



(RESEARCH ARTICLE)



Uncovering COVID-19 conversations: Twitter insights and trends

Selim Molla ^{1,*}, Ehsan Bazgir ², S M Mustaqim ³, Iqtiair Md Siddique ⁴ and Anamika Ahmed Siddique ⁴

¹ Department of Computational Science, University of Texas at El Paso, US.

² Department of Electrical Engineering, San Francisco Bay University, Fremont, CA 94539, US.

³ Department of Mathematical Science, University of Texas at El Paso, US

⁴ Department of Mechanical & Aerospace Engineering, University of Texas at El Paso, US.

World Journal of Advanced Research and Reviews, 2024, 21(01), 836–842

Publication history: Received on 02 December 2023; revised on 07 January 2024; accepted on 10 January 2024

Article DOI: <https://doi.org/10.30574/wjarr.2024.21.1.0121>

Abstract

In this paper, we delve into the public discourse surrounding COVID-19 on Twitter to unearth the collective sentiments, concerns, and spread of information during the pandemic. By leveraging a dataset of relevant tweets and corresponding ISO country codes, our analysis will map out the geographical and digital landscape of these conversations. The significance of this work lies in its potential to inform public health strategies, shape policymaking, and contribute to social research on crisis communication. Stakeholders ranging from health officials to the public have a vested interest in understanding the contours of this dialogue. Our objective is to craft a data-driven narrative through visualizations that reveal how the world engages with the pandemic on the digital front, providing actionable insights into global and local responses to COVID-19 using Machine Learning techniques.

Keywords: COVID-19; Machine Learning; Twitter; Word Cloud.

1. Introduction

Social media platforms, such as Twitter and Facebook, have emerged as influential mediums for individuals to articulate their emotions and responses in relation to past or ongoing events. The COVID-19 pandemic has not only reshaped our daily lives but also revolutionized the way information and opinions are disseminated and shared globally. Amidst the unprecedented challenges posed by the virus, social media platforms like Twitter have become vital channels for individuals to express their thoughts, concerns, and emotions. These platforms have evolved into dynamic forums where discussions, debates, and information exchange related to COVID-19 are constantly unfolding. The insights gleaned from this research are poised to offer substantial benefits to a diverse range of stakeholders. Public health bodies, policymakers, and academics all stand to gain invaluable knowledge from understanding global conversations and regional sentiment variances. These insights can inform and guide decision-making processes, enabling the crafting of health messaging and policies that resonate with the public's voice and effectively address the pressing issues highlighted through social media discourse during the ongoing pandemic. Twitter stands out as a prominent platform where users employ concise tweets, previously limited to a maximum of 140 characters and currently extended to 280 characters, to encapsulate their sentiments and reactions in a succinct manner. This characteristic renders Twitter an exceptional medium for comprehending the feelings and reactions of individuals towards a specific event, surpassing the analytical capabilities of other social media platforms. Considering this, researchers have dedicated their efforts towards developing visualization tools that facilitate the exploration and comprehension of individuals' tweets from various perspectives. For instance, numerous visualization tools have been designed with a primary focus on showcasing sentiment analysis of tweets, as exemplified by [1-10]. Conversely, alternative visualization tools concentrate on the analysis and exploration of tweet data based on its geo-spatial information, as demonstrated by [11-17]. The COVID-19 outbreak initially surfaced in news reports during the month of December 2019. Nonetheless, the

* Corresponding author: Selim Molla

gravity of the pandemic only began to manifest itself in the early months of 2020, when it garnered heightened attention and individuals began to comprehend the severity of the situation. As worldwide lockdowns were implemented in response to the pandemic's rapid spread, individuals increasingly turned to social media platforms to express their sentiments and reactions towards COVID-19.

One study proposed a framework that dynamically identifies key topics with labels from COVID-19 tweets using Latent Dirichlet Allocation (LDA) generated topics [18]. Another study examined linguistic and visual communication in UK public health agency tweets related to COVID-19. It found that the combination of linguistic and visual representations in these tweets affected the overall effectiveness of public health communications [19]. Additionally, a study analyzed the sentiments of Indian people towards COVID-19 vaccines on Twitter. It found that most tweets reflected positive sentiments and identified the most frequently occurring terms used in these tweets [20]. Furthermore, a study used visualization techniques to explore people's tweets about the COVID-19 pandemic over time, providing insights into their feelings and reactions [21]. Lastly, a study proposed an IPSH model for predicting COVID-19 tweet content using parts of speech tags and high-frequency words, which was found to be more efficient than baseline machine learning models [22]. Ahmmed et al. (2023) describes how brain tumor imaging has been detected through the deep learning algorithm and we have plan to extend our current research to this stage in future [24]. Ullah, et al. (2023) try to find out the best scenarios for a production floor that considers covid 19 protocol as well for their industry [31]. They also apply this covid system problem into manufacturing operation that used simulation as well [32]. Moreover, Hossain at. el. (2023) details in his research regarding electricity generation from a moving vehicle which can be the further research proposal from our finding [27]. Supplier selection was the focus, but the author uses comparative weightage method using C algorithm which is important for this research [28,29]. In this study, Fayshal et al. (2023) discusses how human health risk have significant effect on humans and our research will take this into consideration for covid 19 [30]. So, from the above all related works it is undoubtedly true that in this post covid period, to consider the COVID protocol in the daily work is an important task and this research will show how much spread the covid across the various zone in the world from which one can impose restriction for the measurements. In this endeavor, we have harnessed the power of data analytics to not only map sentiments but also identify influential communication channels and highlight prevalent concerns. Nevertheless, it is important to acknowledge the scale of our undertaking, as the sheer volume of Twitter data is vast and ever evolving. Therefore, while we strive to provide comprehensive insights, we must also recognize the inherent limitations posed by the unceasing flow of information in the digital sphere. Molla et al. (2023) describes how the medical textiles sector is influenced by covid-19 protocol and we have future research on this category [34].

This study serves as a vital exploration into the world of COVID-19 conversations on Twitter, offering a nuanced understanding of the ongoing discourse and trends. It is our hope that the findings and analyses presented in the following pages will contribute significantly to our collective knowledge of public sentiment during these challenging times and, in turn, aid in the development of more informed, effective, and responsive strategies for addressing the pandemic's multifaceted impact.

2. Methods / Data Understanding

2.1. Data Collected

The data collected for this study comprises tweets related to COVID-19 sourced from Twitter, along with a dataset containing Wikipedia ISO country codes. The dataset encompasses individual tweets, user information, and corresponding country codes. This comprehensive collection offers a broad perspective on the global discourse surrounding the COVID-19 pandemic [23].

2.2. Data Description

Data provides a description of individual tweets, user information, and country codes, which together offer a broad view of the global discourse on the pandemic.

2.3. Data Manipulation

Cleaning (handling missing values, correcting data types), merging (associating tweets with standardized country codes), and transforming (extracting hashtags, processing dates and times for temporal analysis).

- Cleaning (handling missing values, correcting data types), merging (associating tweets with standardized country codes), and transforming (extracting hashtags, processing dates and times for temporal analysis).

- Data to conduct the study: This data answers the research questions about public sentiment and information dissemination during the COVID-19 pandemic.

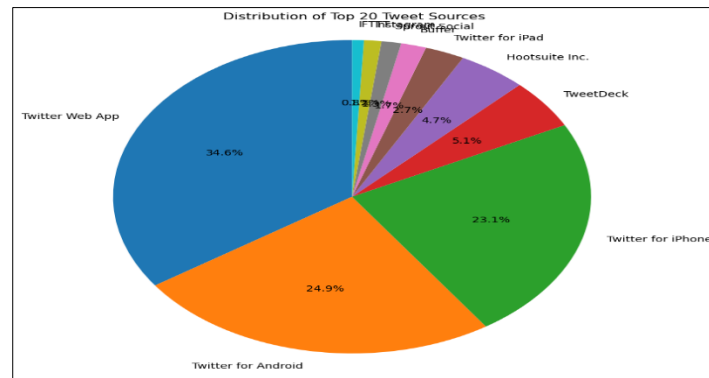


Figure 1 Distribution of the top 10 sources in pie chart

2.4. Data to conduct the study.

The purpose of utilizing this data is to address research questions concerning public sentiment and information dissemination during the COVID-19 pandemic. By diving into the subjects using specific hashtags, examining geographical distribution, evaluating sentiments expressed in tweets, and identifying trends over time, the study aims to gain valuable insights into the public discourse surrounding the pandemic on a global scale. Through these analyses, the study seeks to contribute to a comprehensive understanding of how information flows and sentiments evolve in the context of a global health crisis.

2.5. Visualizations

2.5.1. Bar Graph

Representation of the distribution of tweet sources, which helps in identifying where the most active discussions originate from.

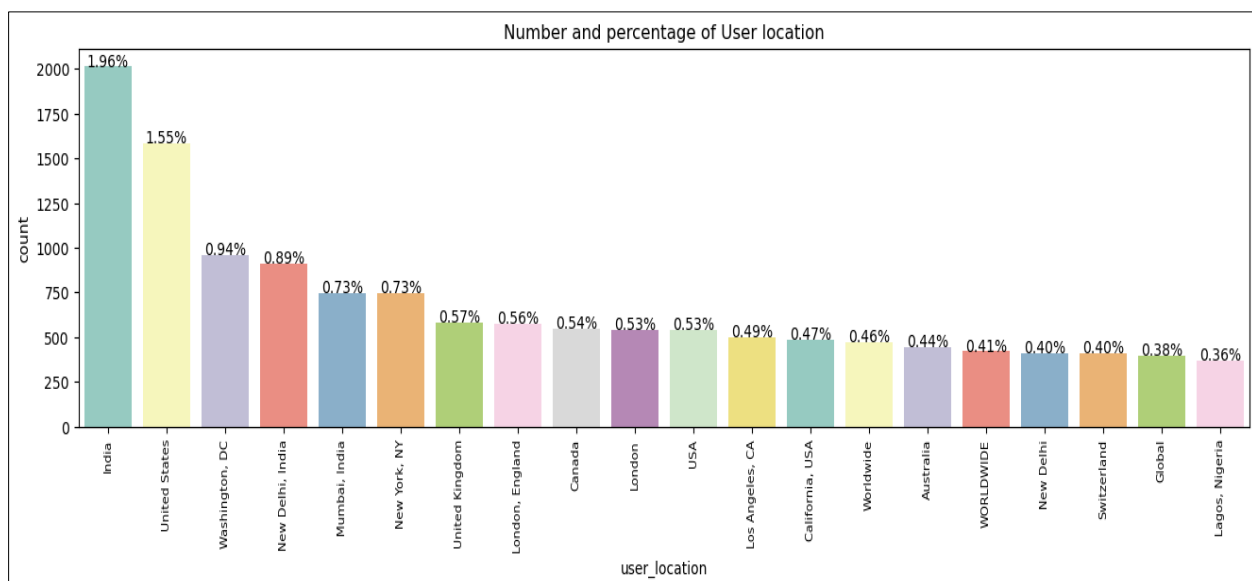


Figure 2 Top User from India: 1.96% and the second Top is USA

2.6. Word-cloud

This visualization highlights the most prevalent words within the tweets, especially from high-volume regions like India, which aids in discerning the dominant themes in the discourse. From the below figure most frequently quoted words are coronavirus and in India it is shown that COVID19 is the most used from the word cloud. Rahman et al (2023)

interprets the bitcoin miners who consider the Covid protocol for their sitting arrangement during mining process [25,33]. Another study reveals that when they try to consider process for their jute mills covid protocol was the significant factor [26].

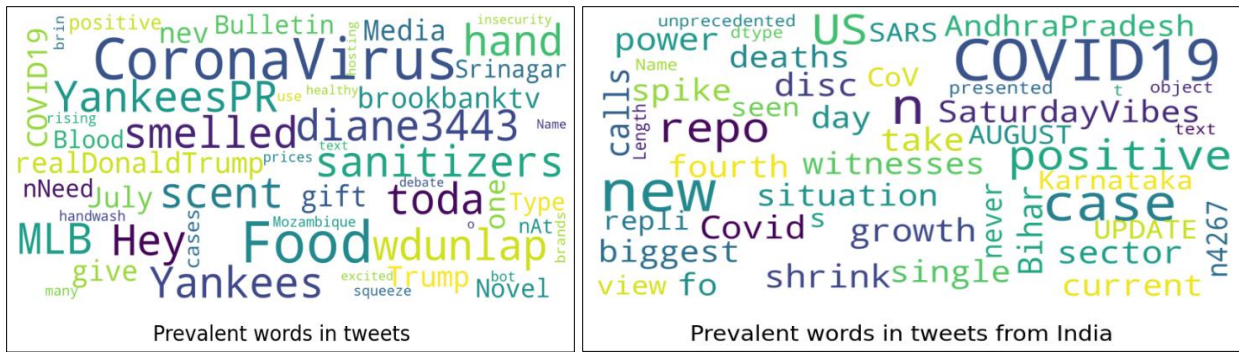


Figure 3 Prevalent words from tweets and from India

2.7. Choropleth map's

The geographic distribution of words is visualized in a choropleth map, where the density or frequency of certain terms used in tweets is represented by varying colors across geographical locations. This allows for a visual interpretation of the spatial aspect of the data, highlighting regions that are more active or have different focuses in the discourse on the pandemic. The choropleth map uploaded could be described as "Fig 4" in this section. It likely represents the geographic distribution of tweet volume, or the frequency of specific words related to COVID-19, as indicated by the color gradient. The map would be interpreted according to the legend, which appears to correlate color intensity with the number of tweets. This form of visualization is a powerful tool for presenting geographic variations in data and can be particularly illustrative of how different regions contribute to the global conversation on Twitter regarding COVID-19.

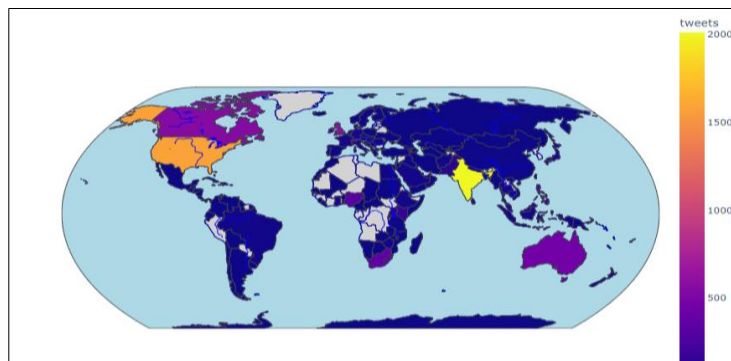


Figure 4 Geographic location for the words in choropleth

2.8. Evaluation

For stakeholders such as public health officials and policymakers, the results might indicate areas where public concern is highest, misinformation is spreading, or where more communication is needed. For researchers, these insights could signal trends worthy of deeper study, such as the relationship between public sentiment and compliance with health guidelines.

- **Public Health Communications:** Tailor messaging to address prevalent concerns and misinformation identified in tweets.
- **Policy Formulation:** Consider the sentiment and topics of discussion when creating or adjusting COVID-19 related policies.
- **Engagement Strategies:** Engage with key influencers and leverage popular hashtags to improve public health message dissemination.
- **Further Research:** Conduct longitudinal studies to see how sentiments and topics change over time, which could inform future crisis communication strategies.

The visualizations would have served as a foundation for these insights and recommendations, providing stakeholders with actionable data to inform their decisions and strategies.

3. Conclusions

This study analyzes Twitter data to grasp public sentiment on COVID-19, identifying key topics and dissemination patterns. In conclusion, this study has undertaken a comprehensive analysis of Twitter data to gain valuable insights into public sentiment regarding COVID-19. By examining a vast corpus of tweets and utilizing ISO country codes, we have successfully mapped sentiments and information dissemination patterns. This effort has allowed us to uncover not only the prevailing concerns but also the influential communication channels shaping global conversations and regional sentiment variances. The significance of these findings cannot be overstated, as they offer crucial guidance to public health bodies, policymakers, and academics. These insights enable more informed decision-making, helping to craft effective health messaging and policies that resonate with the public's voice and address the pressing issues highlighted through social media discourse during the pandemic. However, it is important to acknowledge a limitation of our research, which stems from the sheer volume of data. Despite our best efforts, we were unable to analyze every tweet in the dataset. Nonetheless, the depth and breadth of our analysis provide a robust foundation for understanding the dynamic landscape of COVID-19 conversations on Twitter and offer a valuable resource for future research and action in the field of public health communication and policy development.

Future Work

To expand the COVID-19 Twitter analysis project, incorporating diverse data like news articles and health reports, alongside advanced analytics like sentiment and network analysis, could yield deeper insights. This would enable tracking sentiment trends, topic shifts, and misinformation spread, and could measure influencer impact and the effectiveness of health responses. Access to broader data and advanced tools for processing natural language would be necessary for this enhanced analysis.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

- [1] William B. Claster, Malcolm Cooper, and Philip Sallis. 2010. Thailand – Tourism and Conflict: Modeling Sentiment from Twitter Tweets Using Naïve Bayes and Unsupervised Artificial Neural Nets. In 2010 Second International Conference on Computational Intelligence, Modelling and Simulation. 89–94. <https://doi.org/10.1109/CIMSiM.2010.98>
- [2] Orland Hoeber, Larena Hoeber, Maha El Meseery, Kenneth Odoh, and Radhika Gopi. 2016. Visual Twitter Analytics (Vista): Temporally changing sentiment and the discovery of emergent themes within sport event tweets. *Online Inf. Rev.* 40 (2016), 25–41.
- [3] Shah Rukh Humayoun, Saman Ardalan, Ragaad AlTarawneh, and Achim Ebert. 2017. TExVis: An Interactive Visual Tool to Explore Twitter Data. In *EuroVis 2017 - Short Papers*, Barbora Kozlikova, Tobias Schreck, and Thomas Wischgoll (Eds.). The Eurographics Association. <https://doi.org/10.2312/eurovisshort.20171149>
- [4] Renato Kempter, Valentina Sintsova, Claudiu Cristian Musat, and Pearl Pu. 2014. EmotionWatch: Visualizing Fine-Grained Emotions in Event-Related Tweets. In *ICWSM*.
- [5] Rahman, S. M., Ibtisum, S., Podder, P., & Hossain, S. M. (2023). Progression and Challenges of IoT in Healthcare: A Short Review. *arXiv preprint arXiv:2311.12869*.
- [6] Mondal, M.R.H.; Bharati, S.; Podder, P.; Kamruzzaman, J. Deep Learning and Federated Learning for Screening COVID-19: A Review. *BioMedInformatics* 2023, 3, 691–713. <https://doi.org/10.3390/biomedinformatics3030045>
- [7] Rahman, S. M., Ibtisum, S., Podder, P., & Hossain, S. M. (2023). Progression and Challenges of IoT in Healthcare: A Short Review. *arXiv preprint arXiv:2311.12869*.
- [8] Ibtisum, S. (2020). A Comparative Study on Different Big Data Tools.

- [9] Shenghua Liu, Wenjun Zhu, Ning Xu, Fangtao Li, Xue-qi Cheng, Yue Liu, and Yuanzhuo Wang. 2013. Co-Training and Visualizing Sentiment Evolvment for Tweet Events. In Proceedings of the 22nd International Conference on World Wide Web (Rio de Janeiro, Brazil) (WWW '13 Companion). Association for Computing Machinery, New York, NY, USA, 105–106. <https://doi.org/10.1145/2487788.2487836>
- [10] Yafeng Lu, Xia Hu, Feng Wang, Shamanth Kumar, Huan Liu, and Ross Maciejewski. 2015. Visualizing Social Media Sentiment in Disaster Scenarios. In Proceedings of the 24th International Conference on World Wide Web (Florence, Italy) (WWW '15 Companion). Association for Computing Machinery, New York, NY, USA, 1211–1215. <https://doi.org/10.1145/2740908.2741720>
- [11] Alex Godwin, Yongxin Wang, and John T. Stasko. 2017. TypoTweet Maps: Characterizing Urban Areas through Typographic Social Media Visualization. In Proceedings of the Eurographics/IEEE VGTC Conference on Visualization: Short Papers (Barcelona, Spain) (EuroVis '17). Eurographics Association, Goslar, DEU,25–29. <https://doi.org/10.2312/eurovisshort.20171128>
- [12] Thomas Kraft, Derek X. Wang, Jeffrey Delawder, Wenwen Dou, Li Yu, and William Ribarsky. 2013. Less After-the-Fact: Investigative visual analysis of events from streaming twitter. In 2013 IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV). 95–103.
- [13] Alan M. MacEachren, Anuj Jaiswal, Anthony C. Robinson, Scott Pezanowski, Alexander Savelyev, Prasenjit Mitra, Xiao Zhang, and Justine Blanford. 2011. SensePlace2: GeoTwitter analytics support for situational awareness. In 2011 IEEE Conference on Visual Analytics Science and Technology (VAST). 181–190.
- [14] Sifat Ibtisum, Ehsan Bazgir, S M Atikur Rahman, and S. M. Saokat Hossain, "A comparative analysis of big data processing paradigms: Mapreduce vs. apache spark", World Journal of Advanced Research and Reviews, 2023, 20(01), 1089–1098.
- [15] Mondal, M. R. H., Bharati, S., Podder, P., & Podder, P. (2020). Data analytics for novel coronavirus disease. *informatics in medicine unlocked*, 20, 100374.
- [16] Dennis Thom, Robert Krüger, Thomas Ertl, Ulrike Bechstedt, Axel Platz, Julia Zisgen, and Bernd Volland. 2015. Can twitter really save your life? A case study of visual social media analytics for situation awareness. In 2015 IEEE Pacific Visualization Symposium (PacificVis). 183–190.
- [17] Podder, P., Khamparia, A., Mondal, M. R. H., Rahman, M. A., & Bharati, S. (2021). Forecasting the Spread of COVID-19 and ICU Requirements.
- [18] Shahriar, K. T., Islam, M. N., Moni, M. A., & Sarker, I. H. (2023). A Dynamic Topic Identification and Labeling Approach for COVID-19 Tweets. In *Applied Intelligence for Industry 4.0* (pp. 227-239). Chapman and Hall/CRC.
- [19] Oakey, D., Jones, C., & O'Halloran, K. L. (2022). Phraseology and imagery in UK public health agency COVID-19 tweets. *Discourses, Modes, Media and Meaning in an Era of Pandemic*, 89.
- [20] Mir, A. A., & Sevukan, R. (2022). Sentiment analysis of Indian Tweets about Covid-19 vaccines. *Journal of Information Science*, 01655515221118049.
- [21] Humayoun, S. R., & Chimote, S. (2022, June). An Interactive Visualization Tool to Explore People's Tweets towards COVID-19. In Proceedings of the 2022 International Conference on Advanced Visual Interfaces (pp. 1-3).
- [22] Malla, S., & Alphonse, P. J. A. (2022). An improved machine learning technique for identify informative COVID-19 tweets. *International Journal of System Assurance Engineering and Management*, 1-12.
- [23] <https://www.kaggle.com/code/gpreda/coronavirus-covid-19-tweets/input?select=wikipedia-iso-country-codes.csv> (Accessed Online 17 December, 2023)
- [24] Ahmmed, S., Podder, P., Mondal, M. R. H., Rahman, S. A., Kannan, S., Hasan, M. J., ... & Prosvirin, A. E. (2023). Enhancing Brain Tumor Classification with Transfer Learning across Multiple Classes: An In-Depth Analysis. *BioMedInformatics*, 3(4), 1124-1144.
- [25] Rahman, S. A., Siddique, I. M., & Smith, E. D. (2023). Analyzing Bitcoin's Decentralization: Coefficient of Variation Approach and 21 million Divisibility. *Advancement of IoT in Blockchain Technology and its Applications*, 2(3), 8-17.
- [26] Shakil, M., Ullah, M. R., & Lutfi, M. (2013). Process flow chart and factor analysis in production of jute mills. *Journal of*
- [27] Hossain, Md Zakir, et al. "Evaluating the Effectiveness of a Portable Wind Generator that Produces Electricity using Wind Flow from Moving Vehicles." *Journal of Industrial Mechanics* 8.2 (2023): 44-53.

- [28] Rahman, S. A., and S Shohan (2015). Supplier selection using fuzzy-topsis method: A case study in a cement industry, *IASET: Journal of MechanicalEngineering*,4(1).31-42.
- [29] M Nazma, M Rahman and UK Dey, Comparative Analysis of AHP and Fuzzy-AHP in Supplier Selection: A Case Study on a Cement Industry. *International Conference on Mechanical, Industrial and Energy Engineering*, 2014.
- [30] Fayshal, M. A., Ullah, M. R., Adnan, H. F., Rahman, S. A., & Siddique, I. M. Evaluating Multidisciplinary Approaches within an Integrated Framework for Human Health Risk Assessment., *Journal of Environmental Engineering and Studies*, 8(3), 30-41.
- [31] Md Rahamat Ullah, et al. (2023). Optimizing Performance: A Deep Dive into Overall Equipment Effectiveness (OEE) for Operational Excellence, *Journal of Industrial Mechanics*, 8(3), 26-40.
- [32] Ullah, M. R., Molla, S., Siddique, I. M., Siddique, A. A., & Abedin, M. M (2023). Manufacturing Excellence Using Line Balancing & Optimization Tools: A Simulation-based Deep Framework., *Journal of Modern Thermodynamics in Mechanical System*, 5(3), 8-22.
- [33] Iqtiaar Md Siddique, Anamika Ahmed Siddique, Eric D Smith, Selim Molla. (2023). Assessing the Sustainability of Bitcoin Mining: Comparative Review of Renewable Energy Sources. *Journal of Alternative and Renewable Energy Sources*,10(1),1-12.
- [34] Selim Molla, Md Minhajul Abedin and Iqtiaar Md Siddique. (2023). Exploring the versatility of medical textiles: Applications in implantable and non-implantable medical textile. *World Journal of Advanced Research and Reviews*, 2024, 21(01), 603–615. DOI: 10.30574/wjarr.2024.21.1.0058.