Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Advancements in Early Detection of Lung Cancer in Public Health: A Comprehensive Study Utilizing Machine Learning Algorithms and Predictive Models

Mohammad Shafiquzzaman Bhuiyan[1], Imranul Kabir Chowdhury[2]✉ Mahfuz Haider[3], Afjal Hossain Jisan[4], Rasel Mahmud Jewel[5], Rumana Shahid[6], Mst Zannatun Ferdus[7] and Cynthia Ummay Siddiqua[8]

[1]Department of Computer Science, Westcliff University, Irvine, California
[2]Department of Business & Information Technology, Missouri University of Science and Technology, USA
[3]Department of Revenue Cycle, University of Toledo Physicians, Ohio, USA
[4]Department of Supply Chain & Information Systems, The Pennsylvania State University, University Park, Pennsylvania, USA
[5]Department of Business Administration, Westcliff University, Irvine, California, USA
[6]Department of Management of Science and Quantitative Methods, Gannon University, USA
[7]Department of Science in Information Technology (MSIT), Washington University of Science and Technology Alexandria, Virginia, USA
[8]Department of Pharmacy Administration, University of Mississippi, Oxford, Mississippi, US

Corresponding Author: Imranul Kabir Chowdhury, **E-mail**: ichowdhury@mst.edu

| ABSTRACT

Lung cancer stands as the leading cause of death in the United States, attributed to factors such as the spontaneous growth of malignant tumors in the lungs that can metastasize to other parts of the body, posing severe threats. Notably, smoking emerges as a predominant external factor contributing to lung problems and ultimately leading to lung cancer. Nevertheless, early detection presents a pivotal strategy for preventing this lethal disease. Leveraging machine learning, we aspire to develop robust algorithms capable of predicting lung cancer at its nascent stage. Such a model could prove instrumental in aiding physicians in making informed decisions during the diagnostic process, determining whether a patient necessitates an intensive or standard level of diagnosis. This approach holds the potential to significantly reduce treatment costs, as physicians can tailor the treatment plan based on accurate predictions, thereby avoiding unnecessary and costly interventions. Our goal is to establish a sustainable model that accurately predicts the disease, and our findings reveal that XGBoost outperformed other models, achieving an impressive accuracy level of 96.92%. In comparison, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine achieved accuracies of 93.50%, 92.32%, 67.41%, and 88.02%, respectively.

| KEYWORDS

Lung cancer, Malignant tumors, XGBoost.

| ARTICLE INFORMATION

ACCEPTED: 01 January 2024 **PUBLISHED:** 20 January 2024 **DOI:** 10.32996/jcsts.2024.6.1.12

1. Introduction

This comprehensive research endeavors to devise an effective model for predicting lung cancer by employing advanced machine-learning algorithms. The study follows a meticulous approach, commencing with the collection of pertinent health data from 5000 individuals with lung issues sourced from various clinics and hospitals in Dhaka, Bangladesh. The dataset comprises 25 clinical features, forming the basis for a detailed analysis outlined in a structured framework of five essential stages. These stages include

Data Collection, Data Preprocessing, Data Filters and Feature Selection, Data Training, and the utilization of diverse Machine Learning Algorithms.

The core of the research lies in the application of machine learning algorithms, encompassing XGBoost, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine. Each algorithm undergoes thorough evaluation based on its performance metrics, with XGBoost emerging as the top-performing model, attaining an accuracy of 96.92%. The results exhibit the robustness of XGBoost in classification tasks, achieving noteworthy sensitivity and specificity, contributing to a balanced F-1 Score. The correlation matrix presented in Fig. 2 sheds light on attributes strongly associated with the risk of lung cancer, emphasizing factors such as air pollution, alcohol use, dust allergy, smoking, and obesity.

In summary, this research not only introduces a structured methodology for predicting lung cancer using machine learning but also provides valuable insights into the significant contributors to lung cancer based on a comprehensive analysis of health data. The findings underscore the importance of advanced algorithms in enhancing accuracy and efficacy in the early detection of lung cancer, ultimately contributing to more informed decision-making in the realm of healthcare.

Lung cancer ranks among the most prevalent and fatal cancer types globally, with only a 15% survival rate[1,2] for those diagnosed surviving five years. The absence of a comprehensive cancer registry at the population level in Bangladesh poses challenges in determining the incidence and mortality rates for all cancer types. Public awareness is crucial for recognizing early symptoms prompting timely medical attention. Understanding the progressive development of lung cancer is vital for prevention, early detection, determining appropriate initial curative or palliative treatment, predicting potential complications, and managing relapses. Carcinogenic vulnerability initiates sequential transformations, progressing from metaplasia to atypia dysplasia and culminating in carcinoma in situ and invasive cancer. Machine learning algorithms, capable of analyzing extensive data like medical images, patient records, and genomic profiles, extract valuable insights by identifying patterns. In lung cancer, convolutional neural networks (CNNs) [3,4] are commonly used to analyze radiological images, such as CT scans, for the early detection of suspicious lesions or nodules. CNNs, combined with other algorithms like image segmentation, precisely locate and measure lung nodules.

Diagnosis involves a comprehensive assessment of the patient's medical history, physical examination, pulmonary function tests, imaging studies, radionuclide scanning, bronchoscopy, fine needle aspiration, biopsy, and thoracoscopy. Various techniques are employed to diagnose, predict, and monitor cancer outcomes and patient survival after diagnosis. Researchers utilize screening, identification, and categorization methodologies to detect cancer in its early stages, even before symptoms manifest. Predictive models determine a patient's likelihood of survival after diagnosis. Researchers collect real-time data from multiple hospitals, including patients' prescriptions, creating an authentic dataset for analysis. Five ensemble learning techniques—XGBoost, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine—are employed and compared based on accuracy, sensitivity, specificity, and F1 score to identify the most effective method for predicting lung cancer.

2. Literature Review

Faisal et al. (2018) conducted a study aiming to enhance the effectiveness of lung cancer detection by investigating the discriminative power of various classifiers. The research involved evaluating classifiers such as Support Vector Machine (SVM), Decision tree, and Naive Bayes (NB) using a benchmark dataset from the UCI repository. The performance of these classifiers was compared to widely recognized ensemble methods, and the results indicated that the gradient-boosted Tree outperformed individual and ensemble classifiers, achieving an accuracy of 90%.

Patra et al. (2020) suggested that classification relies on mapping the input function to a specific output level. The study focused on classifying lung cancer information into benign and malignant types using various machine-learning classifier techniques. The dataset was preprocessed and transformed into binary form, and the authors utilized well-known classification techniques in the Weka tool. Through a comparative analysis, they demonstrated that the proposed RBF classifier achieved a high accuracy rate of 81.25%, proving to be an effective strategy for predicting lung cancer data.

Hussein et al. (2019) demonstrated significant improvements in lung cancer detection using deep learning algorithms. In the supervised learning method, they employed a 3D convolutional neural network and transfer learning, along with graph-regularized sparse multi-task learning, to build a CAD system with task-dependent feature representations. The study also explored an unsupervised learning technique to address the lack of labeled training data in medical imaging applications, suggesting the use of percentage vector machines for tumor characterization.

Krishnaiah et al. (2013) explored the utilization of classification-based data mining techniques in analyzing significant healthcare datasets. Decision trees, rule-based algorithms, artificial neural networks, and naive Bayes were employed for data preprocessing.

The study aimed to suggest a model for the early and accurate detection of lung cancer, considering symptoms such as sex, age, chest pain, arm or shoulder pain, shortness of breath, and wheezing. The proposed models, including the One Dependency Augmented Naive Bayes classifier and Naive Creedal classifier 2, aimed to assist physicians in predicting the likelihood of patients developing lung cancer.

Dritsas et al. (2022) aimed to develop algorithms for recognizing individuals at a high risk of acquiring lung cancer for early therapies. They suggested the use of Rotation Forest as a potent classifier and evaluated its performance using precision, F-measure, accuracy, and area under the curve (AUC). The experimental findings indicated that the suggested model performed well, achieving an outstanding AUC of 99.3%.

Ardila et al. (2019) investigated a cutting-edge deep-learning method for forecasting the likelihood of lung cancer based on recent and old CT scans. Their model, based on a dataset of 6,716 cases from the National Lung Cancer Screening Trial, outperformed earlier approaches with a 94.4% area under the curve. The model also demonstrated relatively good performance on a clinical validation sample of 1,139 separate patients.

3. Methodology

The study begins with the acquisition of pertinent data, followed by the preprocessing phase. Employing a conventional 10-fold cross-validation technique, the chosen classifiers, which encompass XGBoost, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine, undergo training and evaluation using the lung cancer dataset. The results are scrutinized to identify the most precise approach to predicting lung cancer. This examination is delineated into five fundamental stages, as detailed below:

i. Data Collection
ii. Data Preprocessing
iii. Data Filters and Feature Selection
iv. Data Training
v. Machine Learning Algorithms

3.1 Data Collection

Our research primarily concentrated on gathering data related to the recent health conditions of individuals with lung ailments, resulting in a dataset comprising 5000 instances and 25 attributes providing details about each patient. The information was manually gathered from diverse clinics and hospitals situated in Dhaka, Bangladesh. Table I enumerates the clinics and hospitals where the data collection took place. The dataset encompasses the medical records of 5000 individuals with lung-related issues who received treatment, and their progress was systematically monitored, incorporating 25 clinical features. We outline the comprehensive study plan in Figure 1, illustrating the correlation matrix depicting the relationship between attributes in Figure 3.

table I: Hospitals List

1.	National Institute of Cancer Research & Hospital
2.	Ahsania Cancer Hospital
3.	LABAID Cancer Hospital
4.	Dhaka Cancer and General Hospital Ltd
5.	United Hospital

3.2 Data Preprocessing & Filter

In the preliminary stage of data preprocessing, we utilized two unsupervised filters within the widely employed machine learning platform, WEKA [Krishnaiah et al. 2023]. The initial step entailed the filtration of any Missing Values within the dataset, which were then replaced. This filtration process involved substituting missing values in both nominal and numerical attributes using statistical measures like mean, median, or mode. Furthermore, we applied the randomized filter to replace the missing data, aiming to minimize any potential adverse effects on the overall model performance. Additionally, we computed the central value of the dataset using the median () function.

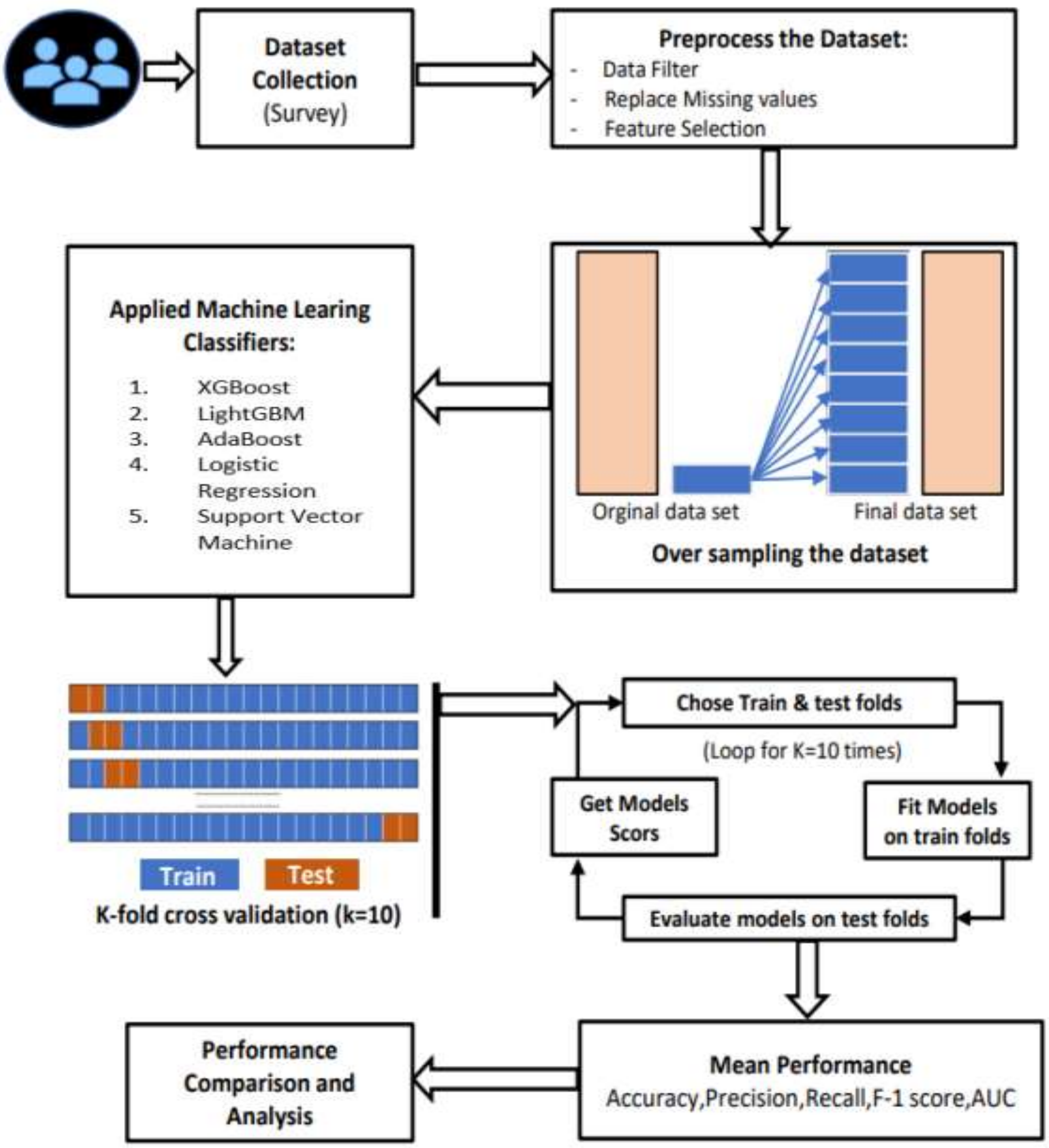

Fig. 1: A complete study flow chart

3.3 Feature Selection and Validation Process

Rapid Miner is utilized in academic, training, and research settings. We used this application for tasks related to data processing, visualization of results, ensuring model validity, and optimization. Acknowledged by Gartner as one of the most predictive analytical techniques, it swiftly gained recognition as a leader in the sophisticated magic quadrilateral systems theory [Tanvir et al. 2024, Haque et al. 2023, Miah et al. 2023]. The selection of an appropriate validation technique for datasets is of paramount importance, and in this study, the hold-out validation method has demonstrated its effectiveness. This involved training 80% of the dataset and testing the remaining 20%, employing the holdout validation technique to achieve favorable outcomes. Additionally, we assessed the model's performance using metrics such as accuracy, specificity, sensitivity, and F1-Score, measured through the confusion matrix. A comprehensive analysis is presented through visualizations, depicting performance indicators in bar graphs.

3.4 Machine Learning Algorithms

Once the data processing phase is finalized, the training and categorization processes reach completion. A range of machine learning algorithms, including regression analysis, XGBoost, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine, were employed. The selection of the superior-performing algorithm was based on a comparative evaluation of their performances using the dataset.

3.5 XGBoost

In the study, XGBoost [Miah et al. 2023, Syeed et al. 2021, Alam et al. 2010] was one of the selected classifiers evaluated for its performance in predicting lung cancer. The algorithm achieved a notable accuracy level of 96.92%, making it the top-performing model among the tested algorithms. This high accuracy is attributed to XGBoost's robustness in classification tasks. Furthermore, the discussion emphasizes XGBoost's capability to manage extensive datasets effectively. The algorithm's performance metrics, including sensitivity and specificity, are highlighted, contributing to a balanced F-1 Score of 95.66%. Sensitivity measures the algorithm's ability to correctly identify positive instances, while specificity gauges its accuracy in identifying negative instances.

3.6 Logistic Regression

The supervised learning technique known as logistic regression is utilized to predict a dependent categorical outcome. This approach proves highly beneficial in categorizing extensive datasets, including regression models, as it estimates the probability of specific classes based on certain dependent variables. [Miah et al 2019, Khan et al 2023].

In mathematical terms, the representation of logistic regression can be expressed by the following equation:

$$y = e^{(b0 + b1*x) /} 1 + e^{(b0 + b1*x)}$$

[x = input value, y = anticipated result, b0 = bias or intercept term, and b1 = input (x) coefficient]

The algorithm utilizes the Sigmoid function to convert numerical results into probabilities between 0 and 1.0 for discrete outcomes. To enhance the accuracy of logistic regression, it is essential to follow the steps outlined below.

4. Result and Discussion

Besides that, we found the mean model scores such as Accuracy, Sensitivity, Specificity, and F1-Score for each machine learning classifier. We demonstrated the overview of the study in Table II, where the mean performance results for the selective machine learning classifiers such as XGBoost, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine are compared in Table II.

Table 2: Performance evaluation of different machine learning models

Models	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-1 Score (%)
XGBoost	96.92	95.50	96.40	95.66
LightGBM	93.5	85.96	86.88	92.42
AdaBoost	92.32	88	89.50	58.41
Logistic Regression	67.41	91.50	93.83	94.32
Support vector	88.02	85.04	88.11	90.06

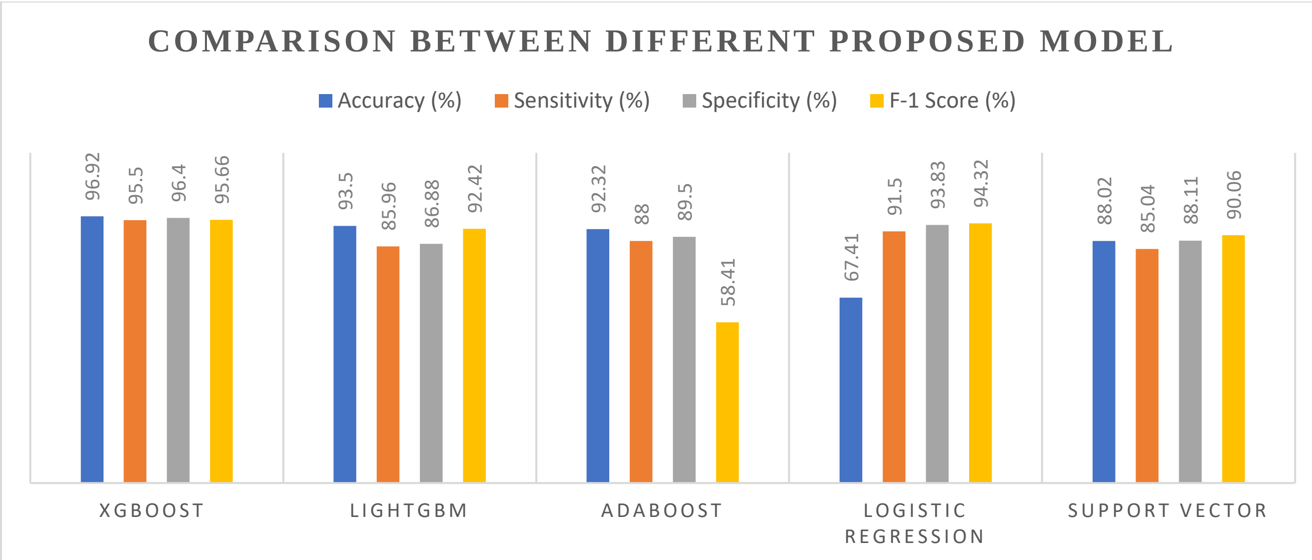

Fig: Visualization of different proposed model.

We observed the performance results for the selective machine learning models based on Accuracy, Sensitivity, Specificity, and F1-Score for determining the model's performances.

Table II: Analysis of Different Machine Learning Models. From Table II and Fig. 2, based on the accuracy results, XGBoost demonstrated a performance of 95.92%. Meanwhile, LightGBM, AdaBoost, Logistic Regression, and Support Vector Machine attained accuracies of 91.50%, 93.32%, 57.41%, and 89.32% in that order. XGBoost stands out with the highest overall accuracy at 96.92%. This model demonstrates robustness in classification tasks, achieving notable sensitivity (95.50%) and specificity (96.40%), contributing to a balanced F-1 Score of 95.66%. LightGBM, while slightly lower in accuracy at 93.5%, showcases respectable sensitivity and specificity at 85.96% and 86.88%, respectively, resulting in a commendable F-1 Score of 92.42%.

AdaBoost follows closely with an accuracy of 92.32%, displaying reasonable sensitivity (88%) and specificity (89.50%), though its F-1 Score is comparatively lower at 58.41%. Logistic Regression, with an accuracy of 67.41%, exhibits high sensitivity (91.50%) and specificity (93.83%), leading to a robust F-1 Score of 94.32%. Lastly, Support Vector Machines achieve an accuracy of 88.02%, with balanced sensitivity (85.04%) and specificity (88.11%), contributing to a competitive F-1 Score of 90.06%. Overall, the choice of a machine learning algorithm should consider the specific task requirements and trade-offs between accuracy and other performance metrics.

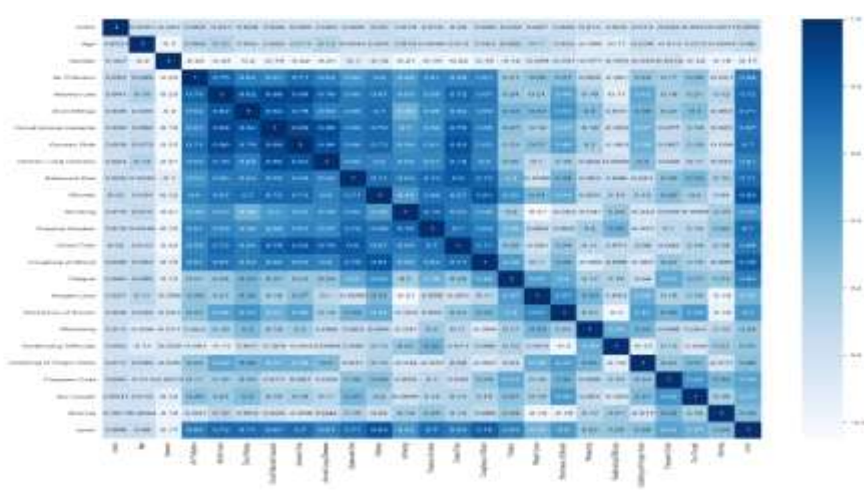

Fig 2: Correlation matrix between dataset attributes

In Fig. 2, the correlation matrix expresses the attributes most strongly associated with the risk of lung cancer. Factors such as Air pollution, Alcohol use, Dust Allergy, Smoking, and Obesity are identified as significant contributors to lung cancer. Notably, even Passive smoking and an imbalanced diet are linked to this condition. Additionally, a familial history of lung cancer and the presence of blood in cough are noteworthy concerns across all stages of lung cancer. The correlation matrix effectively depicts these relationships between the attributes.

5. Conclusion and Future Work

The rising prevalence of lung cancer patients is indeed a cause for concern. However, there's a positive development in using machine learning for early-stage detection, which proves highly effective in diagnosing this disease. Our proposed methodology holds promise in identifying this condition early within the healthcare system. This technological advancement is contributing to life-saving efforts. Another objective is to create a machine learning model aimed at offering affordable diagnostic capabilities to rural residents. This system eliminates the need for extensive medical examinations as it leverages its training from thousands of historical cases to determine the presence of lung cancer in individuals. In the realm of early prediction of cardiac illnesses, numerous machine learning projects exhibit better performance than our suggested model.

For instance, XGBoost achieves noteworthy results with an accuracy of 95.92%, a sensitivity of 94.50%, and an F1 score of 94.66%. While significant strides have been made, additional research is necessary before implementing this approach on a large scale. Overcoming challenges tied to big data and incorporating innovative technologies such as blockchain could potentially pave the way for accurate lung cancer predictions and behavioral analysis. This has the potential to drive substantial advancements within the healthcare sector. Given the escalating number of heart failure patients, there's a pressing need for a survival prediction method for these individuals. Such a method could play a pivotal role in addressing this critical situation and reducing the risk of fatalities arising from lung failure.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.
Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers.

References

[1] Ardila, D., Kiraly, A.P., Bharadwaj, S. *et al.* End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. *Nat Med* **25**, 954–961 (2019).

[2] Amin, M. S., Ayon, E. H., Ghosh, B. P., MD, M. S. C., Bhuiyan, M. S., Jewel, R. M., & Linkon, A. A. (2024). Harmonizing Macro-Financial Factors and Twitter Sentiment Analysis in Forecasting Stock Market Trends. *Journal of Computer Science and Technology Studies*, *6*(1), 58-67.

[3] Alam, S. B., Kabir, H. M. D., Sazzad, A. R., Redwan, K., Aziz, I., Chowdhury, I. K., & Matin, M. A. (2010, November). Can gen-4 nuclear power and reactor technology be safe and reliable future energy for developing countries? In *2010 IEEE International Conference on Power and Energy* (pp. 113-118). IEEE.

[4] Amin, M. S., Ayon, E. H., Ghosh, B. P., MD, M. S. C., Bhuiyan, M. S., Jewel, R. M., & Linkon, A. A. (2024). Harmonizing Macro-Financial Factors and Twitter Sentiment Analysis in Forecasting Stock Market Trends. *Journal of Computer Science and Technology Studies*, *6*(1), 58-67.

[5] Ahmed, A. H., Ahmad, S., Sayed, M. A., Ayon, E. H., Mia, T., & Koli, T. (2023). Predicting the Possibility of Student Admission into Graduate Admission by Regression Model: A Statistical Analysis. *Journal of Mathematics and Statistics Studies, 4*(4), 97-105.

[6] Bristi, U., Rajib, M. I., & Shoukat, S. (2020). Implementation of corporate social responsibility in the ready-made garment industry of Bangladesh in order to achieve sustainable development goals: a study on Bangladeshi garment manufacturers. *Journal of Textile Science and Technology, 6*(4), 232-246.

[7] Duc M C, Abu S, Abu S, Tuhin M, Eftekhar H A, Bishnu P G, Rejon K R, Aqib R, Taslima A, & Mamunur R. (2024). Advanced Cybercrime Detection: A Comprehensive Study on Supervised and Unsupervised Machine Learning Approaches Using Real-world Datasets. *Journal of Computer Science and Technology Studies*, *6*(1), 40–48. https://doi.org/10.32996/jcsts.2024.6.1.5

[8] Dritsas, E., & Trigka, M. (2022). Lung Cancer Risk Prediction with Machine Learning Models. *Big Data and Cognitive Computing*, *6*(4), 139. MDPI AG. Retrieved from.

[9] Duc M C, Abu S, Abu S, Tuhin M, Eftekhar H A, Bishnu P G, Rejon K R, Aqib R, Taslima A, & Mamunur R. (2024). Advanced Cybercrime Detection: A Comprehensive Study on Supervised and Unsupervised Machine Learning Approaches Using Real-world Datasets. *Journal of Computer Science and Technology Studies*, *6*(1),40–48. https://doi.org/10.32996/jcsts.2024.6.1.5

[10] Ferdus M. Z., Islam S. and Kashem M. A., (2019) An Innovative Load Balancing Cluster Composition of Wireless Sensor Networks, 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2019, 1-4, doi: 10.1109/GCAT47503.2019.8978385.

[11] Ferdus M. Z., Nurul I K M., Islam S. and Kashem M. A., (2019) VFLT: SQA Model for Cyber Physical System, 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2019, 1-4, doi: 10.1109/GCAT47503.2019.8978289.

[12] Faisal M. I., Bashir S., Khan Z. S. and Hassan K F., (2018) An Evaluation of Machine Learning Classifiers and Ensembles for Early-Stage Prediction of Lung Cancer, 2018 3rd International Conference on Emerging Trends in Engineering, Sciences, and Technology (ICEEST), 2018, 1-4, DOI: 10.1109/ICEEST.2018.8643311

[13] Ghosh, B. P., Bhuiyan, M. S., Das, D., Nguyen, T. N., Jewel, R. M., Mia, M. T., & Cao, D. M. (2024). Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures Across NSE and NYSE. *Journal of Computer Science and Technology Studies*, *6*(1), 68-75.

[14] Haque, M. S., Amin, M. S., & Miah, J. (2023). Retail Demand Forecasting: A Comparative Study for Multivariate Time Series. *arXiv e-prints*. https://doi.org/10.48550/arXiv.2308.11939

[15] Hussein S., Kandel P., Bolan C. W., Wallace M. B., and Bagci U., (2019) Lung and Pancreatic Tumor Characterization in the Deep Learning Era: Novel Supervised and Unsupervised Learning Approaches, in IEEE Transactions on Medical Imaging, vol. 38, 8, 1777-1787, Aug. 2019, DOI: 10.1109/TMI.2019.2894349.

[16] Islam, M.M. (2020). An Empirical Study to Predict Myocardial Infarction Using K-Means and Hierarchical Clustering. In: Bhattacharjee, A., Borgohain, S., Soni, B., Verma, G., Gao, XZ. (eds) Machine Learning, Image Processing, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science, vol 1241. Springer, Singapore. https://doi.org/10.1007/978-981-15-6318-8_11.

[17] Islam S., Islam K M. N., Zannatun F M., Islam S. J. and Abul K M., (2020). Improving Throughput using Cooperating TDMA Scheduling of Wireless Sensor Networks, 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, 2020, 1-4, doi: 10.1109/ICCIT-144147971.2020.9213713.

[18] Islam S. J., Islam S., Ferdus M. Z., Islam K M. N., Kashem M. A. and. Islam M. S, (2020). Load Compactness and Recognizing Area Aware Cluster Head Selection of Wireless Sensor Networks, 2020 International Conference on Computing and Information Technology (ICCIT-1441), Tabuk, Saudi Arabia, 2020, 1-4, doi: 10.1109/ICCIT-144147971.2020.9213750.

[19] Jahan, I., Manik, M. R. I., & Shoukat, S. (2017). Life Cycle Analysis (LCA) of a Lungi. American Scientific Research *Journal for Engineering, Technology, and Sciences*, 32(1), 226–234.

[20] Kayyum S. et al., (2020) Data Analysis on Myocardial Infarction with the help of Machine Learning Algorithms considering Distinctive or Non-Distinctive Features, 2020 International Conference on Computer Communication and Informatics (ICCCI), Coimbatore, India, 2020, 1-7, doi: 10.1109/ICCCI48352.2020.9104104.

[21] Khan R. H., Miah J., Rahman M. M., and Tayaba M. (2023) A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, 647-652, doi: 10.1109/CCWC57344.2023.10099106.

[22] Khan R. H., Miah J., Abed N S. A. and Islam M., (2023). A Comparative Study of Machine Learning classifiers to analyze the Precision of Myocardial Infarction prediction, 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 2023, 0949-0954, doi: 10.1109/CCWC57344.2023.10099059.

[23] Krishnaiah, V., Narsimha, G., & Chandra, N. S. (2013). Diagnosis of lung cancer prediction system using data mining classification techniques. *International Journal of Computer Science and Information Technologies*, *4*(1), 39-45.

[24] Khan R. H. and Miah J., (2022) Performance Evaluation of a new one-time password (OTP) scheme using stochastic petri net (SPN), 2022 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2022, 407-412, doi: 10.1109/AIIoT54504.2022.9817203.

[25] Khan, R. H., Miah, J., Arafat, S. M., Syeed, M. M., & Ca, D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks. *arXiv preprint arXiv:2310.17729*.

[26] Khan R. H., Miah J., Rahat M. A. R., Ahmed A. H., Shahriyar M. A. and Lipu E. R., (2023). A Comparative Analysis of Machine Learning Approaches for Chronic Kidney Disease Detection," 2023 8th International Conference on Electrical, Electronics and Information Engineering (ICEEIE), Malang City, Indonesia, 2023, 1-6, doi: 10.1109/ICEEIE59078.2023.10334765.

[27] Kabir H. M. D., Anwar S., Alam S. B., Sabidur K. M. R, Abdul M M. and Kabir C I., (2010). Watermarking with fast and highly secured encryption for real-time speech signals, 2010 IEEE International Conference on Information Theory and Information Security, Beijing, China, 2010, 446-451, doi: 10.1109/ICITIS.2010.5689567.

[28] Khan, R. H., Miah, J., Arafat, S. M. Y., Syeed, M. M. M., & Ca, D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks. arXiv preprint, arXiv:2310.17729.

[29] Khan, R. H., Miah, J., Arafat, S. M., Syeed, M. M., & Ca, D. M. (2023). Improving Traffic Density Forecasting in Intelligent Transportation Systems Using Gated Graph Neural Networks. arXiv preprint arXiv:2310.17729.

[30] Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. M. Y. (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. arXiv preprint, 2311.00517.

[31] Miah, J., Cao, D. M., Sayed, M. A., & Haque, M. S. (2023). Generative AI Model for Artistic Style Transfer Using Convolutional Neural Networks. arXiv preprint arXiv:2310.18237

[32] Miah J. Ca D. M, Sayed M. A., Lipu E. R., Mahmud F. and Arafat S. M. Y., (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction, 2023 15th International Conference on Innovations in Information Technology (IIT), Al Ain, United Arab Emirates, 2023, pp. 49-54, doi: 10.1109/IIT59782.2023.10366476.

[33] Miah, J., Ca, D. M., Sayed, M. A., Lipu, E. R., Mahmud, F., & Arafat, S. M. Y. (2023). Improving Cardiovascular Disease Prediction Through Comparative Analysis of Machine Learning Models: A Case Study on Myocardial Infarction. arXiv preprint, 2311.00517.

[34] Miah J., Khan R. H., Ahmed S., and Mahmud M. I. (2023). A comparative study of Detecting Covid 19 by Using Chest X-ray Images– A Deep Learning Approach, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, 0311-0316, doi: 10.1109/AIIoT58121.2023.10174382.

[35] Miah, J., Cao, D. M., Sayed, M. A., Taluckder, M. S., Haque, M. S., & Mahmud, F. (2023). Advancing Brain Tumor Detection: A Thorough Investigation of CNNs, Clustering, and SoftMax Classification in the Analysis of MRI Images. *arXiv preprint arXiv:2310.17720*.

[36] Miah, J., & Khan, R. H. (2019, November). Service Development of Smart Home Automation System: A Formal Method Approach. In *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems* (pp. 161-167).

[37] Patra, R. (2020). Prediction of Lung Cancer Using Machine Learning Classifier. In: Chaubey, N., Parikh, S., Amin, K. (eds) Computing Science, Communication and Security. COMS2 2020. Communications in Computer and Information Science, vol 1235. Springer, Singapore.

[38] Rahman M. M., Islam A. M., Miah J., Ahmad S., and Hasan M. M., (2023). Empirical Analysis with Component Decomposition Methods for Cervical Cancer Risk Assessment, 2023 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2023, 0513-0519, Doi: 10.1109/AIIoT58121.2023.10174477.

[39] Rana, M. S., Hossain, M. M., Jewel, R. M., & Islam, M. R. (2017). Evaluating Customers Satisfaction of Electronic Banking: An Empirical Study in Bangladesh. The SIJ Transactions on Industrial, Financial & Business Management, 5(03), 07-12.

[40] Shoukat, S., & Rabby, S. (2017). A Total Journey of Footwear With Material Analysis. *Journal of Scientificand Engineering Research, 4*(6), 187–191

[41] Syeed, M. M., Khan, R. H., & Miah, J. (2021). Agile Fitness of Software Companies in Bangladesh: An Empirical Investigation. *International Journal of Advanced Computer Science and Applications*, *12*(2).

[42] Shoukat, S. (2016). Now and then: The neckline history of women. American Scientific *Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 26(2), 33-52.

[43] Shoukat, S., & Islam, M. A. (2016). Development of Wedding Costume from Bengal to Bangladesh.

[44] Tanvir I, Eftekhar H A, Bishnu P G, Salim C, Rumana S, Aisharyja R P, Sanjida R, Taslima A, Mamunur R, Mohammad S B, & Tuan N N (2024). Revolutionizing Retail: A Hybrid Machine Learning Approach for Precision Demand Forecasting and Strategic Decision-Making in Global Commerce. *Journal of Computer Science and Technology Studies*, *6*(1), 33–39. https://doi.org/10.32996/jcsts.2024.6.1.4

[45] Uddin, B., Al Mamun, A., Haque, A., & Jewel, R. M. (2017). FACTORS INFLUENCING SELECTION OF HIGHER LEARNING INSTITUTES: AN EMPIRICAL INVESTIGATION IN BANGLADESH. *Aktual'ni Problemy Ekonomiky= Actual Problems in Economics*, (196), 27-101.