Global Journal of Engineering and Technology Advances
eISSN: 2582-5003
Cross Ref DOI: 10.30574/gjeta
Journal homepage: https://gjeta.com/

(REVIEW ARTICLE)

Check for updates

How ML transforms drug discovery

Kallol Kanti Mondal [1,*], Daniel Lucky Michael [2] and Pabitra Mandal [3]

[1] *Institute of Biological Sciences, University of Rajshahi, Rajshahi-6205, Bangladesh.*
[2] *San Francisco Bay University, Fremont, California, USA.*
[3] *Medical Assistant Training School, Bagerhat, Bangladesh.*
[3] *Bandhan Private Hospital, Faridpur, Bangladesh.*

Global Journal of Engineering and Technology Advances, 2024, 21(01), 197-203

Publication history: Received on 02 August 2024; revised on 22 October 2024; accepted on 25 October 2024

Article DOI: https://doi.org/10.30574/gjeta.2024.21.1.0163

Abstract

The process of drug discovery focuses on identifying novel compounds with specific chemical properties to treat various diseases. In recent years, the field has increasingly incorporated computational techniques, driven by the rapid rise of machine learning technologies and their widespread accessibility. With the goals outlined by the Precision Medicine initiative and the emerging challenges in drug discovery, there is a growing need for robust, standardized, and reproducible computational methodologies to meet these objectives. Machine learning-based predictive models have become particularly significant in the early stages of drug development, prior to preclinical studies, offering the potential to significantly reduce both costs and research timelines. This review article explores the application of these advanced methodologies in recent research, providing insights into the current state of the field. By examining recent advancements, it aims to shed light on the future direction of cheminformatics, its limitations, and the positive outcomes achieved so far.

Keywords: Machine Learning; Drug Discovery; Cheminformatics; QSAR; Molecular Descriptors; COVID-19

1. Introduction

The Precision Medicine Initiative defines precision medicine as "an emerging approach for disease treatment and prevention that considers individual variability in genes, environment, and lifestyle" [1]. This innovative approach enables healthcare professionals and researchers to enhance the precision of predicting effective treatment and prevention strategies for specific groups of people. Unlike the traditional "one-size-fits-all" method, which develops strategies based on the average person without accounting for individual differences, precision medicine tailors its strategies to everyone's unique characteristics.

While precision medicine opens up exciting possibilities for developing new treatments, it also presents significant challenges in creating new methodologies. In recent years, an enormous amount of biomedical data has been generated from diverse sources, ranging from small research labs to large-scale international initiatives. These datasets, often referred to as omics data (genomics, proteomics, metabolomics, pharmacogenomics, etc.), provide a vast resource for the scientific community to stratify patients, achieve precise diagnoses, and develop novel treatments [2].

Diagnostic tests have become increasingly common in certain disease areas, as they enable rapid identification of the most effective treatment for individual patients through specific molecular analyses. This targeted approach reduces the reliance on trial-and-error medicine, which can be both costly and frustrating. Furthermore, drugs designed using molecular insights often enhance treatment outcomes while minimizing side effects. For example, in breast cancer treatment, a significant subset of patients with tumors characterized by overexpression of the human epidermal growth

* Corresponding author: Kallol Kanti Mondal

factor receptor 2 (HER2) benefit from treatment with trastuzumab (Herceptin) combined with chemotherapy, reducing the risk of recurrence by over 50% [3].

Machine Learning (ML) has gained significant traction in the pharmaceutical industry, offering tools to accelerate and automate the analysis of the vast quantities of biomedical data now available. ML, a branch of Artificial Intelligence (AI), focuses on developing algorithms capable of learning from raw, unprocessed data to perform specific tasks. These tasks include classification, regression, clustering, and pattern recognition within extensive datasets. Various ML methods, such as Naive Bayes, Support Vector Machines, Random Forest, and, more recently, Deep Neural Networks, have been applied in the pharmaceutical sector for predicting molecular properties, biological activities, drug interactions, and potential adverse effects [4–14].

2. Machine Learning in Drug Design

2.1. Prediction of Target Protein Structures

Proteins are essential to numerous biological processes, and their malfunctions can cause abnormal cellular behavior, leading to various diseases [15]. To selectively target these diseases, small-molecule compounds are often designed based on the three-dimensional (3D) chemical environment around the ligand-binding sites of target proteins [16]. Thus, predicting the 3D structure of target proteins is a critical aspect of structure-based drug discovery. Traditionally, homology modeling has been employed for this purpose, using known protein structures as templates [17]. However, machine learning (ML) approaches have shown considerable potential in predicting protein structures more accurately and efficiently. AlphaFold, developed by DeepMind, leverages deep neural networks (DNNs) to predict 3D protein structures by analyzing amino acid distances and peptide bond angles. It has achieved remarkable accuracy in several protein structure prediction competitions and has significantly advanced the field, revolutionizing structure-based drug discovery [14]. Nevertheless, the complexity of protein structures—such as their ability to form multiple coexisting configurations—poses challenges for accurate prediction [18].

2.2. Prediction of Protein-Protein Interactions (PPIs)

Proteins typically function in collaboration with other proteins, forming intricate networks known as protein-protein interactions (PPIs). PPIs are vital for regulating protein activity, altering specificity, and creating novel binding sites for effector molecules [19]. Understanding and targeting PPIs can enable the design of innovative drugs that modulate complex biological processes.

ML-based methods for predicting PPIs fall into two main categories: structure-based and sequence-based approaches. Structure-based methods utilize knowledge of protein structure similarities, as seen with tools like IntPred, a random forest-based predictor of protein-protein interface sites, and Struct2Graph, a graph attention network (GAT)-based classifier [20,21]. Sequence-based approaches, on the other hand, use protein sequence data to predict physical interactions. For example, DeepPPI employs deep neural networks to learn protein representations from descriptors, achieving high performance metrics on datasets such as S. cerevisiae [22]. Moreover, new sequence-based models, like DELPHI, use deep ensemble approaches to predict PPI-binding sites [23].

2.3. Prediction of Drug-Target Interactions (DTIs)

Drugs exert therapeutic effects by interacting with specific target molecules, such as enzymes, receptors, and ion channels. Predicting drug-target interactions (DTIs) is a crucial step in drug design. Traditional experimental methods for this task are time-intensive and costly, prompting the adoption of ML techniques to predict DTIs more efficiently.

ML-based approaches focus on three key aspects: identifying binding sites, estimating binding affinity, and determining the binding pose of a drug within its target. For example, DeepC-SeqSite uses convolutional neural networks (CNNs) to predict binding residues, outperforming traditional methods like COACH [24]. AGAT-PPIS leverages augmented GAT to improve binding site prediction accuracy [25]. Tools like DEELIG and GraphDelta have been employed to assess binding affinities, while CNN and random forest-based models help identify optimal ligand-target binding configurations [26,27,28]. These methods demonstrate the potential of ML in improving the accuracy and efficiency of DTI prediction.

2.4. De Novo Drug Design

De novo drug design involves creating new molecules from scratch without relying on existing compounds or drug structures. This approach aims to design molecules with specific properties to target particular diseases or conditions

[29,30]. Traditional de novo methods, such as fragment-based approaches, often face challenges with poor drug metabolism, pharmacokinetics, and synthesis complexity [31,32].

ML has introduced transformative methods for de novo drug design, particularly through the use of generative models like variational autoencoders (VAEs) and adversarial autoencoders (AAEs). PaccMannRL integrates a hybrid VAE with reinforcement learning to design anti-cancer molecules using transcriptomic data [33]. DruGAN, another ML-based approach, employs AAEs to generate novel molecules with anti-cancer properties [34]. MedGAN, a model combining Wasserstein GAN and GCN, successfully generates quinoline-scaffold molecules with drug-like properties, demonstrating its ability to produce novel, unique, and effective compounds. To address synthesis challenges, tools like SCScore evaluate the synthetic complexity of generated molecules using neural networks trained on reaction precedents [35]. These advancements highlight the pivotal role of ML in revolutionizing de novo drug design.

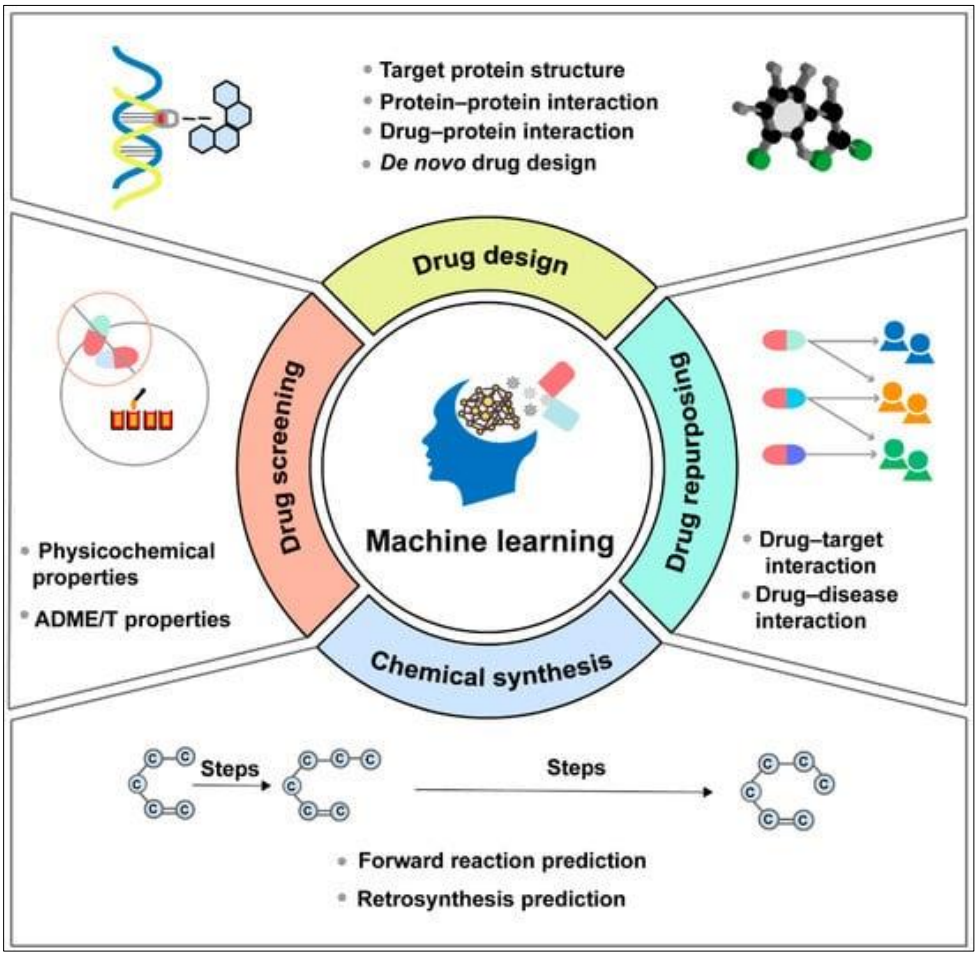

Figure 1 Machine learning can be applied in multiple stages of the drug discovery process [36], mainly including drug design, drug screening, drug repurposing and chemical synthesis [36]

3. Machine Learning in Drug Screening

3.1. Predicting Physicochemical Properties

The physicochemical properties of drugs, such as solubility, ionization, partition coefficient, permeability, and stability, play a crucial role in determining their behavior within biological systems and the environment. These properties influence a drug's bioavailability, absorption, transport, and permeability, as well as its potential risks to human health [37]. During drug screening, evaluating these properties is essential to identify promising candidates for further development.

Francoeur et al. [38] introduced SolTranNet, a molecule attention transformer designed to predict aqueous solubility using the SMILES representation of drug molecules. This tool has demonstrated high sensitivity (0.948) in filtering

insoluble compounds in the SC2 dataset, performing competitively against other methods [38]. Zang et al. [37] used molecular fingerprints and four ML algorithms to create a quantitative structure–property relationship workflow capable of predicting six physicochemical properties of environmental chemicals, including water solubility, melting and boiling points, octanol–water partition coefficient, bioconcentration factor, and vapor pressure [37].

3.2. Predicting ADME/T Properties

After identifying hit or lead compounds in the drug discovery process, it is essential to evaluate their ADME/T properties—absorption, distribution, metabolism, excretion, and toxicity. These pharmacokinetic characteristics are critical for understanding how a compound behaves in the body and its potential to be safe and effective as a drug. Poor ADME/T profiles are a common reason for drug candidates failing in late-stage development or even for the withdrawal of approved drugs [39,40].

Tian et al. [41] created ADMETboost, a web server that employs the XGBoost algorithm to analyze molecular features from fingerprints and descriptors. It predicts key ADME/T properties like blood-brain barrier permeability, CYP2C9 inhibition, and hERG inhibition, achieving top performance in the Therapeutics Data Commons ADMET benchmark for 18 out of 22 tasks [41].

Li et al. [42] designed a multitask autoencoder deep neural network (DNN) model using data from over 13,000 compounds in the PubChem BioAssay Database to predict inhibitors for five major cytochrome P450 (CYP450) isoforms. This multitask model achieved an average accuracy of 86.4% in 10-fold cross-validation and 88.7% on external test datasets.

ML models analyze existing drugs for potential repurposing by screening molecular databases and predicting interactions with SARS-CoV-2 proteins [44-46, 52, 53]. Example: Deep learning models like DeepDrug target protein-ligand binding and accelerate repurposing. ML identifies biomarkers for early detection and progression of chronic kidney disease (CKD). These biomarkers guide the development of targeted therapies [43, 48-50]. ML identifies druggable targets by analyzing lipid metabolism and pathways involved in NAFLD. Integrative ML approaches combine omics data to design therapies for nonalcoholic steatohepatitis (NASH) [54].

4. Conclusion

Advancements in AI have significantly impacted various fields, including cheminformatics and drug discovery, benefiting the pharmaceutical industry. While traditional approaches relied on molecular descriptors derived from small molecules or peptides, modern graph-based models now directly represent molecules, achieving superior results in some cases.

In precision medicine, ML has been instrumental in predicting strong molecular interactions, aiding the discovery of new therapeutic targets. However, translating academic advancements into clinical applications requires standardized frameworks and methodologies to ensure replicability and applicability. Without such standardization, extending results to real-world tasks remains challenging. Throughout this review, this problem has been detected in the different articles reviewed. Therefore, to draw definitive conclusions, this aspect must be deeply influenced. However, the possibilities and advantages offered by ML techniques are immense within the context of precision medicine and drug discovery.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Collins FS, Varmus H. A new initiative on precision medicine. New England J Med 2015;372(9):793–5.

[2] Curtis C, Shah SP, Chin S-F, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486(7403):346–52.

[3] Romond EH, Perez EA, Bryant J, Suman VJ, Geyer Jr CE, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA, et al. Trastuzumab plus adjuvant chemotherapy for operable her2-positive breast cancer. N Engl J Med 2005;353(16):1673–84.

[4] Blanco JL, Porto-Pazos AB, Pazos A, Fernandez-Lozano C. Prediction of high anti-angiogenic activity peptides in silico using a generalized linear model and feature selection. Sci Rep 2018;8(1):1–11.

[5] Munteanu CR, Fernández-Blanco E, Seoane JA, Izquierdo-Novo P, Angel Rodriguez-Fernandez J, Maria Prieto-Gonzalez J, Rabunal JR, Pazos A. Drug discovery and design for complex diseases through qsar computational methods. Current Pharmaceutical Des 2010;16(24):2640–55.

[6] García I, Munteanu CR, Fall Y, Gómez G, Uriarte E, González-Díaz H. Qsar and complex network study of the chiral hmgr inhibitor structural diversity. Bioorganic Med Chem 2009;17(1):165–75.

[7] Liu Y, Tang S, Fernandez-Lozano C, Munteanu CR, Pazos A, Yu Y-Z, Tan Z, González-Díaz H. Experimental study and random forest prediction model of microbiome cell surface hydrophobicity. Expert Syst Appl 2017;72:306–16.

[8] Riera-Fernández P, Munteanu CR, Dorado J, Martin-Romalde R, DuardoSanchez A, Gonzalez-Diaz H. From chemical graphs in computer-aided drug design to general markov-galvez indices of drug-target, proteome, drugparasitic disease, technological, and social-legal networks. Current Computeraided Drug Des 2011;7(4):315–37.

[9] Shirvani P, Fassihi A. Molecular modelling study on pyrrolo [2, 3-b] pyridine derivatives as c-met kinase inhibitors, a combined approach using molecular docking, 3d-qsar modelling and molecular dynamics simulation. Mol Simul 2020:1–16.

[10] B. Suay-Garcia, J.I. Bueso-Bordils, A. Falcó, M.T. Pérez-Gracia, G. Antón-Fos, P. Alemán-López, Quantitative structure–activity relationship methods in the discovery and development of antibacterials, Wiley Interdisciplinary Reviews: Computational Molecular Science e1472..

[11] Fernandez-Lozano C, Gestal M, Munteanu CR, Dorado J, Pazos A. A methodology for the design of experiments in computational intelligence with multiple regression models. PeerJ 2016;4:e2721.

[12] D.S. Wishart, Y.D. Feunang, A.C. Guo, E.J. Lo, A. Marcu, J.R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, et al., Drugbank 5.0: a major update to the drugbank database for 2018, Nucleic acids research 46 (D1) (2018) D1074–D1082.

[13] Kim S, Chen J, Cheng T, Gindulyte A, He J, He S, Li Q, Shoemaker BA, Thiessen PA, Yu B, et al. Pubchem 2019 update: improved access to chemical data. Nucleic Acids Res 2019;47(D1):D1102–9.

[14] Carracedo-Reboredo, Paula, et al. "A review on machine learning approaches and trends in drug discovery." Computational and structural biotechnology journal 19 (2021): 4538-4558.

[15] You, Z.H.; Li, S.; Gao, X.; Luo, X.; Ji, Z. Large-scale protein-protein interactions detection by integrating big biosensing data with computational model. BioMed Res. Int. 2014, 2014, 598129.

[16] Chan, H.C.S.; Shan, H.; Dahoun, T.; Vogel, H.; Yuan, S. Advancing Drug Discovery via Artificial Intelligence. Trends Pharmacol. Sci. 2019, 40, 592–604.

[17] Muhammed, M.T.; Aki-Yalcin, E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. Chem. Biol. Drug Des. 2019, 93, 12–20.

[18] Zhang, Y.; Skolnick, J. The protein structure prediction problem could be solved using the current PDB library. Proc. Natl. Acad. Sci. USA 2005, 102, 1029–1034.

[19] Tang, T.; Zhang, X.; Liu, Y.; Peng, H.; Zheng, B.; Yin, Y.; Zeng, X. Machine learning on protein-protein interaction prediction: Models, challenges and trends. Brief. Bioinform. 2023, 24, bbad076.

[20] Northey, T.C.; Barešić, A.; Martin, A.C.R. IntPred: A structure-based predictor of protein-protein interaction sites. Bioinformatics 2018, 34, 223–229.

[21] Baranwal, M.; Magner, A.; Saldinger, J.; Turali-Emre, E.S.; Elvati, P.; Kozarekar, S.; VanEpps, J.S.; Kotov, N.A.; Violi, A.; Hero, A.O. Struct2Graph: A graph attention network for structure based predictions of protein-protein interactions. BMC Bioinform. 2022, 23, 370.

[22] Du, X.; Sun, S.; Hu, C.; Yao, Y.; Yan, Y.; Zhang, Y. DeepPPI: Boosting Prediction of Protein-Protein Interactions with Deep Neural Networks. J. Chem. Inf. Model. 2017, 57, 1499–1510.

[23] Li, Y.; Golding, G.B.; Ilie, L. DELPHI: Accurate deep ensemble model for protein interaction sites prediction. Bioinformatics 2021, 37, 896–904.

[24] Cui, Y.; Dong, Q.; Hong, D.; Wang, X. Predicting protein-ligand binding residues with deep convolutional neural networks. BMC Bioinform. 2019, 20, 93.

[25] Zhou, Y.; Jiang, Y.; Yang, Y. AGAT-PPIS: A novel protein-protein interaction site predictor based on augmented graph attention network with initial residual and identity mapping. Brief. Bioinform. 2023, 24, bbad122.

[26] Ahmed, A.; Mam, B.; Sowdhamini, R. DEELIG: A Deep Learning Approach to Predict Protein-Ligand Binding Affinity. Bioinform. Biol. Insights 2021, 15, 11779322211030364.

[27] Karlov, D.S.; Sosnin, S.; Fedorov, M.V.; Popov, P. graphDelta: MPNN Scoring Function for the Affinity Prediction of Protein-Ligand Complexes. ACS Omega 2020, 5, 5150–5159.

[28] Nguyen, D.D.; Cang, Z.; Wu, K.; Wang, M.; Cao, Y.; Wei, G.W. Mathematical deep learning for pose and binding affinity prediction and ranking in D3R Grand Challenges. J. Comput. Aided Mol. Des. 2019, 33, 71–82

[29] Nicolaou, C.A.; Kannas, C.; Loizidou, E. Multi-objective optimization methods in de novo drug design. Mini Rev. Med. Chem. 2012, 12, 979–987.

[30] Zhong, F.; Xing, J.; Li, X.; Liu, X.; Fu, Z.; Xiong, Z.; Lu, D.; Wu, X.; Zhao, J.; Tan, X.; et al. Artificial intelligence in drug design. Sci. China. Life Sci. 2018, 61, 1191–1204.

[31] Hessler, G.; Baringhaus, K.H. Artificial Intelligence in Drug Design. Molecules 2018, 23, 2520.

[32] Schneider, G.; Funatsu, K.; Okuno, Y.; Winkler, D. De novo Drug Design—Ye olde Scoring Problem Revisited. Mol. Inform. 2017, 36, 1681031.

[33] Born, J.; Manica, M.; Oskooei, A.; Cadow, J.; Markert, G.; Rodríguez Martínez, M. PaccMann(RL): De novo generation of hit-like anticancer molecules from transcriptomic data via reinforcement learning. iScience 2021, 24, 102269.

[34] Kadurin, A.; Nikolenko, S.; Khrabrov, K.; Aliper, A.; Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. Mol. Pharm. 2017, 14, 3098–3104.

[35] Coley, C.W.; Rogers, L.; Green, W.H.; Jensen, K.F. SCScore: Synthetic Complexity Learned from a Reaction Corpus. J. Chem. Inf. Model. 2018, 58, 252–261.

[36] Qi, X.; Zhao, Y.; Qi, Z.; Hou, S.; Chen, J. Machine Learning Empowering Drug Discovery: Applications, Opportunities and Challenges. Molecules 2024, 29, 903. https://doi.org/10.3390/molecules29040903

[37] Zang, Q.; Mansouri, K.; Williams, A.J.; Judson, R.S.; Allen, D.G.; Casey, W.M.; Kleinstreuer, N.C. In Silico Prediction of Physicochemical Properties of Environmental Chemicals Using Molecular Fingerprints and Machine Learning. J. Chem. Inf. Model. 2017, 57, 36–49.

[38] Francoeur, P.G.; Koes, D.R. SolTranNet-A Machine Learning Tool for Fast Aqueous Solubility Prediction. J. Chem. Inf. Model. 2021, 61, 2530–2536.

[39] Wang, L.; Ding, J.; Pan, L.; Cao, D.; Jiang, H.; Ding, X. Artificial intelligence facilitates drug design in the big data era. Chemom. Intell. Lab. Syst. 2019, 194, 103850.

[40] Yang, X.; Wang, Y.; Byrne, R.; Schneider, G.; Yang, S. Concepts of Artificial Intelligence for Computer-Assisted Drug Discovery. Chem. Rev. 2019, 119, 10520–10594.

[41] Tian, H.; Ketkar, R.; Tao, P. ADMETboost: A web server for accurate ADMET prediction. J. Mol. Model. 2022, 28, 408

[42] Li, X.; Xu, Y.; Lai, L.; Pei, J. Prediction of Human Cytochrome P450 Inhibition Using a Multitask Deep Autoencoder Neural Network. Mol. Pharm. 2018, 15, 4336–4345.

[43] Md Boktiar Hossain, Khandoker Hoque, Mohammad Atikur Rahman, Priya Podder, Deepak Gupta, “Hepatitis C Prediction Applying Different ML Classification Algorithm”, International Conference on Computing and Communication Networks 2024 (ICCCNet 2024).

[44] Maniruzzaman, M., Uddin, M. S., Hossain, M. B., & Hoque, K. (2023). Understanding COVID-19 Through Tweets using Machine Learning: A Visualization of Trends and Conversations. European Journal of Advances in Engineering and Technology, 10(5), 108-114.

[45] Khandoker Hoque, Md Boktiar Hossain, Denesh Das, Partha Protim Roy. Integration of IoT in Energy Sector. International Journal of Computer Applications. 186, 36 (Aug 2024), 32-40. DOI=10.5120/ijca2024923981

[46] Maniruzzaman, Md, Anhar Sami, Rahmanul Hoque, and Pabitra Mandal. "Pneumonia prediction using deep learning in chest X-ray Images." International Journal of Science and Research Archive 12, no. 1 (2024): 767-773.

[47] Kowsher, Md, Anik Tahabilder, Md Zahidul Islam Sanjid, Nusrat Jahan Prottasha, Md Shihab Uddin, Md Arman Hossain, and Md Abdul Kader Jilani. "LSTM-ANN & BiLSTM-ANN: Hybrid deep learning models for enhanced classification accuracy." Procedia Computer Science 193 (2021): 131-140.

[48] Md Shihab Uddin. Addressing IoT Security Challenges through AI Solutions. International Journal of Computer Applications. 186, 45 (Oct 2024), 50-55. DOI=10.5120/ijca2024924107

[49] Rahaman, Md Zahedur, Shammi Akhter, Md Rakibul Islam, Suriya Begum, Kallol Kanti Mondal, Md Mottakin, Md Shahadat Hossain, Sulata Bayen, and Munny Das. "Assessment of thrombolytic, antioxidant and analgesic properties of a medicinal plant of Asteraceae family growing in Bangladesh." Discovery Phytomedicine 7, no. 1 (2020): 47-52.

[50] Mondal, Kallol Kanti. "Potential investigation of anti-inflammatory activity and phytochemical investigations of ethanolic extract of Glycosmis pentaphylla leaves." American Journal of Biomedical Research 3, no. 1 (2015): 6-8.

[51] Azam, Shofiul, Archi Farhana Huda, Kishower Shams, Prawej Ansari, Md Mahmudul Hasan, and Mustafa Khalid Mohamed. "Anti-inflammatory and anti-oxidant study of ethanolic extract of Mimosa pudica." Journal of Young Pharmacists 7, no. 3 (2015): 234.

[52] Mondal, M. R. H., Bharati, S., Podder, P., & Podder, P. (2020). Data analytics for novel coronavirus disease. informatics in medicine unlocked, 20, 100374.

[53] Bharati, S., Podder, P., Mondal, M. R. H., Podder, P., & Kose, U. (2022). A review on epidemiology, genomic characteristics, spread, and treatments of COVID-19. Data Science for COVID-19, 487-505.

[54] Rahman, F., Das, D., Sami, A., Podder, P., & Michael, D. L. (2024). Liver cirrhosis prediction using logistic regression, naïve bayes and KNN. International Journal of Science and Research Archive, 2024, 12(01), 2411–2420.