THE SCIENCE WORLD
A Monthly e Magazine
ISSN:2583-2212
Popular Article
December, 2025 Vol.5(12), 11320-11330

MolEpidPred: A digital intelligent tool for large scale and quick molecular epidemiology of Foot-and-mouth Disease

Samarendra Das*, Bighneswar Barik, Rajeev Ranjan, Jitendra Kumar Biswal, Saravanan Subramaniam
ICAR-National Institute on Foot and Mouth Disease, Arugul, Bhubaneswar 752050, India
*Author for correspondence (Email: samarendra.das@icar.org.in)

DOI:10.5281/ScienceWorld.18174721

Summary

Foot-and-mouth Disease (FMD) is one of the economically devastating transboundary animal disease that affects all the cloven-hoofed animals including wildlife. The etiological agent of FMD is a single-stranded positive-sense RNA virus, which is in global circulation with seven serotypes, multiple topotypes, lineages, and sub-lineages within each serotype. Molecular epidemiology of FMD is crucial to implement its control strategies including vaccination and containment. This primarily deals with knowing serotype/topotype/lineage of the virus causing the outbreak using virus capsid sequence data, which requires advanced computational tools. Thus, this article presents a digital intelligent tool (MolEpidPred), primarily developed for FMD molecular epidemiology. The MolEpidPred tool provides independent accuracies of atleast 98%, 90%, and 80% for serotype, topotype, and lineage prediction, respectively. This tool implements multiple machine learning algorithms for prediction of serotype/topotype/lineage of the virus using *k*-mer features generated from viral capsid sequence. This tool is fast, quick, and scalable for wide molecular epidemiology and is suitable for use across all the FMD-endemic countries in the world.

Keywords: FMD; Virus; TAD; AI/ML; Molecular epidemiology; MolEpidPred

Background

Foot-and-mouth Disease (FMD) is one of the significant and serious viral transboundary animal disease (TAD) in the cloven-hoofed animals. The estimated annual economic losses attributed to the FMD in various regions of the world is $6.5–21 billion [1] [2] (United Kingdom: $3.94 billion; India: $3.48 billion; China: $2.5–7 billion; African countries: $0.83–1.12 billon; Brazil: $0.132–0.271 billon, *etc.*). The causative agent, FMD virus, is a single-stranded positive-sense RNA virus in the genus *Aphthovirus*, family *Picornaviridae*. It is genetically diverse, consisting of seven immunologically distinct serotypes, *viz.* O, A, C, Asia1, SAT1, SAT2, and SAT3. Each serotype has numerous topotypes and lineages. Till date more than 65 topotypes and numerous genetic lineages within these serotypes have been reported globally [2]. There is no cross-protection between different serotypes due to presence of unique antigenic variation; thus, infection or vaccination with one serotype does not protect against others [3]. Therefore, this phenomenon complicates vaccine formulation, vaccine candidate selection and disease control [3]. Thus, identifying FMD virus serotype, topotype, lineage, *etc*. (*i.e*., molecular epidemiology) causing outbreak are crucial for framing its control strategy.

The dynamic nature of FMD virus necessitates real-time tracking and disease epidemiology at molecular-level. This has become an indispensable approach for understanding virus evolution, transmission pathways, and outbreak sources by integrating genomic sequence, and outbreak data [4]. With the growing scale of sequencing data and increasing complexity of FMD outbreaks, conventional methods might be insufficient for timely and scalable analysis. Further, traditional serology-based methods are time and resource consuming, risky due to live virus handling, supply chain issues, possibility of virus spill over, requires sophisticated biocontainment facility, *etc*. Therefore, alternate quick and risk-free solutions might be explored for complex molecular epidemiology of the FMD virus.

Artificial Intelligence (AI) and Machine Learning (ML) are being increasingly employed to analyse high-dimensional epidemiological and biological sequence data [2] [5]. These methods excel in detecting nonlinear relationships, classifying viral strains, and predicting virus types/subtypes/lineages causing the outbreaks offering faster, more accurate tools for surveillance and decision-making [6][7]. The convergence of AI/ML with disease epidemiology holds significant potential in predicting the virus type, sub-type, lineage, *etc*. at molecular level is important for understanding virus evolution and control through appropriate vaccine strain selection. This article highlights an AI/ML solution for molecular epidemiology of the FMD virus, which fast, quick, and large-scale platform.

Artificial Intelligence and Machine Learning in animal health: Digital intelligence is revolutionizing animal health and disease research by making it possible to generate faster, wiser, and more precise insights from high-throughput sequencing data. AI which involves a wide range of ML algorithms, makes data-driven prediction possible without depending on expensive and labour-intensive laboratory processes [8]. Their capacity to process and analyse high-throughput, high-dimensional datasets—such as genomic sequences, clinical data, epidemiological information, and environmental data—has greatly improved disease detection, outbreak prediction, and epidemiological surveillance [8], [9].

In infectious disease control, AI and ML assist in the early detection of emergent pathogens, tracking of disease spread, and the delivery of timely data to support public health interventions. These applications are increasingly utilized in diagnostics, vaccine design, risk factor modeling, epidemiological analysis, and drug discovery pipelines, offering rapid, scalable, and adaptive solutions appropriate for dynamic characteristics of biological systems [10]. Considering the rising global danger exerted by TADS, AI/ML-based methods provide critical assistance in explaining transmission dynamics, providing early warning systems, formulating predictive models, and informing evidence-based policy directions for animal health sectors.

Previous studies indicate that AI/ML methods are capable of attaining comparable performance to or even exceeding conventional epidemiological methods in many applications [11]. This could be due to the natural flexibility of ML algorithms and their minimal dependence on strong assumptions, as they tend to work in nonparametric or semi-parametric models [12]. Nevertheless, given these benefits, incorporation of ML into TAD control and molecular epidemiology is still in its infancy, which highlights the necessity for further development and advancement of computational approaches.

Digital Intelligence in FMD molecular epidemiology: Application of AI/ML in study and management of FMD has gained significant traction in recent years, owing to the increasing availability of high-resolution molecular, epidemiological, and environmental data. FMD, being a rapidly spreading transboundary viral disease with multiple serotypes and complex transmission pathways, presents unique challenges that require advanced analytical solutions. AI/ML techniques offer the ability to integrate diverse datasets and extract meaningful patterns that inform real-time surveillance, outbreak prediction, and decision-making [7][1]

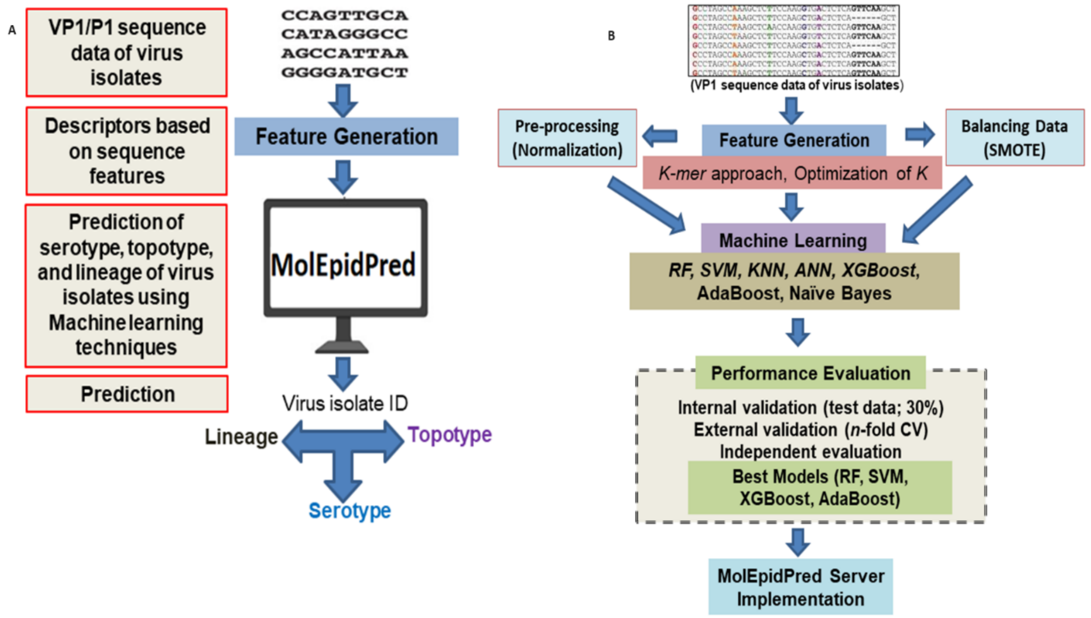
In molecular epidemiology, ML algorithms such as support vector machines, random forests, and deep learning models have been successfully used to classify FMD virus serotypes, predict viral lineage from genomic features, and identify mutations linked to antigenic drift. These approaches enhance the speed and accuracy of diagnosis compared to traditional laboratory methods [5]. Predictive modeling using AI/ML also helps estimate outbreak risk based on spatiotemporal patterns, climate variables, livestock density, and animal movement networks. Such models are invaluable for anticipating hotspots and guiding targeted vaccination and containment strategies.

AI-powered decision-support systems have been integrated into national and regional FMD control programs to optimize resource allocation, simulate intervention scenarios, and evaluate disease mitigation strategies. In some cases, real-time dashboards and early warning systems based on ML have been developed to aid veterinary authorities in rapid response planning. These applications not only reduce the response time during outbreaks but also contribute to long-term control efforts by identifying critical transmission nodes and vulnerable regions. A typical AI application (*e.g.*, MolEpidPred) developed for complex molecular epidemiology of the FMD virus is presented in the following section.

MolEpidPred AI prediction server: Effective control of the FMD relies heavily on molecular epidemiological insights particularly identification of serotype, topotype, and lineage. Traditional methods such serotyping are often biological, involving virus isolation and serotyping procedures that are laborious, time-consuming, and pose biosafety risks due to live virus handling. To address these limitations, MolEpidPred, an AI-based end-to-end computational solution, was developed by ICAR-National Institute on FMD, Bhubaneswar, India, as a quick, rapid, and scalable alternative for conducting molecular epidemiology of FMD using viral capsid (VP1/P1) nucleotide sequence data [2].

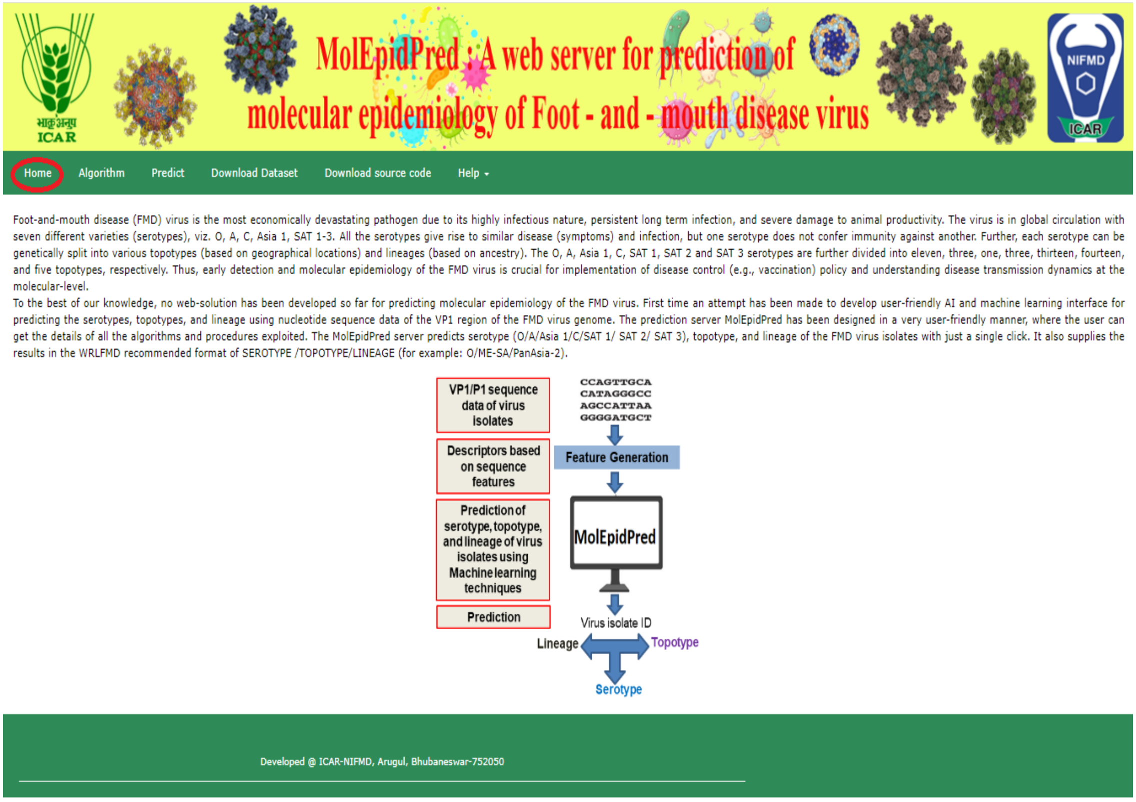
MolEpidPred employs four ensemble learning algorithms trained on curated datasets comprising viral capsid nucleotide sequences of diverse FMD virus strains. The prediction strategy implemented in the MolEpidPred is a three-step process, focusing on predicting serotype, topotype, and lineage of the virus isolates. This predicted information is crucial for understanding specific characteristics of the virus responsible for an outbreak and devising appropriate control strategies, including selection of suitable vaccine candidates. MolEpidPred intends to speed up and improve the effectiveness of decision-making in FMD management and preventive initiatives by offering an intuitive web-based platform for assessing molecular epidemiology.

MolEpidPred is presently available in the form of a web-based AI prediction server available at https://nifmd-bbf.icar.gov.in/MolEpidPred for general use. The tool is user-friendly, which takes VP1/P1 nucleotide sequence data as input and predict various



components of molecular epidemiology including serotype; topotype; and lineage. Schematic presentation of the tool and implemented algorithm is shown in Figure 1. The various web-pages of the MolEpidPred tool are briefly presented in following sections.

***Homepage of MolEpidPred AI prediction server*:** MolEpidPred prediction server's homepage serves as the root landing page that guides users to the function and capability of the AI-based platform. Essentially, this page presents the vision and motive for MolEpidPred: to provide a fast, precise, and extensible computational platform for molecular epidemiology of FMD virus based on VP1/P1 gene sequences. The interface is intuitive and user-oriented, with a straightforward, minimalistic design to ensure simplicity and ease of use. Major options including, *viz.* "Algorithm", "Predict", and "Help" are readily accessible in the navigation bar (Figure 2). So, that the users can access extensive content as and when required. The home page also features brief descriptions of the tool's fundamental functionalities, hyperlinks to the user guide, and details of acceptable input formats.



Foot-and-mouth disease (FMD) virus is the most economically devastating pathogen due to its highly infectious nature, persistent long term infection, and severe damage to animal productivity. The virus is in global circulation with seven different varieties (serotypes), viz. O, A, C, Asia 1, SAT 1-3. All the serotypes give rise to similar disease (symptoms) and infection, but one serotype does not confer immunity against another. Further, each serotype can be genetically split into various topotypes (based on geographical locations) and lineages (based on ancestry). The O, A, Asia 1, C, SAT 1, SAT 2 and SAT 3 serotypes are further divided into eleven, three, one, three, thirteen, fourteen, and five topotypes, respectively. Thus, early detection and molecular epidemiology of the FMD virus is crucial for implementation of disease control (e.g., vaccination) policy and understanding disease transmission dynamics at the molecular-level.
To the best of our knowledge, no web-solution has been developed so far for predicting molecular epidemiology of the FMD virus. First time an attempt has been made to develop user-friendly AI and machine learning interface for predicting the serotypes, topotypes, and lineage using nucleotide sequence data of the VP1 region of the FMD virus genome. The prediction server MolEpidPred has been designed in a very user-friendly manner, where the user can get the details of all the algorithms and procedures exploited. The MolEpidPred server predicts serotype (O/A/Asia 1/C/SAT 1/ SAT 2/ SAT 3), topotype, and lineage of the FMD virus isolates with just a single click. It also supplies the results in the WRLFMD recommended format of SEROTYPE /TOPOTYPE/LINEAGE (for example: O/ME-SA/PanAsia-2).

Figure 2. Homepage of the MolEpidPred web prediction server.

Algorithm page: Scientifically, this page highlights the move away from classical wet-lab-based serotyping towards AI-based molecular diagnostics, highlighting MolEpidPred's power to circumvent time-consuming virus culture and serological testing by means of purely in silico examination. It paves the way for non-bioinformatics users—veterinarians and field epidemiologists, for example—to be able to rely on and successfully communicate with the platform.

This page provides a computationally open description of the ML architecture supporting MolEpidPred. The diagram depicts the three-step hierarchical prediction pipeline used by the server:

(*i*) Serotype Prediction – classification of the antigenic type of the virus.
(*ii*) Topotype Prediction – detection of geographically separate clades in serotypes.
(*iii*) Lineage Prediction – fine-scale evolutionary classification within topotypes.

To execute this multi-level classification task, the platform integrates four powerful ML algorithms: Support Vector Machine (SVM), Random Forest (RFF), Extreme Gradient Boosting (XGB) and Adaptive Boosting (ADB) (Figure 3). The rationale to include these four algorithms was well documented in our previous study [2] [5]. Briefly, these models were trained on a quality-filtered, balanced set of 7,961 global FMD virus isolates, with 3-mer nucleotide features based on the VP1 capsid sequence. The algorithm page details how the models were tested for validity using cross-validation, test data, and independent data sets to ensure generalizability and strength (Figure 3).

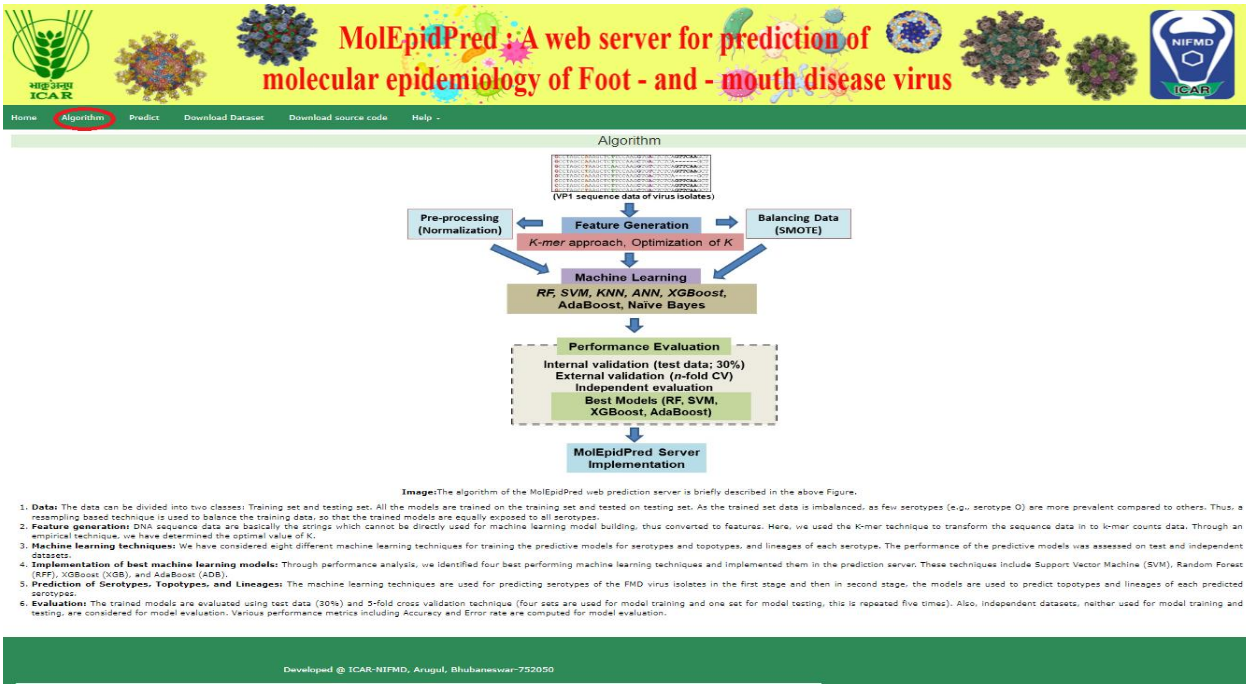


Image: The algorithm of the MolEpidPred web prediction server is briefly described in the above Figure.

1. **Data:** The data can be divided into two classes: Training set and testing set. All the models are trained on the training set and tested on testing set. As the trained set data is imbalanced, as few serotypes (e.g., serotype O) are more prevalent compared to others. Thus, a resampling based technique is used to balance the training data, so that the trained models are equally exposed to all serotypes.
2. **Feature generation:** DNA sequence data are basically the strings which cannot be directly used for machine learning model building, thus converted to features. Here, we used the K-mer technique to transform the sequence data in to k-mer counts data. Through an empirical technique, we have determined the optimal value of K.
3. **Machine learning techniques:** We have considered eight different machine learning techniques for training the predictive models for serotypes and topotypes, and lineages of each serotype. The performance of the predictive models was assessed on test and independent datasets.
4. **Implementation of best machine learning models:** Through performance analysis, we identified four best performing machine learning techniques and implemented them in the prediction server. These techniques include Support Vector Machine (SVM), Random Forest (RFF), XGBoost (XGB), and AdaBoost (ADB).
5. **Prediction of Serotypes, Topotypes, and Lineages:** The machine learning techniques are used for predicting serotypes of the FMD virus isolates in the first stage and then in second stage, the models are used to predict topotypes and lineages of each predicted serotypes.
6. **Evaluation:** The trained models are evaluated using test data (30%) and 5-fold cross validation technique (four sets are used for model training and one set for model testing, this is repeated five times). Also, independent datasets, neither used for model training and testing, are considered for model evaluation. Various performance metrics including Accuracy and Error rate are computed for model evaluation.

Developed @ ICAR-NIFMD, Arugul, Bhubaneswar-752050

Figure 3. Algorithm page of the MolEpidPred web prediction server.

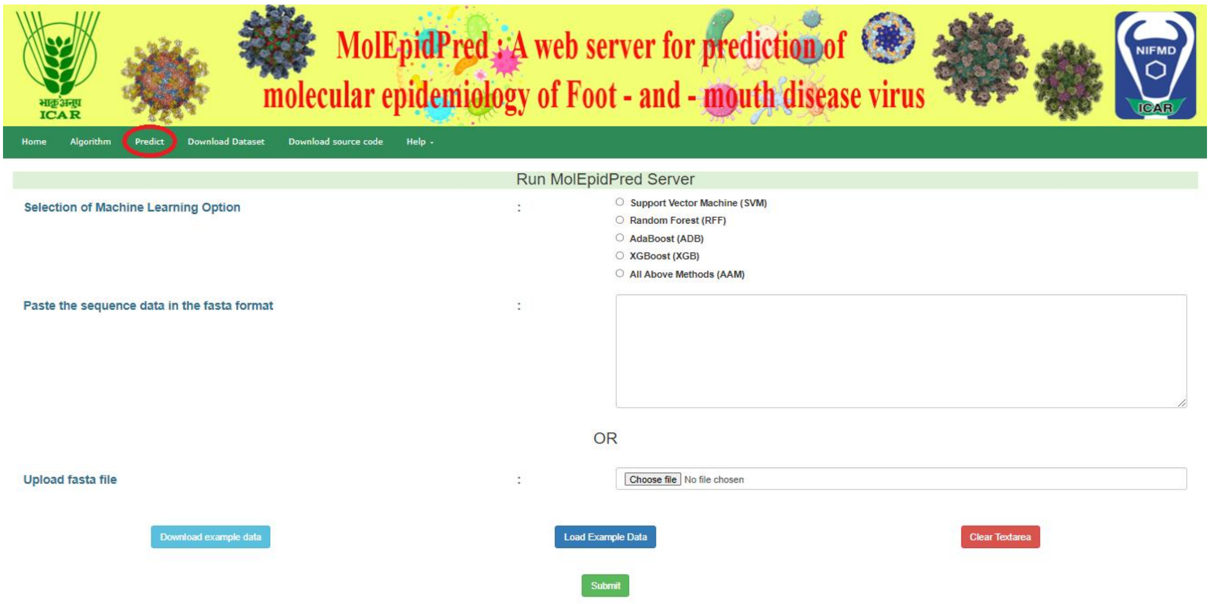
Execution/run page of MolEpidPred: The execution module/run page is the workhorse functionality of the server where users submit data interactively for prediction. The depicted interface enables users to submit the VP1 gene sequences of the FMD virus isolates in FASTA format in following way.

Step 1 *Data input*: Pasting the nucleotide sequences of virus isolates directly in FASTA format in the given window or uploading them as a file (.fasta, .txt). The former option has a limitation of 100 isolates at a time, while the latter gives the user option to use large number of isolates at a time (Figure 4).

Step 2 *Select Algorithm*: This option provides the flexibility to the users about the selection of trained predictive model. The user may choose one or more ML algorithms for prediction or choose to run all models in parallel. The server runs the data through pre-trained models upon submission and makes predictions within seconds (Figure 4).

Step 3 *Execute*: After selecting proper/optional learning algorithm and pasting/uploading input data, the user needs to hit the submit button to get the results. Indeed, this page is an example of democratization of molecular epidemiology to all general users—deprioritizing coding or computational skills. It allows even non-computer-savvy users (*e.g.*, regional veterinary laboratory personnel, policy makers, students, academicians, *etc.*) to conduct high-throughput molecular epidemiologic analysis, thereby expediting outbreak response and vaccine matching in endemic areas.

Step 4 *Output*: After implementing all the above steps (1-3), the results are available at finger tip for downloading. The results can be downloaded in user-editable format for further reporting (Figure 5).



Run MolEpidPred Server

Selection of Machine Learning Option :
○ Support Vector Machine (SVM)
○ Random Forest (RFF)
○ AdaBoost (ADB)
○ XGBoost (XGB)
○ All Above Methods (AAM)

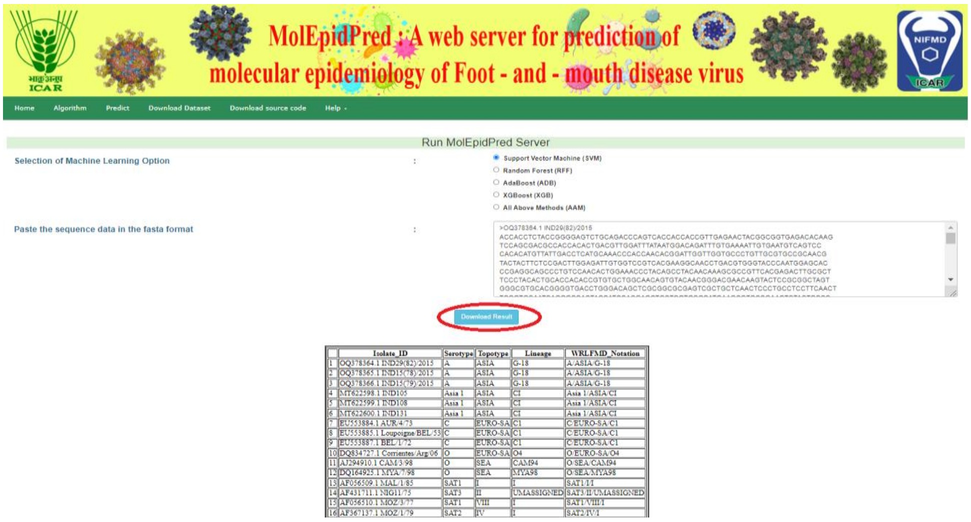
Paste the sequence data in the fasta format :

OR

Upload fasta file : Choose file No file chosen

Figure 4. Execution page of the MolEpidPred web prediction server.

When the model has completed processing the input sequences, the output page displays the prediction outcome in a well-structured tabular format. For each input isolate, the outputs are shown in Figure 5. The interface enables users to download the results in .csv or .txt formats for further analysis, record-keeping, or incorporation into surveillance systems. For multiple model users, the tool also shows model-wise consensus outputs, making more confident interpretation with model agreement or majority voting possible. The output interface supports the conversion of high-dimensional sequence data into epidemiological knowledge that is critical for monitoring viral evolution, regional diversity tracking, and targeted immunization strategy implementation.



Run MolEpidPred Server

Selection of Machine Learning Option :
● Support Vector Machine (SVM)
○ Random Forest (RFF)
○ AdaBoost (ADB)
○ XGBoost (XGB)
○ All Above Methods (AAM)

Paste the sequence data in the fasta format :

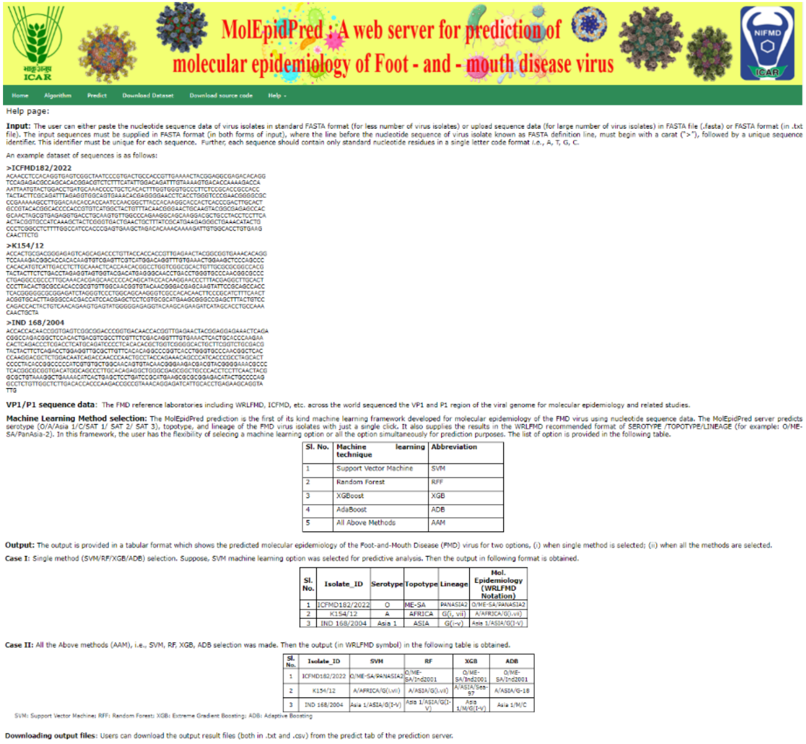
```
>OQ378364.1 IND29(82)/2015
ACCACCTCTACCGGGGAGTCTGCAGACCCAGTCACCACCACCGTTGAGAACTACGGCGGTGAGACACAAG
TCCAGCGACGCCACCACACTGACGTTGGATTTATAATGGACAGATTTGTGAAAATTGTGAATGTCAGTCC
CACACATGTTATTGACCTCATGCAAACCCACCAACACGGATTGGTTGGTGCCCTGTTGCGTGCCGCAACG
TACTACTTCTCCGACTTGGAGATTGTGGTCCGTCACGAAGGCAACCTGACGTGGGTACCCAATGGAGCAC
CCGAGGCAGCCCTGTCCAACACTGGAAACCCTACAGCCTACAACAAAGCGCCGTTCACGAGACTTGCGCT
TCCCTACACTGCACCACACCGTGTGCTGGCAACAGTGTACAACGGGACGAACAAGTACTCCGCGGCTAGT
GGGCGTGCACGGGGTGACCTGGGACAGCTCGCGGCGCGAGTCGCTGCTCAACTCCCTGCCTCCTTCAACT
```

	Isolate_ID	Serotype	Topotype	Lineage	WRLFMD_Notation
1	OQ378364.1 IND29(82)/2015	A	ASIA	G-18	A/ASIA/G-18
2	OQ378365.1 IND15(78)/2015	A	ASIA	G-18	A/ASIA/G-18
3	OQ378366.1 IND15(79)/2015	A	ASIA	G-18	A/ASIA/G-18
4	MT622598.1 IND105	Asia 1	ASIA	CI	Asia 1/ASIA/CI
5	MT622599.1 IND108	Asia 1	ASIA	CI	Asia 1/ASIA/CI
6	MT622600.1 IND131	Asia 1	ASIA	CI	Asia 1/ASIA/CI
7	EU553884.1 AUR/4/73	C	EURO-SA	C1	C/EURO-SA/C1
8	EU553885.1 Loupoigne/BEL/53	C	EURO-SA	C1	C/EURO-SA/C1
9	EU553887.1 BEL/1/72	C	EURO-SA	C1	C/EURO-SA/C1
10	DQ834727.1 Corrientes/Arg/06	O	EURO-SA	O4	O/EURO-SA/O4
11	AJ294910.1 CAM/3/98	O	SEA	CAM94	O/SEA/CAM94
12	DQ164925.1 MYA/7/98	O	SEA	MYA98	O/SEA/MYA98
13	AF056509.1 MAL/1/85	SAT1	I	I	SAT1/I/I
14	AF431711.1 NIG11/75	SAT3	II	UMASSIGNED	SAT3/II/UMASSIGNED
15	AF056510.1 MOZ/3/77	SAT1	VIII	I	SAT1/VIII/I
16	AF367137.1 MOZ/1/79	SAT2	IV	I	SAT2/IV/I

Figure 5. Output/result page of the MolEpidPred web prediction server.

Help page of MolEpidPred AI server: The Help page is structured to provide end-to-end support for users ranging from beginners to experts. It comprises: A step-by-step user guide that features screenshots; Accepted file formats and types; Clarification of error messages and troubleshooting procedures; A FAQs section for technical and methodological questions.

The site also includes contact details of the development team for suggestion, feedback, and queries. Further, the team believes that users are the core of any AI-tool, their suggestion and feedback are paramount for further continuous improvement of the tool. Besides, the web-links to the source code and training datasets of MolEpidPred are also provided in the web-page for further scholarly use. The access to the scholarly publication outlining the detailed methodology [2] is also provided (Figure 6).



Help page:

Input: The user can either paste the nucleotide sequence data of virus isolates in standard FASTA format (for less number of virus isolates) or upload sequence data (for large number of virus isolates) in FASTA file (.fasta) or FASTA format (in .txt file). The input sequences must be supplied in FASTA format (in both forms of input), where the line before the nucleotide sequence of virus isolate known as FASTA definition line, must begin with a carat (">"), followed by a unique sequence identifier. This identifier must be unique for each sequence. Further, each sequence should contain only standard nucleotide residues in a single letter code format *i.e.*, A, T, G, C.

An example dataset of sequences is as follows:

>ICFMD182/2022

>K154/12

>IND 168/2004

VP1/P1 sequence data: The FMD reference laboratories including WRLFMD, ICFMD, etc. across the world sequenced the VP1 and P1 region of the viral genome for molecular epidemiology and related studies.

Machine Learning Method selection: The MolEpidPred prediction is the first of its kind machine learning framework developed for molecular epidemiology of the FMD virus using nucleotide sequence data. The MolEpidPred server predicts serotype (O/A/Asia 1/C/SAT 1/ SAT 2/ SAT 3), topotype, and lineage of the FMD virus isolates with just a single click. It also supplies the results in the WRLFMD recommended format of SEROTYPE /TOPOTYPE/LINEAGE (for example: O/ME-SA/PanAsia-2). In this framework, the user has the flexibility of seleciing a machine learning option or all the option simultaneously for prediction purposes. The list of option is provided in the following table.

Sl. No.	Machine learning technique	Abbreviation
1	Support Vector Machine	SVM
2	Random Forest	RFF
3	XGBoost	XGB
4	AdaBoost	ADB
5	All Above Methods	AAM

Output: The output is provided in a tabular format which shows the predicted molecular epidemiology of the Foot-and-Mouth Disease (FMD) virus for two options, (i) when single method is selected; (ii) when all the methods are selected.

Case I: Single method (SVM/RF/XGB/ADB) selection. Suppose, SVM machine learning option was selected for predictive analysis. Then the output in following format is obtained.

Sl. No.	Isolate_ID	Serotype	Topotype	Lineage	Mol. Epidemiology (WRLFMD Notation)
1	ICFMD182/2022	O	ME-SA	PANASIA2	O/ME-SA/PANASIA2
2	K154/12	A	AFRICA	G(i, vii)	A/AFRICA/G(i,vii)
3	IND 168/2004	Asia 1	ASIA	G(i-v)	Asia 1/ASIA/G(I-V)

Case II: All the Above methods (AAM), i.e., SVM, RF, XGB, ADB selection was made. Then the output (in WRLFMD symbol) in the following table is obtained.

Sl. No.	Isolate_ID	SVM	RF	XGB	ADB
1	ICFMD182/2022	O/ME-SA/PANASIA2	O/ME-SA/Ind2001	O/ME-SA/Ind2001	O/ME-SA/Ind2001
2	K154/12	A/AFRICA/G(i,vii)	A/ASIA/G(i,vii)	A/ASIA/Sea-97	A/ASIA/G-18
3	IND 168/2004	Asia 1/ASIA/G(I-V)	Asia 1/ASIA/G(I-V)	Asia 1/M/G(I-V)	Asia 1/M/C

SVM: Support Vector Machine; RFF: Random Forest; XGB: Extreme Gradient Boosting; ADB: Adaptive Boosting

Downloading output files: Users can download the output result files (both in .txt and .csv) from the predict tab of the prediction server.

Figure 6. Manual/Help page of the MolEpidPred web prediction server.

MolEpidPred AI server in field FMD outbreaks of Eastern States: The utility of the MolEpidPred AI prediction server is demonstrated in the real field FMD outbreak areas in Eastern parts of India (Bihar, Odisha, and West Bengal). Here, the VP1 sequence data of 15 field FMD virus isolates, collected from ten outbreak locations across eastern parts of India, were used. The viruses were isolated from the clinical samples (tongue and foot epithelium) from cattle, buffalo, and pig collected by the state FMD research centres and Veterinary departments, following the World organization of animal health (WOAH) guidelines. The sequence data of virus isolates were generated through following procedure.

First, epithelial tissue samples were ground with sterile mortar and pestle to prepare a 10% suspension in phosphate buffer saline. Then, total RNA was extracted from the suspensions prepared from the FMD suspected clinical samples (tongue or foot epithelium) using the Viral RNA Mini Kit (Qiagen, Germany) as per the manufacturer's instructions. The extracted RNA was used for the synthesis of cDNA using an oligod(T)$_{15}$ primer and Moloney murine leukemia virus reverse transcriptase enzyme (Promega, USA). The VP1 structural protein coding region was amplified and sequenced as per the established protocol (Applied Biosystem, USA) using the Bigdye V3.1 terminator kit. The VP1 sequence data of these 15 isolates are used in the AI server to predict the molecular epidemiology of the FMD virus causing outbreaks in eastern part of India and results are shown in Table 1.

Table 1. On site demonstration of MolEpidPred AI server.

Isolate ID	Outbreak	Animal	Serotype	Topotype	Lineage	Mol.Epid.
IC02/2018	Keonjhar, Odisha	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC03/2018	Keonjhar, Odisha	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC14/2018	Sambalpur, Odisha	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC15/2018	Sambalpur, Odisha	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC16/2018	Sindupur, Odisha	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC67/2018	Galshi, West Bengal	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC70/2018	Medinapur, West Bengal	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC72/2018	Medinapur, West Bengal	Pig	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC117/2018	Patna, Bihar	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC119/2018	Patna, Bihar	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC120/2018	Patna, Bihar	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC128/2018	Bhoipur, Bihar	Buffalo	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC132/2018	Puraini, Bihar	Buffalo	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC123/2019	Nadia, West Bengal	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001
IC124/2019	Bankara, West Bengal	Cattle	O	ME-SA	Ind2001	O/ME-SA/Ind2001

Our predictive analytics suggested that FMD virus from lineage *Ind2001* of *ME-SA* topotype of *O* serotype dominated the FMD outbreaks in Eastern states including Odisha, West Bengal, and Bihar (Table 1). As per WOAH standard notation, the molecular epidemiology of the FMD virus causing outbreaks in cloven-hoofed animals of Eastern states was found as O/ME-SA/Ind2001. This finding is supported by multiple evidences from the published literature. MolEpidPred tool took fraction of a second to provide such valuable insights into the FMD outbreak in the India states, which would help the researchers for further study.

Conclusion: AI/ML tools such as MolEpidPred have demonstrated extremely high potential and fast way for molecular epidemiology of FMD virus. This tool facilitated rapid and scalable prediction of serotype, topotype, and lineage based on sequence-based algorithms. This finding will guide the government agency and policy makers in the respective states to select appropriate vaccine strain to control the FMD outbreak. In general, the use of AI/ML in molecular epidemiology can transform FMD control from reactive to proactive, precision-based strategies. With further development, calibration, and incorporation into surveillance systems, such as MolEpidPred, it can become the hub of real-time, data-driven disease control and global animal health management.

Acknowledgement: The authors acknowledge the financial support from Indian Council of Agricultural Research (ICAR), Department of Agriculture Research & Education, Government of India and Science & Technology Department, Government of Odisha.

References

[1] T. J. D. Knight-Jones and J. Rushton, "The economic impacts of foot and mouth disease – What are they, how big are they and where do they occur?," *Prev Vet Med*, vol. 112, no. 3–4, p. 161, Nov. 2013, doi: 10.1016/J.PREVETMED.2013.07.013.

[2] S. Das, U. Nayak, S. Pal, and S. Subramaniam, "MolEpidPred: a novel computational tool for the molecular epidemiology of foot-and-mouth disease virus using VP1 nucleotide sequence data," *Brief Funct Genomics*, vol. 24, Jan. 2025, doi: 10.1093/bfgp/elaf001.

[3] S. Das, S. Behera, S. K. Pradhan, and N. R. Sahoo, "A Network Biology Approach for Understanding Foot-and-Mouth Disease Vaccination Response in Cattle using High-Throughput Gene Expression Data," *Curr Biotechnol*, vol. 14, no. 2, pp. 108–122, Jun. 2025, doi: 10.2174/0122115501351126250506053135.

[4] N. J. Knowles and A. R. Samuel, "Molecular epidemiology of foot-and-mouth disease virus," *Virus Res*, vol. 91, no. 1, 2003, doi: 10.1016/S0168-1702(02)00260-5.

[5] S. Das *et al.*, "FMDVSerPred: A Novel Computational Solution for Foot-and-mouth Disease Virus Classification and Serotype Prediction Prevalent in Asia Using VP1 Nucleotide Sequence Data," *Curr Bioinform*, vol. 19, no. 9, pp. 794–809, 2024, doi: 10.2174/0115748936278851231213110653.

[6] L. Sun, X. Kong, J. Xu, Z. Xue, R. Zhai, and S. Zhang, "A Hybrid Gene Selection Method Based on ReliefF and Ant Colony Optimization Algorithm for Tumor Classification," *Sci Rep*, vol. 9, no. 1, Dec. 2019, doi: 10.1038/S41598-019-45223-X.

[7] D. Li, C. Zhang, and J. Cao, "Short-Term Passenger Flow Prediction of a Passageway in a Subway Station Using Time Space Correlations between Multi Sites," *IEEE Access*, vol. 8, pp. 72471–72484, 2020, doi: 10.1109/ACCESS.2020.2988030.

[8] H. J. Duggirala *et al.*, "Artificial intelligence and machine learning in veterinary medicine: a regulatory perspective on current initiatives and future prospects," *Am J Vet Res*, vol. 86, no. S1, pp. S16–S21, Mar. 2025, doi: 10.2460/AJVR.24.09.0285,.

[9] A. L. Beam and I. S. Kohane, "Big Data and Machine Learning in Health Care," *JAMA*, vol. 319, no. 13, pp. 1317–1318, Apr. 2018, doi: 10.1001/JAMA.2017.18391.

[10] J. Guitian, M. Arnold, Y. Chang, and E. L. Snary, "Applications of machine learning in animal and veterinary public health surveillance," *OIE Revue Scientifique et Technique*, vol. 42, pp. 230–241, 2023, doi: 10.20506/RST.42.3366,.

[11] G. Kapalaga *et al.*, "Enhancing random forest predictive performance for foot and mouth disease outbreaks in Uganda: a calibrated uncertainty prediction approach for varying distributions," *Front Artif Intell*, vol. 7, p. 1455331, Nov. 2024, doi: 10.3389/FRAI.2024.1455331/XML.

[12] Z. Souaissi, T. B. M. J. Ouarda, and A. St-Hilaire, "Non-parametric, semi-parametric, and machine learning models for river temperature frequency analysis at ungauged basins," *Ecol Inform*, vol. 75, p. 102107, Jul. 2023, doi: 10.1016/J.ECOINF.2023.102107.