White-Box Machine learning approach to Plasmodium falciparum Malaria Detection Using Multi-Channel Colour Scheme

Nancy Ipole
Department of Computer Science
Bingham University
Karu, Nigeria
nancy.ipole@binghamuni.edu.ng

Oluwasegun Ishaya Adelaiye
Department of Computer Science
Bingham University
Karu, Nigeria
oluwasegun.adelaiye@binghamuni.edu.ng
0000-0002-4193-5641

Musa Yusuf
Department of Computer Science
Bingham University
Karu, Nigeria
yusufmusa@binghamuni.edu.ng
0000-0002-7893-9531

Anah Bijik Hassan
Department of Computer Science
Bingham University
Karu, Nigeria
annebjk@binghamuni.edu.ng
0000-0001-6718-1078

Adamu Sulaiman Usman
Department of Computer Science
Bingham University
Karu, Nigeria
adodass@binghamuni.edu.ng
0000-0002-4917-1455

Yakubu Aliyu Ibrahim
Department of Computer Science
Bingham University
Karu, Nigeria
yakubu.ibrahim@binghamuni.edu.ng
0000-0002-5529-5243

Ageebee Silas Faki
Department of Computer Science
Bingham University
Karu, Nigeria
faki.silas@binghamuni.edu.ng
0000-0003-1531-0119

Jenom Ezekiel Maikori
Department of Computer Science
Bingham University
Karu, Nigeria
jenommaikori@binghamuni.edu.ng

***Abstract*—Malaria, a severe global health threat, remains a challenge despite ongoing efforts. This paper reviews various studies, highlighting disease severity and combat strategies. The substantial public health burden of malaria is evident, affecting millions globally. In 2019, 409,000 deaths were reported, 67% being children, necessitating preventive measures. Various strategies, such as laboratory networks, computing-based methods, and Geographic Information Systems, target malaria's impact. Despite progress, malaria persists in West Africa, requiring significant support and research. The research utilizes multi-channel colour scheme in generating key features. This paper focuses on white box algorithms for malaria diagnosis which aligns with providing a transparent and accurate healthcare solutions. With F1-score of 0.95 for "Infected," 0.95 for "Uninfected," and overall accuracy of 95.42%, the model detects Plasmodium falciparum. The study underscores comprehensive strategies to combat malaria's global burden. The solution in addition to aiding in accurate diagnosis of malaria also presents a medically transparent and trusted solution.**

Keywords—Malaria Diagnosis, Machine Learning, Explainable Artificial Intelligence, Colour Scheme, Blood Smears, Plasmodium Falciparum.

I. INTRODUCTION

Malaria is a severe human parasite illness that remains a significant concern globally. Various studies by researchers like [1] and [2] have highlighted the disease's seriousness and the ongoing efforts to combat it. Malaria imposes a substantial public health burden, affecting millions worldwide. In 2019, [3] reported 409,000 malaria-related deaths, with a concerning 67% being children, underscoring the vulnerability of young populations and the need for effective preventive measures.

Researchers have explored different strategies to mitigate malaria's impact. [3] discussed the potential of a laboratory network for improved early detection and targeted treatment. [4] suggested using computing-based methods like blood smear image analysis for enhanced diagnosis accuracy. [5] proposed a Geographic Information System (GIS) approach in India for localized malaria control, which could be applicable to similar challenges elsewhere.

Despite progress, malaria remains a major concern in West Africa and developing nations, with [6] highlighting its ongoing impact as a leading cause of death. A variant of the malaria parasite known as Plasmodium Falciparum is tagged the most dangerous [7]. Limited resources and infrastructure in some areas make control and prevention challenging, necessitating ongoing support. The COVID-19 pandemic added complexities, as patients with both diseases faced higher mortality risks, stressing the need to address co-infections [17].

The fight against malaria requires a comprehensive approach involving research, technology, healthcare infrastructure, and global collaboration. While advancements have been made in diagnosis, treatment, and prevention, malaria remains a pressing global health issue. With the use of Artificial Intelligence solutions, the black-box approach of deep learning models raises concerns about transparency and explainability which is key for medical diagnosis [4]. This paper focuses on implementing a white box machine learning algorithms in healthcare, specifically for malaria diagnosis. This integration aims to enhance doctors' work by providing transparent, explainable, and accurate diagnostics. Transparency and explainability are crucial in healthcare algorithms to foster trust and informed treatment decisions, ultimately improving patient care.

The subsequent sections of this paper are structured as follows: Section two reviews the literature, Section three outlines the methodology, and Section four presents results and discussions. The paper concludes in Section five.

II. LITERATURE REVIEW

Malaria, caused by Plasmodium parasites transmitted by infected mosquitoes, remains a global health threat. Parasites undergo a complex life cycle in humans, multiplying in the liver and infecting red blood cells. Symptoms include recurring fever, chills, headache, and muscle aches resembling flu-like illness, known as malaria "paroxysms." Severity varies by Plasmodium species and immune response. Plasmodium falciparum is the most dangerous, and responsible for most malaria-related deaths globally [7].

Cerebral malaria, a severe complication of malaria, is most pronounced with Plasmodium falciparum infection. Parasites sequester in brain blood vessels, causing neurological symptoms and possible fatality [8]. Pregnant women and children are more susceptible due to weaker immunity [9].

[4] and [10] both address the need for improved malaria diagnostic approaches. [4] adopted Convolutional Neural Network (CNN)-based deep learning models, achieving an accuracy of 95.03%. In contrast, [10] used the Grad-CAM technique to generate heatmaps with an impressive 99.9% accuracy. While both approaches show promising results, the black-box nature of deep learning models in [4] work, raises concerns about the explainability of the approach, while the utilization of Grad-CAM by [10], provides insights into decision-making but lacks complete understanding.

[11] and [12] focus on achieving high accuracy using ensemble models. [11] combined pre-trained CNN models (VGG19, ResNet50, MobileNetV2) to achieve a 99% accuracy, demonstrating the power of transfer learning. Similarly, [12] used a CNN-based approach with microscopic blood smears, achieving a 96.97% accuracy. Both studies highlight the potential of ensemble methods, but the interpretability of black-box algorithms remains a challenge.

[13] and [11] presented the limitations of existing diagnostic methods while proposing new approaches. [13] emphasized the need for an automated and accurate diagnosis to combat malaria's impact, highlighting the challenge of biased and error-prone approaches by clinical experts. On the other hand, [11] emphasized the drawbacks of Rapid Diagnostic Test (RDT) and Polymerase Chain Reaction (PCR) techniques. Both studies underline the significance of advanced AI-based diagnostic tools to overcome these limitations.

[4], [10], and [13] employed various computer vision and deep learning techniques for malaria diagnosis. [4] utilized light microscopy, filters for pre-processing, and CNN-based deep learning, achieving a high accuracy of 95.03%. [10] focused on Grad-CAM for heatmap generation with a remarkable 99.9% accuracy. [13] work involved an innovative computer-assisted diagnosis system using improved k-SVD denoising, fuzzy type II-based segmentation, and local and global feature extraction, achieving an accuracy of 98.02%. While each study demonstrates effectiveness, they utilize different techniques to tackle malaria diagnosis challenges.

[13] highlighted the importance of classifying various species of malaria for comprehensive diagnosis. This aspect is essential in understanding the malaria disease variations and guiding appropriate interventions. While other studies [4],[11],[12], and [10] focused on overall malaria diagnosis, [13] work emphasized species classification as crucial in tailoring treatments and prevention strategies for specific malaria types.

Decision trees (DT) offer transparent decision-making by visually depicting the path from root to leaf nodes, aiding understanding [14]. This clarity highlights influential features and the prediction process, enabling debugging and insights [14]. When errors occur, tracing the decision path helps pinpoint problematic features [14]. DT aid fairness and bias reduction [15], [16], thus finding utility where interpretability is vital.

III. METHODOLOGY

The project adopts Agile methodology, known for its systematic approach to ensure effective and reliable solution development. Agile's stages are designed for seamless implementation, making it well-suited due to its adaptability. This approach allows quick responses to changing requirements and continuous performance improvement. The project's execution follows Agile Principles Figure 1, ensuring an evolving, dependable tool for malaria parasite detection in blood smear images using DT. The following sections detail these stages, explaining their components and contributions to the research objective.

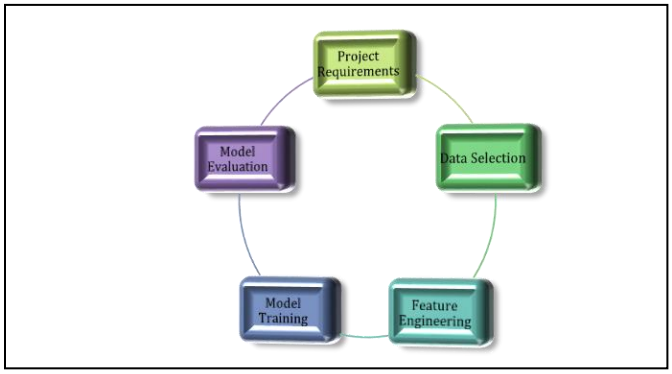


Fig 1 *Adapted Agile methodology*

The training dataset was acquired from the National Library of Medicine, NIH, USA, including 150 infected and 50 uninfected patient's thick blood smear images. Derived from white blood cells via thick smears, this dataset aligns with the project's malaria detection focus, serving as a valuable resource for model training. Figure 2 and 3 presents an infected blood smear image and an uninfected blood smear image respectively.

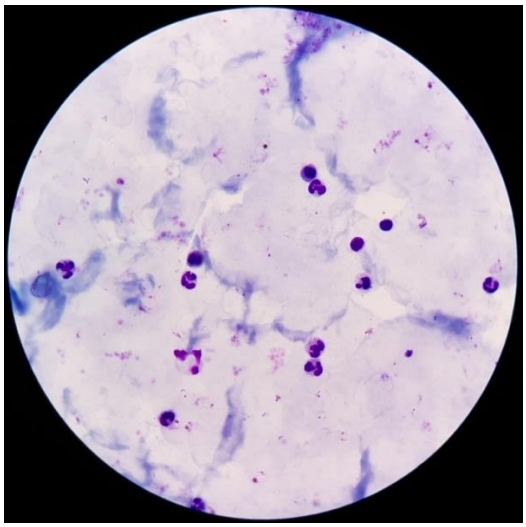


Fig 2 *Infected cell image*

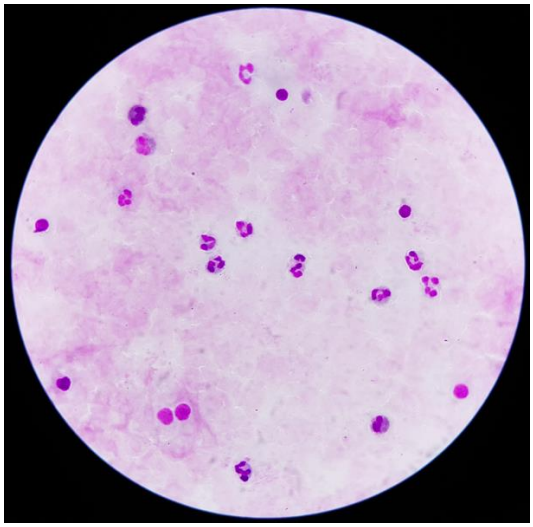


Fig *3 Uninfected cell image*

Using various color spaces (BGR, HSV, CIE LAB, and Grayscale) offers a diverse array of color representations for blood smear images. Each space captures distinct data about color, texture, and intensity. BGR and grayscale color spaces are displayed in Figures 4 and 5, showcasing the utilized representations.

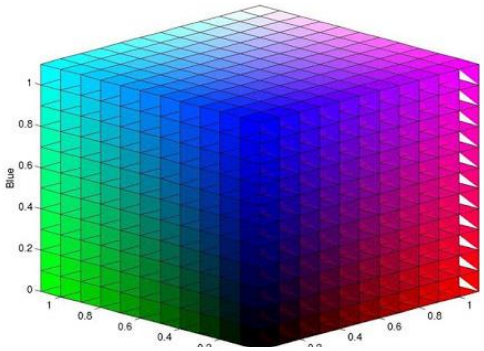


Fig *4 BGR Colour Space (Source: Dynamsoft)*

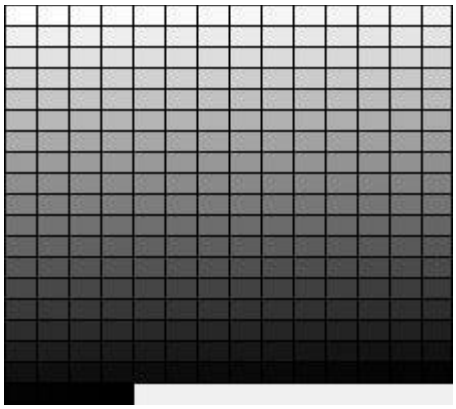


Fig *5 Grayscale Colour Space (Source: AI Shack)*

From the color channels, essential statistical metrics like mean, minimum, and standard deviation are computed, serving as key features for the blood smear images. These metrics provide insights into image attributes, such as brightness and color variation. The pixel values of an image presented as a matrix *I*, with *I(i,j)* denoting the pixel value at the specific position *(i,j)* within the image. Equations 1, 2, and 3 provide details of the formulas for these calculations.

$$\text{Mean } (\mu) = (\Sigma I(i,j)) \,/\, (\text{width} * \text{height}) \quad (1)$$

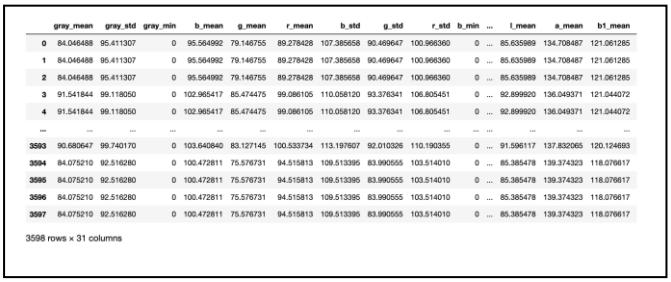
$$\text{Minimum (min)} = \min(I(i,j)) \quad (2)$$

$$\text{Standard Deviation } (\sigma) = \sqrt{\frac{((\Sigma(I(i,j) - \mu)\text{^}2)}{(\text{width} * \text{height}))}} \quad (3)$$

IV. RESULTS AND DISCUSSION

The Python script utilizes the OpenCV library to extract features from blood smear images stored in the 'Uninfected' and d ‘infected’ directories. It employs different color spaces (gray, BGR, HSV, CIE LAB) to compute statistics such as mean, standard deviation, and minimum pixel values for each color channel. These features are then stored in a CSV file named 'uninfected.csv' and ‘infected.csv’ converting the images seen in Figure 2 to its numerical output seen in Figure 4. The script serves as a crucial step in a larger project involving malaria

parasite detection, where the extracted features would likely be used to train a machine learning model for accurate diagnosis.



	gray_mean	gray_std	gray_min	b_mean	g_mean	r_mean	b_std	g_std	r_std	b_min	...	l_mean	a_mean	b1_mean
0	84.046488	95.411307	0	95.564992	79.146755	89.278428	107.385658	90.469647	100.966360	0	...	85.635989	134.708487	121.061285
1	84.046488	95.411307	0	95.564992	79.146755	89.278428	107.385658	90.469647	100.966360	0	...	85.635989	134.708487	121.061285
2	84.046488	95.411307	0	95.564992	79.146755	89.278428	107.385658	90.469647	100.966360	0	...	85.635989	134.708487	121.061285
3	91.541844	99.118050	0	102.965417	85.474475	99.086105	110.058120	93.376341	106.805451	0	...	92.899920	136.049371	121.044072
4	91.541844	99.118050	0	102.965417	85.474475	99.086105	110.058120	93.376341	106.805451	0	...	92.899920	136.049371	121.044072
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
3593	90.680647	99.740170	0	103.640840	83.127145	100.533734	113.197607	92.010326	110.190355	0	...	91.596117	137.832065	120.124693
3594	84.075210	92.516280	0	100.472811	75.576731	94.515813	109.513395	83.990555	103.514010	0	...	85.385478	139.374323	118.076617
3595	84.075210	92.516280	0	100.472811	75.576731	94.515813	109.513395	83.990555	103.514010	0	...	85.385478	139.374323	118.076617
3596	84.075210	92.516280	0	100.472811	75.576731	94.515813	109.513395	83.990555	103.514010	0	...	85.385478	139.374323	118.076617
3597	84.075210	92.516280	0	100.472811	75.576731	94.515813	109.513395	83.990555	103.514010	0	...	85.385478	139.374323	118.076617

3598 rows × 31 columns

Fig 6 Output from feature extraction of Malaria cell images

The confusion matrix shows the performance of a classification model with 344 true positive predictions and 343 true negatives, along with 22 false negatives and 11 false positives. The precision, recall, and F1-score for both "Infected" and "Uninfected" classes are reported, with overall accuracy at 95.42%. The macro average F1-score is 0.95, and the weighted average is also 0.95, indicating good overall model performance. The execution time is 12.07 seconds. The results are presented in Table 1.

WHITE-BOX APPROACH RESULTS USING DT

Class	Precision	Recall	F1-Score	Accuracy
Infected	0.94	0.97	0.95	
Uninfected	0.97	0.94	0.95	
Macro Avg	0.95	0.95	0.95	95.42%
Weighted Avg	0.95	0.95	0.95	

The classification results, particularly the precision, recall, and F1-scores, are crucial when considering the detection of specific malaria species like Plasmodium falciparum. Given that Plasmodium falciparum is one of the most dangerous and life-threatening species of malaria, achieving high precision is paramount to accurately identifying its presence. The high precision for the "Infected" class (0.94) suggests that when the model predicts an image as infected, it is often correct, minimizing false positives. Similarly, the high recall for the "Uninfected" class (0.94) indicates the model's ability to correctly capture most instances of uninfected cases, minimizing false negatives. These metrics are crucial for detecting Plasmodium falciparum, as its accurate identification is vital for timely and effective medical interventions. The overall balanced performance and consistent F1 scores further underscore the model's potential for reliable Plasmodium falciparum detection.

V. CONCLUSION

In conclusion, this research has illuminated the gravity of malaria as a global health concern, shedding light on its impact and the efforts aimed at combating it. The diverse approaches highlighted in the reviewed studies, encompassing technological advancements, diagnostic methods, and healthcare infrastructure improvements, underscore the multidimensional nature of the fight against malaria. Despite considerable progress, the disease remains a significant challenge in vulnerable regions, necessitating ongoing support and investment. The interplay between malaria and other health crises, exemplified by the COVID-19 co-infection complexities, further emphasizes the need for comprehensive strategies. The utilization of machine learning algorithms for malaria diagnosis, as pursued in this research, provides both enhanced accuracy and transparency in healthcare decision-making. The integration of white box machine learning algorithms aligns well with the imperative for understandable and reliable diagnostic tools, contributing to the collective efforts in alleviating the global burden of malaria.

This paper presents an innovative approach to malaria detection, emphasizing transparency and interpretability in medical artificial intelligence-based solutions. It achieves an impressive 95.42% accuracy, emphasizing its potential for malaria diagnosis focusing on the most dangerous variant plasmodium falciparum using blood smear images and presented an easy method for early detection which is key in reducing the mortality rate. This endeavor not only advances malaria detection but also sets a precedence for interpretable medical artificial intelligence-based solutions, contributing to healthcare trust and acceptance. Future work will require the utilization of multiple datasets for a comprehensive evaluation and a real-time implementation in validating the solution.

REFERENCES

[1] F. Nosten, D. Richard-Lenoble, M. D.-L. P. Médicale, and undefined 2022, "A brief history of malaria," Elsevier, Accessed: Aug. 11, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0755498222000239

[2] J. A. Garrido-Cardenas, L. González-Cerón, F. García-Maroto, J. Cebrián-Carmona, F. Manzano-Agugliaro, and C. M. Mesa-Valle, "Analysis of fifty years of severe malaria worldwide research," mdpi.comJA Garrido-Cardenas, L González-Cerón, F García-Maroto, J Cebrián-CarmonaPathogens, 2023•mdpi.com, 2023, doi: 10.3390/pathogens12030373.

[3] J. Yin, M. Li, H. Yan, S. Zhou, Z. X.-F. of Medicine, and undefined 2022, "Laboratory diagnosis for malaria in the elimination phase in China: efforts and challenges," SpringerJ Yin, M Li, H Yan, S Zhou, Z XiaFrontiers of Medicine, 2022•Springer, vol. 16, no. 1, pp. 10–16, Feb. 2022, doi: 10.1007/s11684-021-0889-7.

[4] S. Shambhu, D. Koundal, P. Das, … V. H.-C., and undefined 2022, "Computational methods for automated analysis of malaria parasite using blood smear images: recent advances," hindawi.comS Shambhu, D Koundal, P Das, VT Hoang, K Tran-Trung, H TurabiehComputational Intelligence and Neuroscience, 2022•hindawi.com, Accessed: Aug. 11, 2023. [Online]. Available: https://www.hindawi.com/journals/cin/2022/3626726/

[5] B. S. Baliga et al., "Indigenously developed digital handheld Android-based Geographic Information System (GIS)-tagged tablets (TABs) in malaria elimination programme in Mangaluru city, Karnataka, India," Malar J, vol. 18, no. 1, Dec. 2019, doi: 10.1186/S12936-019-3080-8.

[6] J. Novignon, G. Aryeetey, J. Nonvignon, … K. M.-H., and undefined 2023, "Efficiency of malaria service delivery in selected district-level hospitals in Ghana," Taylor & FrancisJ Novignon, G Aryeetey, J Nonvignon, K Malm, NY Peprah, SA Agyemang, S AmonHealth Systems, 2023•Taylor & Francis, vol. 12, no. 2, pp. 198–207, 2021, doi: 10.1080/20476965.2021.2015251.

[7] K. Adhikary, A. Chatterjee, S. C.- Viral, undefined Parasitic, undefined Bacterial, and undefined 2023, "Malaria: Epidemiology, pathogenesis,

and therapeutics," Elsevier, Accessed: Aug. 11, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780323857307000229

[8] O. A. Ndunge, N. Kilian, M. S.-A. Science, and undefined 2022, "Cerebral Malaria and Neuronal Implications of Plasmodium Falciparum Infection: From Mechanisms to Advanced Models," Wiley Online Library, vol. 9, no. 36, Dec. 2022, doi: 10.1002/advs.202202944.

[9] A. K. Heuschen et al., "Impact of the COVID-19 pandemic on malaria cases in health facilities in northern Ghana: a retrospective analysis of routine surveillance data," Malar J, vol. 21, no. 1, Dec. 2022, doi: 10.1186/S12936-022-04154-1.

[10] M. Islam, M. Nahiduzzaman, M. Goni, A. S.- Sensors, and undefined 2022, "Explainable transformer-based deep learning model for the detection of malaria parasites from blood cell images," mdpi.comMR Islam, M Nahiduzzaman, MOF Goni, A Sayeed, MS Anower, M Ahsan, J HaiderSensors, 2022•mdpi.com, Accessed: Aug. 11, 2023. [Online]. Available: https://www.mdpi.com/1424-8220/22/12/4358

[11] M. Alnussairi, A. İ.-C. and E. Engineering, and undefined 2022, "Malaria parasite detection using deep learning algorithms based on (CNNs) technique," Elsevier, Accessed: Aug. 11, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790622005389

[12] D. Uzun Ozsahin, M. Taiwo Mustapha, B. Bartholomew Duwa, and I. Ozsahin, "Evaluating the performance of deep learning frameworks for malaria parasite detection using microscopic images of peripheral blood smears," mdpi.comD Uzun Ozsahin, MT Mustapha, B Bartholomew Duwa, I OzsahinDiagnostics, 2022•mdpi.com, 2022, doi: 10.3390/diagnostics12112702.

[13] G. Madhu, "Computer Vision and Machine Learning Approach for Malaria Diagnosis in Thin Blood Smears from Microscopic Blood Images," pp. 191–209, 2020, doi: 10.1007/978-981-15-3689-2_8.

[14] V. G. Costa and C. E. Pedreira, "Recent advances in decision trees: an updated survey," Artif Intell Rev, vol. 56, no. 5, pp. 4765–4800, May 2023, doi: 10.1007/S10462-022-10275-5.

[15] H. Chen, H. Zhang, … D. B.-… C. on M., and undefined 2019, "Robust decision trees against adversarial examples," proceedings.mlr.press Chen, H Zhang, D Boning, CJ HsiehInternational Conference on Machine Learning, 2019•proceedings.mlr.press, Accessed: Aug. 11, 2023. [Online]. Available: http://proceedings.mlr.press/v97/chen19m.html?ref=https://githubhelp.com

[16] X. Meng, P. Zhang, Y. Xu, & H. X.-I. J. of E. P., and undefined 2020, "Construction of decision tree based on C4. 5 algorithm for online voltage stability assessment," Elsevier, Accessed: Aug. 11, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0142061519324251

[17] Azuaba, E., Samson, B. J., Eseigbe, E. E., Abraham, T. J., & Musa, Y. (2021). Modeling the Risk As-sessment of COVID-19 Pandemic in Bingham University of Nigeria. Int J Clin Biostat Biom, 7, 039.