilitating the identification of meaningful patterns and relationships within the data. This combined approach enables a comprehensive analysis of big data, leveraging both statistical and machine learning techniques to uncover hidden knowledge. The proposed methodology offers a unified and adaptable framework for organizations seeking to navigate the challenges of big data analytics efficiently. Through case studies and empirical validation, the effectiveness of the SEMMA-CRISP-DM-FCA fusion in knowledge discovery is demonstrated, highlighting its potential to drive innovation and informed decision-making in various domains.

Key words: Formal concept analysis; Knowledge discovery; Data mining; Big data; CRISP-DM model; SEMMA model

INTRODUCTION

The value of DA - Data Analytics has been broadly utilized by various academic and industrial communities. Big Data and Data Mining concepts transform our lives in all dimensions from businesses to consumers. Therefore, this transformation takes place at each level from different microeconomics to macroeconomics, where the consumers andmanufacturers connect with each other in order to realize emerging forms of value that lead to societies and economic function. Further, data analytics and big data are considered ever-growing technology that attracts more attention in an industrial view and scientific literature with various news media. Most of the companies using these technologies believe that analytics and management deliver values and indicate basic competitive source benefits in the industry. However, due to the lack of DA insights in the system, most companies are not aware of the potential veiled behind certain unexploited available data that creates with the advanced information methods. Nevertheless, changing the real- time data into a data-driven one becomes challenging [13]. Moreover, identifying the activities of the human is considered an important factor in the customized context-aware systems design and development applied in living applications, further this enables the appropriate assistance. Normally the system will gather information with behavior analysis to identify the activities processed and the continuous behaviors that are considered vital to complete the activity performed. At the same time, it is important to pick the appropriate feedback in order to help the smart environment residents. However, identifying human activities using the sensor-based smart home is considered a challenging task, especially in the scientific community due to the human-object interactions, different behavioral patterns, and reliable data. Further, the difficulty during the recognition of activities increases due to multiple residents, especially in a smart environment. In these scenarios, multiple inferences should be employed to the similar sensors at the same place and exact time [12].

FCA - Formal Concept Analysis is considered a powerful and efficient tool support system. Recently, several tools on Big Data exist and failed to identify the issues occurring between the current tools and appropriate processing that should be exhibited to address the FCA's current needs. It is significant for researchers to find scalable and efficient solutions, especially for FCA that enable the proper prototyping for executing FCA cross-tables even on a large-scale [10]. The topic detection task is completely targeted at exploring the main important topics to be addressed by a whole range of documents and the topics are outlined as self-contained, cohesive, and thematically similar. Hence, this task is widely examined from the clustering and probabilistic methods. The FCA concept is an exploratory method utilized for data analysis [11] and these FCA-based techniques are used for topic detection that is employed in the literature by providing the stability concept for the selection of topic. Certainly, the FCA-based concept could overcome the issues of clustering and probabilistic methods. With the increasing capability to accumulate, analyze, and store the data generated with increasing frequency, the data science field started to grow rapidly. Similarly, due to the rapid increase in the organization using data analysis allows accumulating of larger amounts of data. Therefore, these analyses keep on increasing which becomes a team activity, instead of work performed by a single data scientist. During this process, the growth in the utilization of big data has exceeded the knowledge of supporting teams that are required to do big data tasks. The algorithms utilized in the research help develop insightful analysis, frameworks, tools, and techniques that allow teams to process effectively to perform big data projects. Still, operations research, business intelligence, and software development are discussed in existing research that provides proper insights on big data process utilization [3]. The utilization of attribute reduction techniques for formal decision contexts is explored, employing a simplified discernibility matrix to develop formal concepts. This approach indicates reduced computation time and storage space compared to the original technique. Additionally, approximation algorithms are employed to minimize redaction within formal decision contexts based on graph theory. Various experiments are conducted to validate the effectiveness of this method, showing improved performance in storage speed and space utilization when applied to larger datasets. However, the reduction approach in formal decision contexts faces challenges in terms of discernibility matrix construction and computational complexity, leading to significant time requirements [6]. The data mining concepts, and knowledge discovery techniques used in recent research have been discussed with different degrees of success. Both DM - Data Mining & KDD - knowledge discovery in databases, DM, and KDD terms are utilized to define the techniques, tools, and research utilized to extract appropriate and useful information from a large volume of information or data. The process which has been completely executed in the data extraction process is known as the KDD process. The existing research such as CRISP-DM - Cross-Industry Standard Process for DM - Data Mining [11] that is used in DM development projects. Therefore, these techniques define the activities to develop a DM project and each activity consists of tasks. These tasks generate output and required inputs are detailed and these methods are used to resolve the issues with CRISP-DM in existing DM development projects.

LITERATURE REVIEW

An examination of the k-nearest neighbor (k-NN) algorithm as a data mining (DM) technique is conducted to acquire high-quality data, referred to as smart data, by addressing issues such as noise, missing values, and redundant data through data preprocessing methods. Various data preprocessing approaches based on k-NN techniques are discussed and experimented with under Apache Spark, aiming to expedite the processing of techniques while maintaining data quality. The empirical analysis conducted on collected behavioral datasets enables non-experts and practitioners to define smart data preprocessing methods for large datasets. However, in many cases, the accuracy is compromised, and the reduction in redundancy is substantial [7], leading to a trade-off between accuracy and reduction

The state-of-the-art principles and DM methods are reviewed with missing information or process outliers in different industrial applications. The data preprocessing methods are compared with the normalization and data cleaning [5]. The robust statistical process modeling methods have been utilized on the PCA basis. The different statistical DM methods are examined with various process characteristics such as dynamics, non-linear, and non-Gaussian. a fuzzy concept is employed to reduce the number of concepts in Formal Concept Analysis (FCA) by integrating fuzzy attributes. This involves computing the weight of Fuzzy Formal Concepts (FFC) using Shannon entropy. Subsequently, the weight of FFC is minimized by selecting the computed weight granulation. The outcomes from this method, utilizing interval valued FFC and the Levenshtein distance technique, demonstrate improved results with minimal computational complexity [8]. Interval pattern structures are employed to categorize work into three distinct categories. This includes biclusters, which have found application in recommender systems, yet lack effective techniques and appropriate semantics. Additionally, a pattern-based classifier is introduced, which defines numerical patterns, closed patterns, and generators, facilitating the initialization of association rules during processing. These patterns can then be utilized in supervised classification tasks. Furthermore, k-anonymity is implemented alongside projections to enhance the security of data. The presented generators and closed patterns are applied to datasets to address critical issues observed on the web [4].

The study incorporates a comprehensive examination of knowledge discovery methods, data mining techniques, and process models, providing a detailed overview of the Knowledge Discovery in Databases (KDD) process. Special emphasis is placed on delineating the distinctive features, advantages, and disadvantages of KDD processes. Furthermore, various data mining techniques tailored to specific tasks are elucidated within each method, facilitating a thorough interpretation of the entire KDD process. Comparative analysis is conducted with a novel data mining and KDD process, termed as a refined data mining process, aimed at developing specific tasks using analyzed techniques. Moreover, a domain-driven data mining approach proposed in prior work aims to construct an actionable knowledge discovery framework. The primary objective is to enhance the implementation of learned rules by analyzing requisite actions in business tasks, determining their sequential execution, and investigating task-specific constraints and data. However, findings indicate that knowledge derived from rule learning or domain expertise may be insufficient for task execution, necessitating further in-depth analysis [1,2]. Recently, data mining has shown impressive effectiveness in handling big data and has achieved success when exploited in tasks like data analytics. Moreover, the insertion of attention mechanisms in data mining have made it feasible to consider and extract the related important data rather than extracting the available information which are invaluable for big data analytics. These characteristics have motivated us to develop a data mining-based framework and explore its robustness and efficacy for performing effective big data analytics. Furthermore, the excellence of FCA in knowledge discovery has motivated this research to integrate FCA with data mining for effectively handling big data. This study offers a thorough analysis of Formal Concept Analysis (FCA) in knowledge discovery for data analytics, emphasizing its relevance in managing vast datasets. It addresses the challenges associated with adopting data mining and knowledge discovery techniques, particularly in handling big data without preprocessing. Additionally, it introduces a novel concept for modeling the FCA-based knowledge discovery process, aiming to contribute to existing research endeavors. Through its comprehensive examination and innovative approach, this study aims to enhance understanding and utilization of FCA in the context of data analytics, providing valuable insights for researchers and practitioners alike.

PROBLEM STATEMENT

Data mining and knowledge discovery (KD) refers to the nontrivial extraction of implicit, previously unknown and potentially useful information from data stored in databases. Data mining helps in the generation and extraction of valuable non-trivial knowledge and data from unstructured and structured information. Data mining and KD is used in the visualization, categorization and explication of information and hence are considered as the key elements for enhancing the effectiveness of business processes and main enablers of competitive benefits. However, the success rate of data mining is very low and the value of data in most of the organizations are not available. Various researchers have discussed the enhancement of knowledge discovery and data mining using Formal Concept Analysis (FCA). It can be inferred from existing works that the main challenges associated with the adoption of data mining and KD are the lack of availability of structured techniques for data analytics. Several data mining models powered by FCA have been proposed and there has been a trivial interest in this domain. However, an analysis of current techniques has suggested that the performance of FCA is not satisfactory and even appropriate FCA based techniques cannot be successfully applied for large scale applications. In addition, most of the existing works have used FCA to deal with numerical data**,** and conventional techniques cannot analyze this data without using preprocessing techniques [12].

The KD model acts as a data centric model which acts as an iterative and interactive data analytics process. However, while data-related tasks are detailed, there is a lack of business perspective in this outline model. Most of the existing KD approaches do not focus on performing data mining for extra large-scale datasets. This is mainly because these approaches suffer from high overhead problems while dealing with large scale data processing. This restricts the adaptability of the KD approaches for applications requiring larger datasets. When a software model is used, the KD process becomes an important step in the software development framework that encompasses project management and development processes. Hence, it is essential to develop an efficient model for data mining and KD which is suitable for a dynamic environment and performs well for the volatile objectives which are used in the exploratory analysis. Else, the classical software development approaches might become too rigid and cumbersome to follow. The work presented in [1] shows that despite the availability of several research works, there is a lack of an effective research approach which can overcome the intrinsic limitations of traditional process models and achieve better performance in agile technologies. There is a great demand for a holistic approach to enhance the data mining and KD process and offer better support to analytics project management. Also, it is important to reduce the risk of analytics project failures by accompanying better technologies which can help in designing better processes and provide better understanding of the potential impact of analytics models.

In this context, this research aims to provide a comprehensive analysis of Knowledge Discovery and data mining to address the limitations inferred from existing works. The primary goal of the proposed research is to devise a methodology for handling large-scale data through data mining and knowledge discovery, leveraging formal concept analysis (FCA). The research objectives encompass several key areas: firstly, to investigate the utilization of interactive knowledge discovery and data mining techniques with numeric FCA; secondly, to devise an efficient strategy for managing big data by integrating knowledge discovery with data mining practices; and finally, to enhance the scalability and efficacy of the proposed approach, enabling its applicability to larger datasets while bolstering resilience to missing or erroneous data elements, as well as facilitating visualization and exploration of mining outcomes.

PROPOSED METHODOLOGY

This research endeavors to integrate a data mining and knowledge discovery approach tailored for big data applications. Specifically, it seeks to merge two prominent KDD models, namely SEMMA and CRISP-DM. Existing literature highlights the absence of a standardized methodology for enhancing business-related tasks within organizations, suggesting that conventional big data processing tools may not suffice. Consequently, there is a growing demand for robust models that can adequately address the complexities of analytical projects. Moreover, the proliferation of domain-specific methodologies underscores the need for a standardized KDD process. As organizations increasingly adopt their methodologies, there is a heightened emphasis on data analytics models, including data mining tasks. The experimental findings suggest that data analysts and business management professionals often explore various methodologies to tailor them to their specific requirements. Considering this, the research endeavors to develop a novel methodology by integrating two effective KDD models to better manage big data efficiently. Biswas et al. (2024) consider machine learning algorithm for her data analysis in twitter for covid time and we have considered the coding concept from this paper for our research [14]. CRISP-DM is based on a waterfall life cycle model which has a hierarchical process which helps in providing guidance on how to perform a task. In general, there are six important stages which constitute the top level of hierarchy, which are understanding the business, understanding the data, data preparation, modeling, performance evaluation, and model deployment. These phases incorporate generic tasks wherein the inputs and corresponding outputs are defined clearly. The generic tasks cover the entire data mining process and are suitable for all data mining applications. This process is more stable since it is applicable for all novel and unknown technologies and processes. Though the process involved in the waterfall life cycle is linear and sequential, the CRISP-DM [22] considers an iterative process and feedback loops. It is one of the excellent and well documented process models and its clear and comprehensive documentation plays an important role in improving the functioning of the standard knowledge discovery process models. The CRISP- DM model incorporates all updated elements from the previous models and these elements are considered as the reference for performing all future processes. There are two main explicit steps involved in the CRISP-DM model i.e., business understanding and data understanding. Data Understanding: The dataset provides information about traffic accidents, including weather conditions, severity, location, and more. Since all the data are not required for my storytelling for the questions of interest hence, I have dropped some of the column from the data set using coding. These two steps are considered as an edge for making any data mining task successful since they help in gaining more insights into the objectives of business processes and the data availability. The SEMMA process model consists of 5 main stages namely, sampling, exploring, modifying, modeling, and assessing. These stages focus on modeling the development aspects of data mining. It can be considered as a fundamental plan for modeling the KDD process and it works perfectly with all tasks of KDD in parallel. The workflow stages of the SEMMA model are linked to the SAS Enterprise Miner software and hence is considered as one of the topmost processing models used for data analytics. Though the SEMMA model can be used extensively, it is challenging to perform iterations and carryout interactions between multiple tasks in the original model. To overcome this drawback of the SEMMA model, this research intends to combine it with the CRISP–DM (Cross Industry Standard Process for DM) as shown at below figure.

In [87]:
```
from pprint import pprint
def sanity_check(df):
    pprint('-'*70)
    pprint('No. of Rows: {0[0]}        No. of Columns : {0[1]}'.format(df.shape))
    pprint('-'*70)
    data_profile = pd.DataFrame(df.dtypes.reset_index()).rename(columns = {'index' : 'Attribute',
                                                                     0 : 'DataType'}).set_index('Attribute')

    data_profile = pd.concat([data_profile,df.isnull().sum()], axis=1).rename(columns = {0 : 'Missing Values'})
    data_profile = pd.concat([data_profile,(df.isnull().mean()*100).round(2)], axis=1).rename(columns = {0 : 'Missing %'})
    data_profile = pd.concat([data_profile,df.nunique()], axis=1).rename(columns = {0 : 'Unique Values'})

    pprint(data_profile)
    pprint('-'*70)

sanity_check(df)
```

```
'----------------------------------------------------------------------'
'No. of Rows: 1048575        No. of Columns : 39'
'----------------------------------------------------------------------'
                    DataType  Missing Values  Missing %  Unique Values
TMC                    int64               0       0.00             21
Severity               int64               0       0.00              4
Start_Time            object               0       0.00         522149
End_Time              object               0       0.00         514612
Start_Lat            float64               0       0.00         410611
Start_Lng            float64               0       0.00         398820
Distance(mi)         float64               0       0.00           3043
Description           object               0       0.00         735158
Number               float64          642388      61.26          23517
Street                object               0       0.00          87922
Side                  object               0       0.00              3
City                  object              33       0.00           8607
County                object               0       0.00           1364
State                 object               0       0.00             49
Zipcode               object             145       0.01         182360
Timezone              object             609       0.06              4
Temperature(F)       float64           15281       1.46            709
Wind_Chill(F)        float64          471320      44.95            743
Humidity(%)          float64           16597       1.58            100
Pressure(in)         float64           12179       1.16            925
Visibility(mi)       float64           19534       1.86             64
Wind_Direction        object           15124       1.44             24
Wind_Speed(mph)      float64          104745       9.99            106
Precipitation(in)    float64          495462      47.25            218
Weather_Condition     object           18838       1.80            108
Amenity                 bool               0       0.00              2
Bump                    bool               0       0.00              2
Crossing                bool               0       0.00              2
Give_Way                bool               0       0.00              2
Junction                bool               0       0.00              2
No_Exit                 bool               0       0.00              2
Railway                 bool               0       0.00              2
Roundabout              bool               0       0.00              2
Station                 bool               0       0.00              2
Stop                    bool               0       0.00              2
Traffic_Calming         bool               0       0.00              2
Traffic_Signal          bool               0       0.00              2
Turning_Loop            bool               0       0.00              1
Sunrise_Sunset        object              38       0.00              2
'----------------------------------------------------------------------'
```

In [89]:
```
df.dropna(subset=['Visibility(mi)', 'Wind_Direction', 'Description', 'Humidity(%)', 'Weather_Condition', 'Temperature(F)',
                  'Pressure(in)', 'Sunrise_Sunset', 'Street', 'Zipcode'], inplace=True)
```

In [90]:
```
# Calculate the mean values for each column
mean_1 = df['Precipitation(in)'].mean()
mean_2 = df['Wind_Chill(F)'].mean()
mean_3 = df['Wind_Speed(mph)'].mean()

# Impute missing values in each column with their respective means
df['Precipitation(in)'].fillna(mean_1, inplace=True)
df['Wind_Chill(F)'].fillna(mean_2, inplace=True)
df['Wind_Speed(mph)'].fillna(mean_3, inplace=True)
```

In [91]:
```
print("Number of rows:", len(df.index))
df.drop_duplicates(inplace=True)
print("Number of rows after dropping duplicates:", len(df.index))
```

```
Number of rows: 1018600
Number of rows after dropping duplicates: 1017844
```

Figure 1: Data Preparation and Cleaning

There are two main stages involved in this process for handling big data which are data discovery and model deployment. This phase follows the SEMMA model process, encompassing several key steps. Initially, it involves analyzing requests, akin to understanding business requirements and objectives. Subsequently, data preparation is undertaken, focusing on collecting and transforming data from various sources to ensure its relevance and validity for processing within the SEMMA and CRISP-DM models, a time-consuming yet crucial aspect of the knowledge discovery and data mining process. Following this, data exploration occurs, utilizing modern visualization tools to refine business data, identify missing data, and apply suitable analytical approaches. The next step involves data modeling, where analytical and machine learning algorithms are employed to uncover relationships within the data to address business queries. Finally, the deployment phase entails implementing the model derived from the data discovery phase, leveraging automated and appropriate data models. This stage is pivotal as it enables strategic and operational decisions based on the deployed model's performance, with feedback informing further deployment iterations.

LIMITATIONS AND RECOMMENDATIONS

While the proposed method endeavors to overcome the drawbacks associated with conventional techniques, several limitations of this study exist. Firstly, there is a lack of emphasis on the implementation of contemporary big data tools like Apache Spark or Apache Cassandra for managing large-scale datasets. Incorporating these tools may necessitate additional computational resources and could escalate computational costs. Furthermore, the study is confined to the examination of Formal Concept Analysis (FCA) for big data knowledge discovery, without delving into pertinent issues such as data quality, storage, validation, and the implications of accumulating data from diverse sources. For future work, it is recommended to explore the integration of modern big data processing tools to enhance the scalability and efficiency of the proposed approach. Additionally, addressing the issues surrounding data quality, storage, and validation will be crucial for ensuring the reliability and effectiveness of big data analytics. Furthermore, investigating strategies for handling data accumulation from various sources and mitigating its impact on analysis outcomes could provide valuable insights for advancing the methodology.

CONCLUSION

In conclusion, this study introduced a novel approach for handling big data by combining the SEMMA and CRISP-DM models with Formal Concept Analysis (FCA) in knowledge discovery. While the proposed methodology offers a structured framework for extracting insights from complex datasets, certain limitations were identified, including a lack of focus on modern big data processing tools and overlooking key issues such as data quality and storage. Despite these limitations, the study contributes to the advancement of big data analytics by providing a unified framework for systematic data analysis and mining. Future research should aim to address these limitations by incorporating modern tools and addressing pertinent issues surrounding data quality and storage, thus enhancing the robustness and applicability of the proposed methodology in real-world big data scenarios. Overall, this study lays the groundwork for further exploration and refinement of techniques for effective big data management and knowledge discovery.

Acknowledgments

The author declares that no funds, grants, or other support were received during the preparation of this manuscript.

Disclosure of conflict of interest

The author has no relevant financial or non-financial interests to disclose.

REFERENCES

[1]. Fatima, F., Talib, R., Hanif, M. K., & Awais, M. (2020). A paradigm-shifting from domain-driven data mining frameworks to process-based domain-driven data mining-actionable knowledge discovery framework. IEEE Access, 8, 210763-210774.

[2]. Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137-166.

[3]. Saltz, J. S. (2015, October). The need for new processes, methodologies, and tools to support big data teams and improve big data project effectiveness. In 2015 IEEE International Conference on Big Data (Big Data) (pp. 2066-2071). IEEE.

[4]. Kaytoue, M. (2010). Mining Patterns in Numerical Data with Formal Concept Analysis.

[5]. Zhu, J., Ge, Z., Song, Z., & GAO, F. (2018). Review and big data perspectives on robust data mining approaches for industrial process modeling with outliers and missing data. Annual Reviews in Control, 46, 107-133.

[6]. Chen, J., Mi, J., Xie, B., & Lin, Y. (2019). A fast attribute reduction method for large formal decision contexts. International journal of approximate reasoning, 106, 1-17.

[7]. Triguero, I., García-Gil, D., Maillo, J., Luengo, J., García, S., & Herrera, F. (2019). Transforming big data into smart data: An insight on the use of the k-nearest neighbors' algorithm to obtain quality data. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(2), e1289.

[8]. Singh, P. K., Cherukuri, A. K., & Li, J. (2017). Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy. International Journal of Machine Learning and Cybernetics, 8(1), 179-189.

[9]. Castellanos, A., Cigarrán, J., & García-Serrano, A. (2017). Formal concept analysis for topic detection: a clustering quality experimental analysis. Information Systems, 66, 24-42.

[10]. Tamburri, D. A. (2020). Design principles for the General Data Protection Regulation (GDPR): A formal concept analysis and its evaluation. Information Systems, 91, 101469.

[11]. Hao, J., Bouzouane, A., & Gaboury, S. (2018). Recognizing multi-resident activities in non-intrusive sensor-basedsmart homes by formal concept analysis. Neurocomputing, 318, 75-89.

[12]. Rotondo, A., & Quilligan, F. (2020). Evolution paths for knowledge discovery and data mining process models. SNComputer Science, 1(2), 1-19.

[13]. Das, S., Biswas, J., Siddique, M. I., (2024). Mechanical characterization of materials using advanced microscopy techniques. World Journal of Advanced Research and Reviews, 2024, 21(03), 274–283. 10.30574/wjarr.2024.21.3.0742.

[14]. Biswas, J., (2024). Decoding COVID-19 Conversations with Visualization: Twitter Analytics and Emerging Trends. Journal of Computer Science and Software Testing, Volume- 10, Issue- 1.

[15]. Biswas, J., Das, S., Siddique, I. M., & Abedin, M. M. (2024). Sustainable Industrial Practices: Creating an Air Dust Removal and Cooling System for Highly Polluted Areas. European Journal of Advances in Engineering and Technology, 11(3), 1-11. https://doi.org/10.5281/zenodo.10776875.

[16]. Hasan, M. I., Tutul, M. T. A., Das, S., & Siddique, I. M. (2024). Adaptive Risk Management and Resilience in Automated Electronics Industry. Journal of Scientific and Engineering Research, 11(2), 82-92.

[17]. Joyeshree Biswas, S M Mustaquim, S.M. Saokat Hossain, & Iqtiar Md Siddique. (2024). Instantaneous Classification and Localization of Eye Diseases via Artificial Intelligence. European Journal of Advances in Engineering and Technology, 11(3), 45–53. https://doi.org/10.5281/zenodo.10813807.

[18]. Bazgir, Ehteshamul Haque, Md. Maniruzzaman and Rahmanul Hoque, "Skin cancer classification using Inception Network", World Journal of Advanced Research and Reviews, 2024, 21(02), 839–849.

[19]. Rahmanul Hoque, Suman Das, Mahmudul Hoque and Ehteshamul Haque, "Breast Cancer Classification using XGBoost", World Journal of Advanced Research and Reviews, 2024, 21(02), 1985–1994.

[20]. Rahman, M. A., Bazgir, E., Hossain, S. S., & Maniruzzaman, M. (2024). Skin cancer classification using NASNet. International Journal of Science and Research Archive, 11(1), 775-785.

[21]. Bazgir, E., Haque, E., Sharif, N. B., & Ahmed, M. F. (2023). Security aspects in IoT based cloud computing. World Journal of Advanced Research and Reviews, 20(3), 540-551.

[22]. Buian, M. F. I., Arde, R. A., Billah, M. M., Debnath, A., & Siddique, I. M. (2024). Advanced analytics for predicting traffic collision severity assessment. World Journal of Advanced Research and Reviews, 21(2), 2007-2018.