Machine Learning-Integrated Computational Drug Discovery for Novel Multi-Target Drugs Against emerging viral outbreaks

Project Reference No.: *48S_BE_0464*

College : Reva University, Bengaluru
Branch : Bioelectronics
Guide(s) : Dr. Raje Siddiraju Upendra
Dr. Karthik Rajendra
Student(s): Mr. M S Upamanyu
Ms. Yashasvi V
Ms. Shalen Janet
Ms. Prerana M

Keywords:
Monkeypox (mpox), Machine Learning (ML), Artificial Intelligence (AI), Computer Vision (CV), Medical Image Processing, Drug Development, Molecular Dynamics (MD).

Introduction:

The recent times have seen various viral outbreaks that have caused mass destruction and huge loss for humanity. Some of the most noticed outbreaks are the COVID-19 and mpox. The most recent development in these contagious viruses is the endemic outbreak of mpox observed in certain regions of Central and West Africa [1]. Figure 1 displays the recent spread of mpox throughout the globe as noted by World Health Origination (WHO) in a 2022 report [2]. This indicates the increased threat due to the spread of mpox virus, which can further escalate from endemic to pandemic. Hence, there is a need to design an effective pipeline for accurate disease diagnosis and further develop an effective drug against mpox disease.

Therefore, the present project delves into the prediction, detection and development of effective drug against the emerging mpox virus aided with AIML techniques. The present project intends to predict the mpox disease using computer vision integrated with algorithms like ResNet, EfficientNet, and Inception, etc [3]. using image data as input for training and testing these algorithms. Further, develop a novel pipeline for the development of the drugs (copyright). Further the developed pipeline is implemented

for finding an effective drug against mpox and further validated the lead drug candidate through in-vitro and in-vivo.

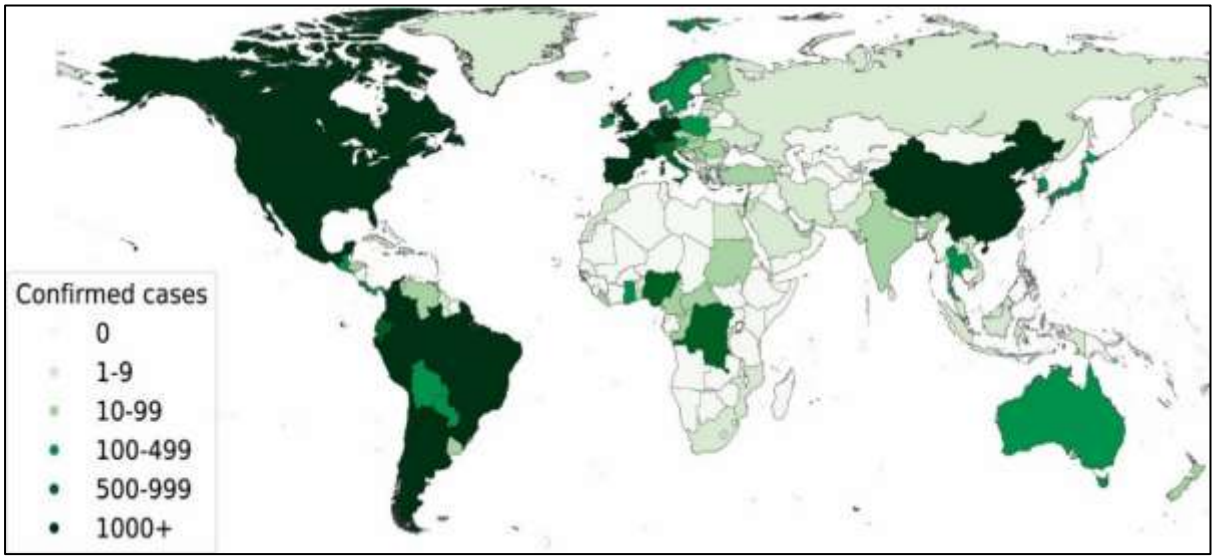

Figure 1: Indicates the spread of mpox throughout the globe in 2022 [2].

Objectives:

To create an intuitive GUI for the drug discovery pipeline, with target protein and ligand molecules as user inputs for streamlined analysis.

a. **Target Identification:** Utilize ML algorithms to analyse viral genome sequences and host-pathogen interactions to identify potential multi-target sites for drug intervention.

b. **Compound Screening:** Implement ML models for virtual screening of existing compound libraries to identify candidates that can bind to multiple targets associated with the viral lifecycle.

c. **Structure-Activity Relationship (SAR) Modeling:** Use ML to model the relationship between the chemical structure of compounds and their biological activity, helping to optimize lead compounds.

To understand the metabolic behaviours of the drug performing molecular dynamics studies

a. **Predictive Modelling:** Develop predictive models for the efficacy and toxicity of identified compounds using historical data and simulation techniques.

b. **Optimizing Drug Design:** Integrate ML-driven insights into drug design processes, facilitating the creation of novel compounds with enhanced pharmacological profiles.

Methodology

Early Prediction

- The study began by retrieving 770 skin lesion images from a publicly available Kaggle dataset, which were further classified into 04 groups i.e. Normal, monkeypox, measles, and chickenpox in order to ensure structured categorization for subsequent analysis [4].
- To overcome the limitation of a small dataset, the study later employed TensorFlow's Keras ImageDataGenerator (i.e. Data augmentation). By applying transformations such as rotations, shifts, zooms, flips, and shearing, the original dataset of 770 images was expanded to 5,000 images [5].
- Five distinct Computer Vision–Deep Learning (CV-DL) models along with three ensemble models (combinations of EfficientNet with ResNet, VGG-16, and Inception) was trained and tested using the training (80% data) and testing (20% data) datasets [6].
- The models were evaluated on different parameters like accuracy, precision, recall, and F1 score. Further, the best model was validated through the implementation of a 10-fold cross-validation [7].

Computational drug discovery

- Once the predictive model was completed, next step was to develop the drug through bioinformatics tools [8-9].
- ML techniques were used to filter out 114 molecules from a total database of 6000 ligands. Further these 114 molecules underwent molecular docking and 02 molecules among 114 molecules displayed better docking score than the standard drug molecule [10].
- The best molecule out of the 02 lead molecules was evaluated through molecular dynamics.

Results

Early prediction

- The model trained on EfficientNet algorithm outperformed all the other algorithms with an accuracy of 98.41%, precision of 98.44%, recall of 98.41%, and an F1 score of 0.98 (Figure 2 - Figure 4). This increases the accuracy value by 16.74% when compared to other similar studies.

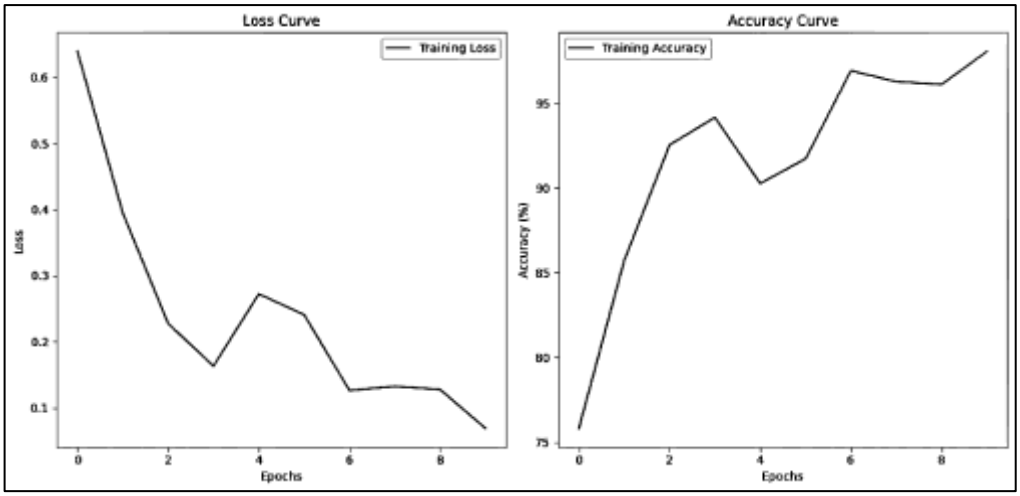

Figure 2 Loss and accuracy curves obtained for the EfficientNet algorithm.

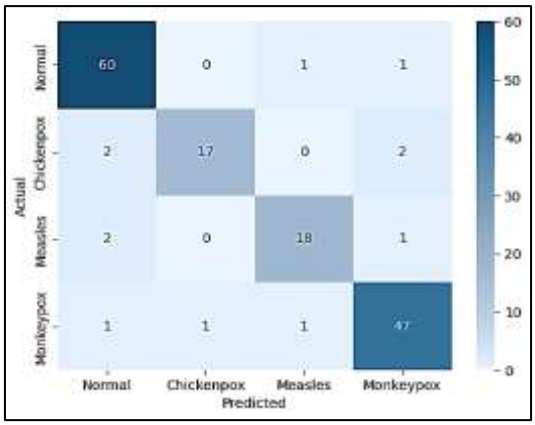

Figure 3 Confusion matrix obtained from the EfficientNet algorithm.

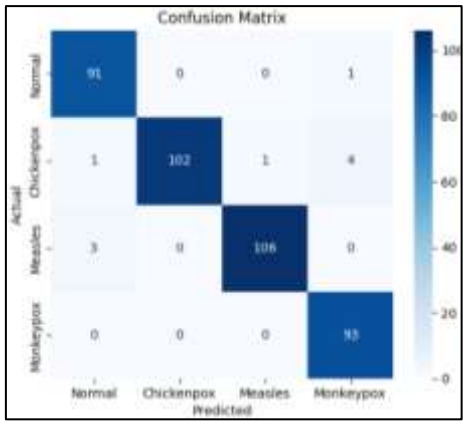

Figure 4. Confusion matrix obtained after the 10-fold cross-validation

- The findings of the present study represented the potential of the EfficientNet model for real-world applications, including integration into the Internet of Medical Things (IoMT), and contribute significantly to public health surveillance and early mpox detection efforts.

Computational Drug Discovery

- Initially 6000 drug molecules were screened using the novel approach as presented in Figure 5. The proposed algorithm utilized the smiles to identify similar compounds. This approach reduced the number of compounds to 114.

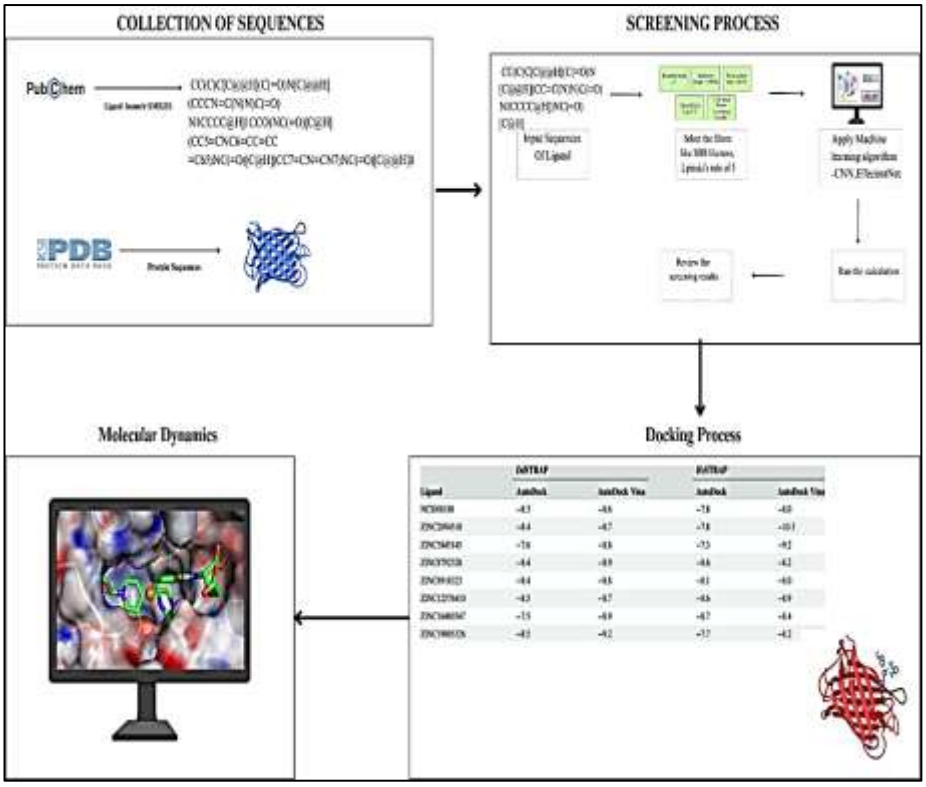

Figure 5. Workflow for novel drug discovery against mpox

- All 114 screened molecules were docked with the identified antigenic protein A29L. The Table 1 shows the docking scores of the different ligands docked with the target protein, A29L.

Table 1. Docking scores of screened molecules with A29L

Lead molecules	Docking Scores
P1	-10.066
P2	-10.307
P3	-10.03
P4	-9.836
P5	-10.395
P6	-10.004
P7	-9.99
P8	-10.172
P12	-10.568

P22	-10.652
P33	-10.57

- The best lead molecule P22 with an PubChem ID 129016360 displayed negative highest docking score of -10.652 kcal/mol (Figure 6). Hence the P22-A29L complex was further subjected to molecular dynamics studies.

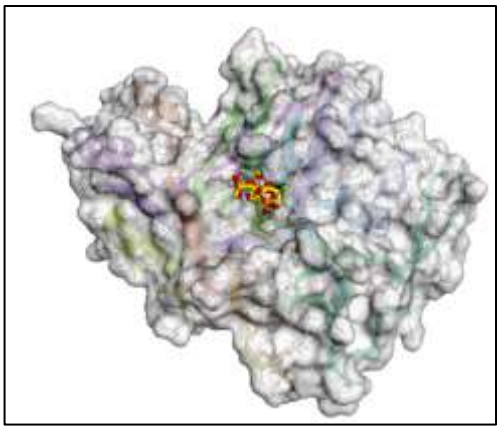

Figure 6. The molecular docking between lead molecule and A29L protein

Outcomes

A Graphical User Interface (GUI) combining both early detection and lead molecule prediction for mpox disease.

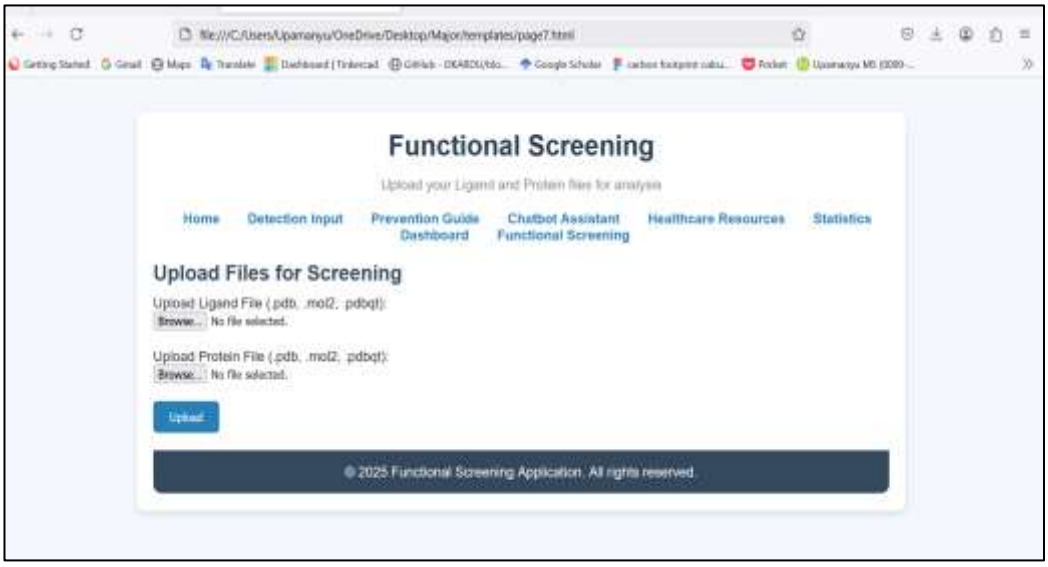

Figure 7. The web application for drug discovery and early prediction of mpox

A copyright for the proposed pipeline

Figure 8. Copyright Submission Proof (11/12/2024)

A patent was filled for the ML model trained and used in early prediction of mpox.

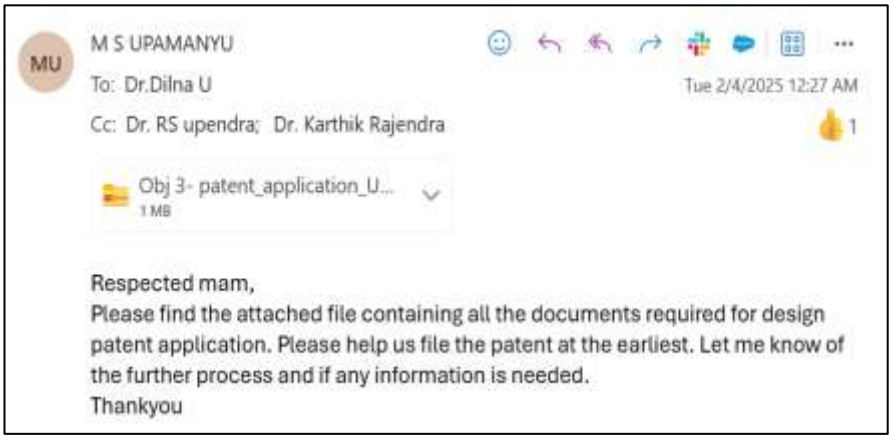

Figure 9. The application form for copyright registration (2/4/2025)

An IEEE conference paper was submitted based on the project.

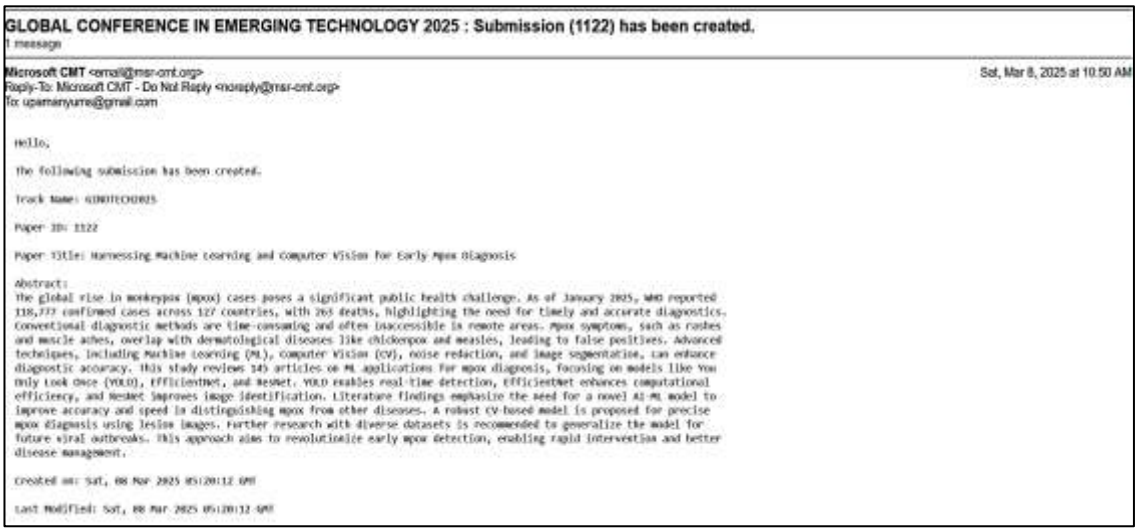

Figure 15. Confirmation mail by CMT for successful submission of conference paper.

Conclusion and Future scope

The utilization of modern techniques like CV enabled with ML algorithms and ensembled models to predict mpox opens up new possibilities and avenues. The model trained on EfficientNet achieved the highest accuracy of 98.41% with a precision of 98.44%, a recall value of 98.41%, and an F1 score of 0.98. This promises for rapid and accurate mpox diagnosis, particularly in resource-limited settings. Furthermore, the novel GUI based drug discovery pipeline enables a faster and accurate drug discovery process. Using the process, the present study obtained 02 lead molecules of which the P22, PubChem ID 129016360 displayed the highest docking score of -10.652 kcal/mol. The obtained lead molecule in the present study must be validated through in-vitro and in-vivo testing.

References

[1] Adegboye, O., Alele, F., Pak, A., Alakunle, E., Emeto, T., Leggat, P., & Okeke, M. (2024). Monkeypox Outbreak 2022, from a Rare Disease to Global Health Emergence: Implications for Travellers. Poxviruses, 355-368.

[2] World Health Organization, & World Health Organization. (2022). Monkeypox outbreak: global trends. World Health Organization, 15.

[3] Maqsood, S., Damaševičius, R., Shahid, S., & Forkert, N. D. (2024). MOX-NET: Multi-stage deep hybrid feature fusion and selection framework for monkeypox classification. Expert Systems with Applications, 255, 124584.

[4] Dipu, I. U. (2022). Monkeypox Skin Image Dataset [Data set]. Kaggle. https://www.kaggle.com/datasets/dipuiucse/monkeypoxskinimagedataset.

[5] Yadav, S., & Qidwai, T. (2024). Machine learning-based monkeypox virus image prognosis with feature selection and advanced statistical loss function. Medicine in Microecology, 19, 100098.

[6] Khan, S. U. R., Asif, S., Bilal, O., & Ali, S. (2024). Deep hybrid model for Mpox disease diagnosis from skin lesion images. International Journal of Imaging Systems and Technology, 34(2), e23044.

[7] Pal, M., Mahal, A., Mohapatra, R. K., Obaidullah, A. J., Sahoo, R. N., Pattnaik, G., ... & Rabaan, A. A. (2023). Deep and transfer learning approaches for automated early detection of monkeypox (Mpox) alongside other similar skin lesions and their classification. ACS omega, 8(35), 31747-31757.

[8] Dharani, A., Ezhilarasi, D. R., Priyadarsini, G., & Abhinand, P. A. (2023). Multi-epitope vaccine candidate design for dengue virus. Bioinformation, 19(5), 628.

[9] Upendra, R. S., Srinivas Nagar, S., & Vasudevan, K. (2024). Computational screening of phytocompound isolated from the plant *Toddalia asiatica* and *Coleonema album* as potential inhibitors against enzyme DPP-4 as a treatment for Type-2 Diabetes mellitus. Chemical Papers, 78(3), 1833-1847.
[10] Munıkannaıan, R., Upendra, R. S., & Nagar, S. S. (2022). Significance of Insilco approaches in finding novel biomolecules challenging newly emerging, resurging, deliberately emerging global outbreaks. The Eurasia Proceedings of Science Technology Engineering and Mathematics, 20, 134-141.