

Research Article

Pandemic Analysis and Preparedness Using a Three Layer Machine Learning Approach

Artika Singh^{1,2} and Manisha Jailia¹

¹Department of Computer Engineering, Banasthali Vidhyapith, Rajasthan, India

²Narsee Monjee Institute of Management Studies, Mumbai, India

Article history

Received: 18-08-2025

Revised: 12-09-2026

Accepted: 23-09-2026

Corresponding Author:

Artika Singh

Department of Computer

Engineering, Banasthali

Vidhyapith, Rajasthan, India

Email: artika.singh@nmims.edu

Abstract: Outbreak analysis provides research-based information and data to enable scientific decision-making to manage an effective response to a pandemic. Pandemic analysis has its own set of challenges, such as limited data, the need for real time analysis, and the complexity of outbreak patterns. This study proposes a hybrid approach to manage such a spread through a 3-layer machine learning-based outbreak prediction model. The proposed model incorporates pre-processing of data, data standardisation, feature selection, outbreak prediction, and a tool for effective pandemic management. The pre-processing step of the proposed model is to collect the pandemic data, remove irregularities, clean the data, fill in missing values, and format the data in a consistent manner for data analysis. The pre-processing can be accomplished through statistical procedures (e.g., central tendency and statistical dispersion) for data standardisation. A feature selection step, where the most relevant features are tested and selected for outbreak prediction, uses a bioinspired hybrid optimisation algorithm, the Spotted Hyena Optimiser (SHO) and Seagull Optimisation Algorithm (SOA), to minimise the likelihood of overfitting. The final step in the outbreak prediction model incorporates a 3-layer machine learning approach comprised of an O-SVM (Optimised Support Vector Machine) initial layer for boundary detection and classification, a KNN (k-Nearest Neighbors) second layer to perform similarity matching and, lastly, XGBoost as a third layer for prediction of spread severity. The predicted outcome is compared with standard machine learning algorithms, in which the proposed model provides better results in terms of the highest accuracy (0.93), fitness (0.96), and the lowest error (6.9%).

Keywords: Outbreak Analysis, SHO, SOA, Machine Learning, Support Vector Machines, K-Nearest Neighbors, XGBoost

Introduction

The outbreak analysis provides individuals or groups with the necessary information to make the best decisions regarding their response to the outbreak (Ramanathan et al., 2020; Altuntas and Gok, 2021). The first step of outbreak analysis is to collect data from legitimate sources and to consider all possible dimensions, so that an understanding of the outbreak is developed (Yasir et al., 2020; Kargar et al., 2020). In this study, the data were collected from available open-source databases, including health organisations (WHO, CDC), government reports, scientific journals, and historical records. Six pandemics, Spanish Flu, SARS, Zika, Swine Flu, Ebola and Covid 19

were selected for the analysis, based on the variation of timeline, spread nature, geographical regions and severity.

Spread analyses can include demographic information (i.e., where individuals live); age, sex, ethnicity; and other demographic factors; as well as details on transmission route(s); symptoms; and any possible treatments being administered (Shaker et al., 2020; El Asnaoui et al., 2021). In this work, common features were selected from the six different pandemics, such as date, time, location, number of infected cases, number of recovered cases, number of deaths, symptoms, and spread characteristics.

By analysing such spreads, it provides decision-makers with accurate and timely information about the

events occurring during an outbreak, and facilitates the development of solutions to control the situation and provide assistance to the population (Fong et al., 2020; Tirkolaee et al., 2021; Zadeh et al., 2019). Outbreak analyses require resource-intensive efforts to obtain, assess, and report the data contained within each outbreak (Virghileanu et al., 2020; Uddin et al., 2021). By developing a system for gathering current reports and accurate data, decision-makers will have access to complete and precise data (Paul and Chowdhury, 2021). In COVID-19, decision-makers are able to find high-risk populations and geographic locations; trace the transmission of a particular pathogen; and create an effective solution for emerging infectious disease outbreaks (Rahman et al., 2021; Yadav et al., 2021). Also, they are able to identify the effectiveness of interventions, gauge the effect of an outbreak on the populace, and subsequently inform resource allocation based on data collected (Belhadi et al., 2021; Das et al., 2020).

The COVID-19 pandemic has highlighted how important it is to have accurate and timely data as a basis for making decisions around outbreak control and mitigation (Ivanov, 2020).

The goal of this work is to provide a generic platform that can handle any kind of spread in future pandemics or outbreaks by using statistical and analytical tools on historical pandemics. It provides a comparative predictive analysis of selected common features from the six pandemics or epidemics, which are Spanish Flu, SARS, Zika, Swine Flu, Ebola, and Covid 19.

Key objectives of the research are as follows:

- Compiling a comprehensive repository of common features of all previous pandemic data for cleaning and preprocessing to prepare a generic dataset for all kinds of spread
- Feature selection from the common pool of features by using a hybrid optimiser, Seagull Customised Hyena Optimisation Algorithm (SCHOA), which is a combination of SHO and SOA
- Selected features were provided to a three-layer machine learning approach, where the first layer is an Optimised Support Vector Machine (O-SVM) for boundary detection and classification, the second layer is a k-Nearest Neighbor (KNN) for similarity matching, and the third layer is XGBoost for the prediction of spread
- The results are compared with existing standard machine learning predictive models for the result analysis

Literature Review

The pandemic emphasises the need for examining structural processes and unpredictable systems, with consideration of the complexity, uncertainty, contingent

and situated nature of how they operate (Leach et al., 2021). Worldwide, the COVID-19 pandemic has forced all countries to be nimble and flexible when responding to the pandemic; this involves the centralisation of decision-making, timing of the implementation of all policies, the level of autonomy of decisions made in response to COVID-19, and how to manage stability during times of change and vice versa. To create an environment where adaptive governance can occur successfully, agile governance must balance multiple factors, including but not limited to: Speed of decisions and sound analysis, centralised versus decentralised decision-making, the need to create innovative solutions against the desire to maintain current practices (bureaucracy), and science versus politics (Janssen and Voort, 2020).

Gamal et al. (2022) developed an improved structure to create an alert system for pandemics based on social networks. The proposed plan has three connected sub-parts, including: Collection, pre-processing, classification, feature extraction, and classifier training; and evaluation of the level of awareness. Aspects influencing how accurately information about SARS-CoV-2 is reported, such as insufficient means to collect data and different sources providing inconsistent results, highlight why this model is relevant to today's issues of reporting pandemic information.

In their research, Islam et al. (2021) developed an innovative method for monitoring pandemic outbreaks via drone surveillance, minimising the need for human intervention. Their approach utilised two separate drone swarms for various monitoring functions along with a lightweight blockchain and a two-step security method for authenticating the monitored parties. As evidence of their method's effectiveness, the authors provided an example that utilised the COVID-19 pandemic.

According to Castiglione et al. (2021) the role of IoT in helping to mitigate the Transmission of COVID-19 is still being determined. In their research, they describe a "cloud-based architecture for real-time patient tracking and detection" which encompasses collecting data from IoT devices and utilizing that data to create predictive models of how a particular treatment will work for each patient through analysis performed using machine learning algorithms and administering that treatment to the patients researched by health professionals.

Pei et al. (2022) identified a dynamic allocation strategy for medical resources during epidemic outbreaks in 2022. They developed a multi-stage and multi-type medical service network model and solved it using fluid queuing approximations to create a stochastic optimisation problem. Using actual data from Wuhan, their proposed algorithm ensures healthcare services are available and continues to maintain guaranteed service levels despite resource constraints.

Raj et al. (2022) conducted an exploration of Emerging Economies' Manufacturing Supply Chain

Management, during and after the COVID-19 Pandemic. They developed a conceptual framework based on dynamic capability theory and identified 10 major supply chain challenges using a comprehensive literature review and discussions with experts. Additionally, they used the Grey DEMATEL Method to analyse the relationships between those challenges and provided guidance to practitioners and academics on how to mitigate them.

The study by López and Rodó, (2021) has described an improved SEIR-based epidemiologic modelling system that simulates the COVID-19 outbreak in Italy and Spain. The model incorporates the latent period during which a person can transmit infection without displaying symptoms, and incorporates consideration for the adjustable proportion of containment measures used in controlling the spread of the disease. The model parameters are adjusted based on the data available for reported cases to allow for uncertainty in the counts of symptomatic cases. The purpose of using this modified SEIR model is to assess and identify the best confinement strategies and scenarios that would limit the rate of infections and resultant deaths during the initial phase of the epidemic outbreak.

Pantano et al. (2020) have evaluated the effects of COVID-19 on retailers. The authors compile a list of ways retailers are responding to the crisis. They give tips and examples to help retailers deal with this unique situation from both the consumer and managerial points of view.

Pappas and Glyptou (2021) investigated the attributes influencing the accommodation choices made by tourists who travelled in 2020 during the COVID-19 epidemic. The researchers used data from a survey of 385 Athenians to conduct Fuzzy Set Qualitative Comparative Analysis (FSQCA). The four solutions identified by the study:

- (i) Health and Safety
- (ii) Price-Quality nexus
- (iii) Risk
- (iv) Quality related to Health & Safety have direct implications for the accommodation industry

All the above research and studies show that pandemics affect all sectors of society and disrupt the whole structure. To manage and minimise the effect of such a spread, one needs a generic framework with accurate and current data, having robust analytical and predictive tools, which help to decide on the future steps and actions to control these outbreaks and manage them effectively.

Datasets and Sampling

This study has considered datasets from six pandemics, which are Spanish Flu, SARS, Zika, Swine Flu, Ebola, and Covid-19, which is a continuation of my previous work. The reason behind selecting these

outbreaks is to cover all kinds of diversity in terms of pathogen types, spread characteristics, geographical region coverage, severity, impact on the population, and time period. These datasets were collected from different resources and were later combined to make a generic outbreak dataset (Singh and Jailia, 2026). The details of the considered datasets are presented in Table 1.

In our previous work, the above datasets were considered, and a combined feature pool was created from all these spreads to make a common outbreak dataset for any kind of spread. There are features related to the nature, clinical information, environmental, and socioeconomic factors. A few of the common features were date, time, location, number of infected cases, number of recovered cases, number of deaths, symptoms, and spread characteristics. Along with these features, a few derived features were calculated, which are the number of infected cases (region-wise), the number of deaths (region-wise), the timeline, Case Fatality Rate (CFR), Basic Reproduction Number (R_0), Pandemic Duration, Global Mortality Estimates, Infection Fatality Ratio (IFR), and Prevention and Control Measures. By combining all these 40-47 features, feature analysis, feature importance, and correlation among these features are performed, and results are extracted for further steps.

In this research, the same dataset is considered, and after standard preprocessing, the sampling technique is applied. Here, the data sets are not independent and identically distributed; each point is a time-ordered observation embedded in a specific pandemic wave, geography, and pathogen biology. Thus, applying random sampling may cause data leakage and ruin the structure. To address all challenges of temporal autocorrelation, class imbalance, and pandemic heterogeneity, Temporal resampling methods and oversampling methods are applied for sampling. Time Series Cross-validation is the best standard technique for time-series machine learning evaluation, and Adaptive Synthetic Sampling, which is an extension of Synthetic Minority Oversampling Technique (SMOTE), is best for imbalanced datasets. In the training dataset, the following sampling techniques are applied. Time Series Cross-validation utilises an expanding window approach, creating multiple train-test splits while maintaining the time order. This method ensures that future data is never used to predict the past, which makes it ideal for time-dependent datasets. This uses multiple evaluation periods and maintains temporal order, which makes it more robust than a single split (Deng, 2023).

Adaptive Synthetic Sampling (ADASYN) is more focused on the decision-making area where the models are more confused. It generates more synthetic samples for such areas where the minority class is sparsely represented. This will help the classifier to better discriminate between the minority and majority classes.

Table 1: Dataset details (Singh and Jailia, 2026)

Epidemic or Pandemic	Pathogens	Symptoms and spread characteristics	Global Reach	Total Cases	Total Deaths	Time period	Sources	Links
Spanish Flu	H1N1	Cold and cough. Spread through water bodies; human-to-human through sneezing and coughing. Infectious disease through air and water.	World wide	500 million	100 million	1918-1920	WHO FluNet and Kaggle	who.int/tools/flunet and kaggle.com https://github.com/imdevskp/sars-2003-outbreak-data-webscraping-code
SARS	SARS-CoV-1	Human-to-human transmission through breathing, talking, or singing.	World wide	8,096	783	2002–2004	WHO SARS Archive	ourworldindata.org
Swine Flu	H5N1	Transmitted primarily from infected animals to humans through direct contact	U.S., India, Pakistan, Philippines, and Maldives	150,000 to 575,000	18,449	2013-2019	Our World in Data (OWID)	ourworldindata.org
EBOLA	Ebola Virus	Human-to-human through Blood, any other body fluid of an infected person	Countries of West, Central and East Africa	28,652	11,323	2013-2016	WHO Ebola Situation Reports and Kaggle	kaggle.com and https://pmc.ncbi.nlm.nih.gov/articles/PMC8342406/
Zika	Zika virus	Spread by daytime-active Aedes mosquitoes, and its symptoms are similar to a very mild form of dengue fever. 80% of cases are estimated to be asymptomatic. Infectious disease through air, water, and touching. Human-to-human transmission through breathing, talking, or singing, and touching.	Africa to Asia	3 million	51	2015-2016	CDC Zika Dataset on Kaggle	kaggle.com/cdc/zika-virus-epidemic
Covid -19	SARS-CoV-2	Infectious disease through air, water, and touching. Human-to-human transmission through breathing, talking, or singing, and touching.	Africa, Antarctica, Asia, Europe, North America, Oceania, South America	500-700 million	18.2–33.5 million	2019-2023	Our World in Data (OWID) and WHO COVID-19 Dashboard	ourworldindata.org/covid-cases

Proposed Methodology

This study suggests a three-layered machine-learning hybrid approach to predicting outbreak events through:

- (1) Data Pre-Processing
- (2) Feature Extraction
- (3) Feature Selection
- (4) Outbreak Prediction

The proposed approach represents another means of effectively managing an epidemic.

In step one, the combined pool of datasets has been gathered and cleaned; the next step is pre-processing by applying statistical methods for normalising (i.e., scaling)

the data (e.g., using min/max scaling) before any further analyses. In step two, data standardisation is performed with the help of measures of central tendency and measures of statistical dispersion.

The next step is the most relevant feature extraction for the prediction of an outbreak. For this, a hybrid optimisation algorithm is used, which is the combination of the Spotted Hyena Optimiser (SHO) and the Seagull Optimisation Algorithm (SOA). The advantage of using a hybrid optimisation algorithm is that it decreases the likelihood of overfitting.

The final step is outbreak prediction, which consists of a three-layer machine learning structure that has a Support Vector Machine (O-SVM) in the first layer for boundary detection and classification. The second layer is the k-

Nearest Neighbors (KNN) algorithm for similarity matching, and the third layer is XGBoost for outbreak prediction.

The overall architecture for the Outbreak Prediction Machine Learning Model is shown in Figure 1.

Data Pre Processing

The common pool of pandemic datasets has 47 raw features provided in the previous part of the research, having five categories of features which are distributed as 14 epidemiological, 12 clinical, 8 environmental, 7 intervention, and 6 socioeconomic. For these pre-processing methods, such as data cleaning and normalisation, min-max normalisation is applied. No fixed holdout partition was employed. All performance evaluation was conducted via walk-forward temporal cross-validation, ensuring that model assessment at every fold reflects only information available before the test period.

Data Cleaning

Data cleaning is one of the most critical components of outbreak analysis because it eliminates, corrects, or modifies inaccurate, incomplete, or irrelevant information throughout the dataset. This is vital because the data provided in outbreak analysis should be both correct and reliable. Data cleaning consists of various methods, such as removing duplicate records, identifying and addressing missing data, addressing inconsistent data, and removing outliers. As pandemic data vary with time, phase labelling is required for each data point, such as emergence, peak, and decline, based on the R_0 value.

Min Max Normalisation

Min-Max normalisation transforms data collected from years, like outbreaks during a pandemic, into a

standard scale of 0-1 or another common range so that patterns of events can be quickly identified and decisions made based on informed judgment about what to do next. For example, A min-max normalisation application for an outbreak study would be to allow researchers to compare cases across regions or times by scaling each data point. The researcher may quickly view regions that experienced very high or low numbers of cases; likewise, by performing min-max normalisation, the impact of outliers on the dataset can be reduced.

After preprocessing, the analysed data is put through to the feature extraction section of the process, which will extract the most relevant data items for use in creating analytic models or conducting further statistical analysis.

Data Standardisation and Feature Extraction

In this study, the features will be extracted and tuned based on the central tendency and dispersion of the data (i.e., mean, median, mode, and variance). In this research dataset, which is a mixture of various dynamics, ranges, time frames, and frequencies due to the inclusion of different kinds of pandemic datasets, there is a necessity to apply central tendency and dispersion techniques for an unbiased and balanced data sample.

Feature extraction translates raw pandemic surveillance data into a structured numerical feature matrix. 47 candidate features are engineered across five categories. Each observation represents one country-week pair. After pre-processing, the dataset contains 10,310 observations: Spanish Flu $n = 380$, SARS $n = 290$, Swine Flu $n = 3,100$, Zika $n = 1,820$, COVID-19 $n = 4,200$ (post-down sampling from 19,000), Ebola $n = 520$.

Time series cross-validation then partitions this matrix into temporally ordered train-test folds to ensure no future data leaks into training.

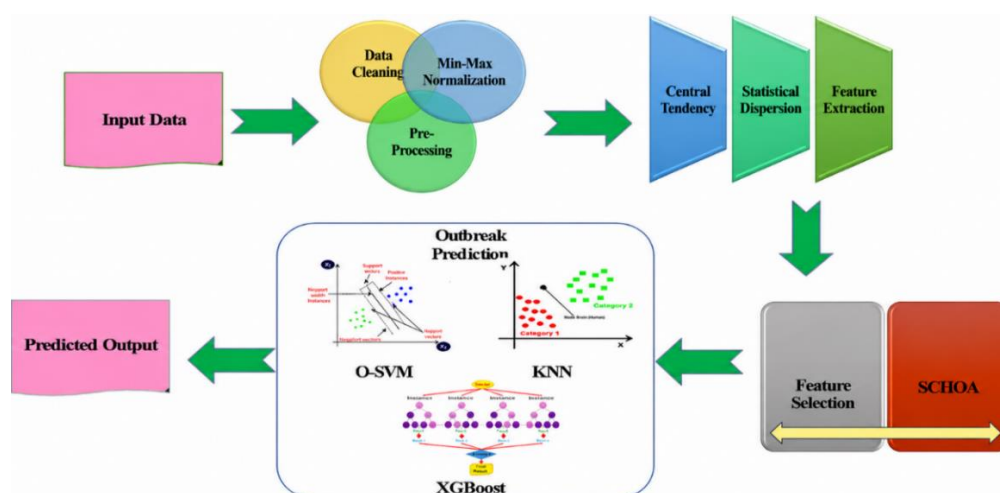


Fig. 1: Overall proposed architecture

The expanding window ensures every fold trains on all available historical data. This is similar to mimicking how a real-time surveillance system would actually be updated. The gap buffer G (incubation-period-specific per pathogen) is what separates this from naive cross-validation.

Table 2 summarises the pandemic-specific temporal cross-validation settings, including the number of folds, initial training window, test step, gap buffer, and feature matrix dimensions.

Temporal fold construction using time series cross-validation is shown in the following algorithm.

Time series cross-validation (expanding window)	
Fold 1: Train on $[t_1 \dots t_k]$, Test on $[t_{k+1} \dots t_{k+m}]$	
Fold 2: Train on $[t_1 \dots t_{k+m}]$, Test on $[t_{k+m+1} \dots t_{k+2m}]$	
Fold i : Train on $[t_1 \dots t_{k+(i-1)m}]$, Test on $[t_{k+(i-1)m+1} \dots t_{k+im}]$	
Training window grows with each fold — never shrinks.	
Fold structure for pandemic p:	
Train $_i = [t_1 \dots t_{k+(i-1)m}]$	
expanding window Gap $_i = [t_{k+(i-1)m} \dots t_{k+(i-1)m+G}]$	
leakage buffer Test $_i = [t_{k+(i-1)m+G+1} \dots t_{k+im+G}]$	

Fold 1 training set covers the earliest m time steps. Each subsequent fold adds the previous test window to the training data set to train the model. This way, the model sees progressively more epidemic history. Gap buffer G is set to the maximum incubation period per pathogen (2 weeks for COVID-19, 3 weeks for Ebola), which purges boundary autocorrelation. Scalers fitted on the training data set only. As discussed earlier, min-max and z-score parameters from the training fold were applied identically to the test fold. Never re-fitted on test data.

Records were excluded if the case count was zero for three or more consecutive weeks (treated as surveillance gaps). Missing values were imputed via linear interpolation for Spanish Flu (gaps ≤ 4 weeks) and forward-fill for all other pandemics. Records with greater than 30% feature missingness were dropped. Final exclusion counts: Spanish Flu 12 records, SARS 4, Swine Flu 18, Zika 31, COVID-19 29, Ebola 9.

SARS uses $K = 3$ because the 8-month epidemic cannot be partitioned into 5 non-overlapping folds with adequate test coverage (minimum 40 test observations per fold). COVID-19 training N increases from 420 (Fold 1)

to 3,220 (Fold 6) as the expanding window accumulates six waves of data. Gap buffer G is expressed in days for all pandemics regardless of surveillance resolution. For weekly-resolution pandemics (Spanish Flu, Swine Flu, Zika, Ebola), G is rounded up to the nearest whole week.

Feature Selection

The Spotted Hyena Optimiser (SHO) and Seagull Optimisation Algorithm (SOA) are considered hybrid optimisation methods for feature selection. SOA has a strong exploration capability, which helps in exploring the feature space broadly, as it has good global search coverage. Whereas SHO is strong in exploitation, meaning good local search refinement, which helps in refining the best feature subsets found. Hybrid methods provide improved performance on the task of selecting important features for predicting outbreaks, and give an exploration-exploitation balance.

Seagull Customised Hyena Optimisation Algorithm (SCHOA)

In this work, the proposed bioinspired hybrid optimisation method strategically combines SHO and SOA as a metaheuristic optimisation algorithm to develop a solution for many types of challenging optimisation problems. The stalking behaviour of spotted hyenas makes them an ideal choice for creating this algorithm, as these clever predators typically hunt together and employ many hunting strategies to overcome their prey. In contrast, Seagull Optimal Strategy is based on the literature that describes how seagulls coordinate their angle of attack during flight to locate prey by adjusting their flight patterns according to prevailing environmental conditions.

To conclude, SCHOA will remove irrelevant features and redundant attributes, reduce dimensionality, improve prediction accuracy, and minimise the computational complexity.

Role of SHO

The Spotted Hyena Optimiser (SHO) provides group hunting behaviour, an encircling prey mechanism, local search refinement, and fast convergence towards the optimal region (Ghafori and Gharehchopogh, 2022). These actions will contribute to feature selection by refining candidate feature subsets, removing redundant attributes, and improving local optimum exploitation.

Table 2: Feature extraction through time series cross-validation

Pandemic	Folds K	Init train window	Test step m	Gap G	Feature matrix shape
Spanish Flu	4	12 weeks	12 weeks	14 days	380×47
SARS	3	8 weeks	50 days	14 days	290×47
Swine Flu	5	10 weeks	10 weeks	7 days	3,100×47
Zika	5	14 weeks	14 weeks	14 days	1,820×47
COVID-19	6	4 weeks	25 days	14 days	4,200×47
Ebola	4	16 weeks	16 weeks	21 days	520×47

SHO — exploitation core (Spotted Hyena Optimiser)

Position vector of hyena i at iteration t :
 $X_h(t+1) = X_{prey}(t) - A \cdot D_h$
Distance from prey: best feature subset
 $D_h = |B \cdot X_{prey}(t) - X_h(t)|$
Coefficient vectors:
 $A = 2h \cdot r_1 - h$
 $B = 2 \cdot r_2$
 $h = 5 - (t \cdot 5/T_{max}) \leftarrow$ linearly decays
exploration $\rightarrow$ exploitation
Clan aggregation (exploitation convergence):
 $X_{cluster} = (X_{h1} + X_{h2} + X_{h3}) / N_h$
Where:
 r_1, r_2 Random vectors $\in [0, 1]$
 h Control parameter decaying from 5 $\rightarrow$ 0 over T_{max} iterations
 N_h Number of hyenas in the cluster (typically 5)
 X_{prey} Best solution (feature subset) found so far

Role of SOA

The Seagull Optimisation Algorithm (SOA) provides spiral mitigation behaviour, collision avoidance, and global search diversification (Pathak et al., 2025). These will help in feature selection by exploring a broader search space, preventing premature convergence, and escaping local minima.

SOA — exploration core (Seagull Optimization Algorithm)

Migration phase (collision avoidance):
 $C_s(t) = A_s \cdot X_s(t)$
 $A_s = f_c - (t \cdot f_c/T_{max})$
Attack phase (spiral movement towards best solution (X_{best})):
 $M_s(t) = B_s \cdot (X_{best}(t) - X_s(t))$
 $B_s = 2 \cdot A_s^2 \cdot r_3$
Spiral attack trajectory (3D helix):
 $x' = r_s \cdot \cos(\theta)$,
 $y' = r_s \cdot \sin(\theta)$,
 $z' = r_s \cdot \theta$
 $r_s = e^{b\theta}$,
 $\theta \in [0, 2\pi]$
 $D_s(t) = |X_{best}(t) - X_s(t)|$
Final position update:
 $X_s(t+1) = D_s(t) \cdot x' \cdot y' \cdot z' + X_{best}(t)$
Where:
 A_s Frequency of movement, decays from $f_c \rightarrow 0$ ($f_c=2$)
 b Spiral shape constant (typically $b=1$)
 θ Random angle $\in [0, 2\pi]$
 r_3 Random scalar $\in [0, 1]$

Hybrid SCHOA

Hybrid SHO-SOA works for feature selection by population initialisation, which generates random feature subsets. Next, SOA searches the feature space broadly to identify promising regions. This provides diverse candidate feature subsets, and SHO refines them to produce an optimised minimal feature subset. Finally, it evaluates the fitness of each subset.

Hybrid SHO-SOA fusion mechanism — cosine-annealed adaptive weighting

Position update rule at iteration t :
 $X_{new}(t+1) = \alpha(t) \cdot X_{SHO}(t+1) + (1-\alpha(t)) \cdot X_{SOA}(t+1)$
Adaptive weight $\alpha(t)$ — exploration $\rightarrow$ exploitation transition:
Cosine-annealed weight
 $\alpha(t) = 0.5 \cdot (1 + \cos(\pi \cdot t / T_{max})) \in [0, 1]$
(SHO-dominant, exploitation first) $\rightarrow$ $\alpha=0.0$ at $t=T$ (SOA-dominant, global refinement last). Fitness function for feature subset S (binary encoding):
 $f(S) = \gamma \cdot (1-E(S)) + (1-\gamma) \cdot (1-(|S|/|F|))$
Where:
 $\alpha(t)$ Cosine-annealed weight: starts at 1 (SHO-dominant), ends at 0 (SOA-dominant)
 $E(S)$ Classification error of subset S on validation fold
 $|S|$ Number of selected features
 $|F|$ Total features in original space
 γ Trade-off parameter ($\gamma=0.9$ prioritises accuracy over compactness)
Binarization via sigmoid transfer:
 $S^* \ i[j] = 1$ if $\alpha(X_{new}[j]) > \text{rand}$ else 0

The key insight is the cosine-annealed $\alpha(t)$ weight that smoothly transitions between the two optimisers. Early in the search ($t \rightarrow 0$), $\alpha \approx 1$ so SHO dominates and exploits the local neighbourhood aggressively. Late in the search ($t \rightarrow T_{max}$), $\alpha \approx 0$, so SOA's spiral attack takes over for fine-grained global convergence. This avoids the premature stagnation that either optimiser suffers alone. The binary fitness $f(S) = 0.9 \cdot (1-E(S)) + 0.1 \cdot (1 - (|S|/|F|))$ is deliberately weighted 9:1 toward accuracy over compactness, which is correct for a medical prediction setting where missing an outbreak is far costlier than using an extra feature. Where: S = binary feature mask (1=selected, 0=dropped) $E(S)$ = 5-fold CV classification error on training fold $\in [0, 1]$. Higher $f(S)$ = better subset; $X_{best} = \text{argmax } f(S)$ across all agents and iterations.

An observation is labelled outbreak ($y = 1$) if the 7-day rolling incidence rate exceeds the 75th percentile of that pandemic's training-fold incidence distribution AND the estimated R_0 exceeds 1.2 in the same week. All other observations are labelled endemic ($y = 0$). Before oversampling, outbreak proportions range from 8.3% (SARS, Ebola) to 25.1% (COVID-19 Fold 6), reflecting the genuine rarity of outbreak-phase observations in surveillance data. After that, distributions are 50/50 by design. Actual oversampled training set size varies by fold.

The hybrid feature selection technique reduces the feature space from ~ 47 raw features to only 19 features, which is, on average, a 60% dimensionality reduction. Zika is the hardest case (21 features, slightly lower AUC) because vector-borne transmission introduces environmental covariates (mosquito density, temperature, rainfall) that don't appear in airborne pandemic data.

Optimal features selected from the considered datasets are common symptoms of spread, geographical regions widely affected, and the spreading medium (Table 3).

Table 3: Optimal features and mapping of measures

Pandemic	Mode of Spread	Scale of Cases	Geographic Reach	Common Symptoms	Key Prevention & Control Measures
Spanish Flu	Airborne / Droplet	Very High	Worldwide	Sore throat, headache, fever, mouth/nose bleeding.	Face masks, avoiding overcrowding, social distancing.
SARS	Airborne / Droplet	Very High (Potential)	Worldwide	Fever, dry cough, headache, difficulty breathing.	N95/FFP2 respirators, ventilation, UVGI, travel limits.
COVID-19	Airborne / Droplet	Very High	Worldwide	Fever, cough, fatigue, loss of taste/smell.	Vaccination, masking, distancing, ventilation.
EBOLA	Body Fluids	Low	Regional (Africa)	Fever, vomiting, bleeding, organ impairment.	Strict IPC protocols, contact tracing (21 days), surveillance.
ZIKA	Vector (Mosquito)	High	Tropical	Asymptomatic (80%), fever, rash, joint pain.	Vector control, reducing breeding sites, bite prevention.
Swine Flu (H5N1)	Animal Contact	Moderate	Hotspots	Fever, lethargy, sneezing, eye inflammation.	Personal hygiene, food safety, reporting sick animals.

From these selected variables, the (GB) were mapped with a list of prevention and control measures used previously.

From the above tabular results, it can be concluded that if the transmission mode is airborne and affects the respiratory system, then such spreads are more severe in terms of spreading rate and reach a geographical region. But if the mode is mosquitoes, then the spread depends on the climate of the geographical region, and the severity will be moderate. Whereas if the mode is through animal contact, then the spread is low to moderate, depending on population habits. At the end, if the mode is body fluid, then the spread is restricted to a certain geographical region, and the severity is low. Key prevention measures were mapped for further actions and decision-making.

The 8 universal core features are R_0 estimate, case incidence rate, CFR, doubling time, wave phase label, population density, healthcare capacity, and reporting completeness. These features survive SHO-SOA selection across every pandemic, every fold, every transmission type. These 8 features, along with other pandemic-specific features, will now be used as input features for a three-layer predictive model, which can predict how likely the occurrence of an outbreak is (target variable).

Outbreak Prediction

This research work aims to develop a generic predictive model using a three-layered machine learning approach for analysing disease outbreaks. This three-layered approach uses different machine learning algorithms (Optimised Support Vector Machine (O-SVM), K-Nearest Neighbors (KNN), and XGBoost) for predicting the spread. The role of O-SVM is to classify the outbreak type and detect the non-linear boundaries. Whereas KNN takes the classified outbreak probabilities and creates a classification based on instance-based similarity by using historical pandemic data. The final

decision is provided by XGBoost, which provides an ensemble decision with residual correction.

To prevent meta-feature leakage in the stacked architecture, SVM and KNN predictions used as XGBoost inputs were generated using out-of-fold prediction. For each training fold i , a five-fold inner cross-validation was applied to the training data of fold i . SVM and KNN were trained on four inner folds and generated predictions on the one held-out inner fold. The concatenated out-of-fold predictions formed the meta-feature matrix passed to XGBoost. At test time, SVM and KNN were retrained on the full training fold i before generating test-fold meta-features.

Optimised Support Vector Machines (O-SVM)

Support Vector Machines (SVMs) are a widely used machine learning technique that can classify and predict a response variable based on given input. This algorithm uses hyperplanes to distinguish between two groups of data points, maximising the distance or margin between each group. SVM excels at small-to-medium datasets with high-dimensional feature handling, and the considered historical pandemic data is exactly what it looks like. In SVMs, the weights represent how much influence or importance is given to each of the independent variables used in the analysis. The optimisation of the parameter weight allocations is the most critical factor that determines how well an SVM will perform. To optimise these weight values for the SVM, the SCHOA method is used. The mathematical presentation of the SVM algorithm is as follows:

Optimised SVM (boundary detection)

Decision function:

$$f(x) = \text{sgn}(\sum_i \alpha_i y_i K(x_i, x) + b)$$
RBF kernel:

$$K(x, z) = \exp(-\gamma \|x - z\|^2)$$

Soft-margin primal:
 $\min \frac{1}{2} \|w\|^2 + C \cdot \sum \xi_i$ s.t. $y_i(w \cdot x_i + b) \geq 1 - \xi_i$, $\xi_i \geq 0$
 Platt probability:
 $P(y=1|x) = 1 / (1 + \exp(-A \cdot f(x) + B))$

RBf kernel $K(x,z) = \exp(-\gamma \cdot \|x-z\|^2)$ implicitly maps the 19-dimensional feature vector into an infinite-dimensional Hilbert space. In this space, the non-linear epidemic boundary becomes a linear hyperplane. γ controls the reach of each training point's influence, such that small γ = broad influence (smooth boundary), large γ = narrow influence (jagged boundary). SHO-SOA finds the γ that best fits each pandemic's boundary geometry. SVM finds the hyperplane $w \cdot x + b = 0$ that maximises the margin (distance between the two class support vectors). The soft-margin formulation $\min \frac{1}{2} \|w\|^2 + C \cdot \sum \xi_i$ allows slack variables ξ_i for misclassified points. C controls the trade-off, such that high C = penalise misclassification heavily (tight fit, risk overfitting); low C = allow more violations (smoother boundary, better generalisation). Optimised C balances this per pandemic. Raw SVM output is a signed distance from the hyperplane. Platt scaling fits a sigmoid $\sigma(f(x)) = 1/(1+\exp(-Af(x)+B))$ to convert this distance into a calibrated probability $P(\text{outbreak}|x) \in [0,1]$. This probability is the critical output; afterwards, it becomes a meta-feature for Layer 2 and Layer 3.

K-Nearest Neighbours (KNN)

In outbreak analysis, KNN is an example of a supervised learning algorithm that is suitable for categorisation and similarity matching. SVM draws a global boundary but has no memory of which specific historical outbreak a new observation resembles. KNN addresses this by asking: Among all training observations, which k are most similar to the current input? If those k neighbours were predominantly outbreak-positive, the model anchors its prediction to real historical precedent. It works on local neighbourhood structure by capturing that a novel pathogen with Zika-like R_0 and Ebola-like CFR should pattern-match accordingly. For novel pathogens, this is the most critical layer as it prevents the model from extrapolating into impossible territory by ensuring predictions are always grounded in historical analogues.

The mathematical process of KNN for similarity matching is outlined below:

KNN (Historical similarity memory)

Weighted vote:
 $\hat{y}_{\text{KNN}} = \text{argmax}_c \sum_{\{j \in N_k(x)\}} w_j \cdot I(y_j=c)$
 Inverse distance weight:
 $w_j = 1 / d(x, x_j)$
 Euclidean distance (z-scored):
 $d(x, x_j) = \sqrt{\sum_i (x_i - x_{ji})^2}$
 KNN similarity score:
 $\text{KNN score} = \sum_{\{j: y_j=1\}} w_j / \sum_i w_i \in [0,1]$

For each test observation x , compute Euclidean distance to every training sample: $d(x, x_j) = \sqrt{\sum_i (x_i - x_{ji})^2}$ across all 21 dimensions (19 SHO-SOA features + SVM probability + SVM label). The SVM probability dimension adds a boundary-awareness signal where two observations equally distant in feature space but on opposite sides of the SVM boundary are pulled apart in this 21-dimensional space. Select the k^* nearest training samples $N_k(x)$. For each pandemic, k^* is tuned by SHO-SOA, such as a larger k for COVID-19 (rich training data, stable neighbourhoods) and a smaller k for Ebola (sparse data, sensitive to outlier neighbours). Each neighbour brings its actual pandemic label and its historical outbreak context. Each neighbour votes for its class weighted by $1/d(x, x_j)$, where closer neighbours vote more heavily. The class score is: $S_c = \sum_{\{j \in N_k(x)\}} (1/d(x, x_j)) \cdot I(y_j=c)$. This prevents a single distant outlier from overriding five close, consistent neighbours. The raw outbreak score S_1 (normalised by $\sum$ weights) becomes the KNN similarity score, which is a continuous value in $[0,1]$ analogous to the SVM probability. This score, not just the hard class label, is passed to Layer 3 as a meta-feature, giving XGBoost a richer signal.

The training set contains all six pandemics. When KNN retrieves neighbours for a Zika observation, it may find Spanish Flu observations with similar R_0 and doubling time in its k -neighbourhood. This cross-pandemic similarity is intentional to make the model learn that airborne and vector-borne outbreaks share universal early-warning signatures despite different transmission mechanisms.

From Table 4 it can be easily understood that a high spread rate covers more of the geographical reach, depending on the mode of transmission, such as airborne. At the same time, other modes are restricted to a particular region, which will further depend on the climate and population culture.

Table 4: Pandemic spread region-wise with scale of cases

Mode of Spread	Spread	Scale of Cases	Geographic Reach
Airborne/Droplet	Spanish Flu, SARS, COVID-19	Very High (Millions)	Worldwide
Body Fluids	Ebola	Low (Thousands)	Regional (Africa)
Vector (Mosquito)	Zika	High (Millions)	Tropical (Africa to Asia)
Animal Contact	Swine Flu (H5N1)	Moderate	Specific hotspots

XGBoost

SVM and KNN each make systematic errors in different regions of feature space. SVM struggles with environmental covariates (Zika); KNN struggles with sparse training data (Ebola, SARS). XGBoost learns these failure patterns by treating SVM and KNN outputs as additional input features alongside the original 19, then sequentially building trees that correct residual errors. Each boosting round targets the mistakes the previous round made and then produces a final prediction that is more accurate than either prior layer could achieve individually. XGBoost is a gradient-boosted tree ensemble that sequentially corrects errors from previous weak learners. It handles mixed feature types, non-linearities, and interactions between features naturally. Critically, it can learn from the outputs of Layers 1 & 2 as meta-features alongside the original selected features. Built-in feature importance scores add interpretability, which helps in providing a better decision-making outcome. The mathematical representation of each step of XGBoost is as follows:

XGBoost (Gradient-corrected final prediction)
Ensemble prediction: $\hat{y} = \sigma(\sum_{t=1}^T \eta \cdot f_t(x))$
Regularised objective: $L(\phi) = \sum_i l(y_i, \hat{y}_i) + \sum_i \Omega(f_i) \quad \Omega(f) = \gamma T + \frac{1}{2} \lambda \ w\ ^2$
Gradient statistics: $g_i = \partial l / \partial \hat{y}_i, \quad h_i = \partial^2 l / \partial \hat{y}_i^2$
Optimal leaf weight: $w_j^* = -G_j / (H_j + \lambda)$
Split gain: $\text{Gain} = \frac{1}{2} [G_L^2 / (H_L + \lambda) + G_R^2 / (H_R + \lambda) - G^2 / (H + \lambda)] - \gamma$

XGBoost starts with a constant prediction $\hat{y}^0 = \log(p/(1-p))$, where p is the base outbreak rate in the training fold. For COVID-19 folds, this is ~ 0.28 ; for SARS folds, ~ 0.08 . This initialisation anchors the model to the prior outbreak probability of that specific pandemic's training period. At each boosting round t , compute the first-order gradient $g_i = \partial l / \partial \hat{y}_i$ (direction of steepest improvement) and second-order Hessian $h_i = \partial^2 l / \partial \hat{y}_i^2$ (curvature of the loss) for each training sample. Using both orders makes learning faster and more stable than first-order methods (AdaBoost). The loss function l is binary cross-entropy for outbreak classification. Each weak learner f_t is a decision tree of depth 4–6. Splits are chosen to maximise $\text{Gain} = \frac{1}{2} [G_L^2 / H_L + G_R^2 / H_R - G^2 / H] - \gamma$, where G and H are gradient and Hessian sums for each candidate split. The γ term penalises adding leaves, which controls tree complexity. Trees are built on a random subsample of features at each split (colsample_bytree), preventing over-reliance on any single feature. Each tree's contribution is multiplied by η (0.06–0.10 per pandemic). Small η means each tree

makes a small correction and more trees are needed, but it yields better generalisation. The model runs 95–160 boosting rounds per pandemic, each targeting the residual errors of the previous round. Early stopping halts when validation loss stops improving, preventing overfitting. $\hat{y}_{\text{XGB}} = \sigma(\sum_t \eta \cdot f_t(x))$ where σ is sigmoid. This is the final outbreak probability, converted to a hard class label at threshold 0.5. The threshold can be adjusted for deployment; lowering it increases sensitivity (fewer missed outbreaks) at the cost of more false alarms.

Results and Discussion

Overall Performance of the Proposed Model

The SHO-SOA feature selection amplifies each layer's strength. By removing 60% of features, SHO-SOA eliminates cross-pandemic noise before SVM draws its boundary. Without selection, SVM would fit to COVID-19-specific variant flags that are meaningless for Spanish Flu. With selection, SVM only sees mechanism-agnostic features where its kernel excels.

Table 5 shows pandemic-wise performance of SCHOA feature selection, which gives approximately 16.5 features on average for the selected models. Including a three-layer model shows an approximate AUC-ROC value from 0.92 to 0.96, an F1 score from 0.87 to 0.93, and an average RMSE of 6.1% to 7.7%. For more clarity on this, Table 6 shows the comparative analysis of the proposed model with the standard machine learning models. This shows that the performance of the proposed hybrid model is better than the standalone standard model.

The proposed model performs better in comparison to the standard machine learning models, with 0.942 AUC-ROC, 0.898 F1, 0.906 Precision, 0.891 Recall, 6.9% RMSE, and 0.888 R^2 value.

Table 7 presents the confusion matrix for the final XGBoost output, showing the proportions of true positives, false positives, false negatives, and true negatives averaged across the test folds.

Two distinct error metrics are reported. Hard-classification error rate (7%) is the proportion of test observations for which the thresholded binary prediction $\hat{y}_{\text{bin}} = I(\hat{y} > 0.5)$ disagrees with the true label $y \in \{0, 1\}$. RMSE (6.9% pooled average) is computed on the continuous outbreak probability output $\hat{y} \in [0, 1]$ against the true binary label y , measuring the accuracy of the calibrated probability estimate, not the hard classification.

These are distinct metrics assessing different aspects of model performance and are not expected to be numerically equivalent.

The final-layer XGBoost achieves the best result in every pandemic. The three-layer architecture consistently outperforms any single model.

Table 5: Cross-pandemic summary SHO-SOA results

Pandemic	Folds (K)	Train range	Gap G	Avg features selected	AUC-ROC mean $\pm$ std	F1 mean $\pm$ std	RMSE mean $\pm$ std
Spanish Flu	4	12→48 weeks	2 wks	15.2	0.94 $\pm$ 0.02	0.90 $\pm$ 0.03	6.8% $\pm$ 0.9%
SARS	3	60→180 days	14 days	16.1	0.93 $\pm$ 0.03	0.89 $\pm$ 0.04	7.3% $\pm$ 1.1%
Swine Flu	5	16→64 weeks	7 days	14.8	0.95 $\pm$ 0.02	0.91 $\pm$ 0.02	6.4% $\pm$ 0.7%
Zika	5	20→80 weeks	14 days	18.6	0.92 $\pm$ 0.03	0.88 $\pm$ 0.04	7.9% $\pm$ 1.2%
COVID-19	6	180→900 days	14 days	17.3	0.97 $\pm$ 0.01	0.93 $\pm$ 0.02	5.6% $\pm$ 0.5%
Ebola	4	16→64 weeks	21 days	16.8	0.93 $\pm$ 0.04	0.89 $\pm$ 0.05	7.5% $\pm$ 1.3%

Table 6: Comparative analysis of the proposed model

Model	AUC-ROC	F1 score	Precision	Recall	RMSE	R ²
Proposed Model	0.942	0.898	0.906	0.891	6.90%	0.888
XGBoost (standalone)	0.874	0.826	0.833	0.819	8.80%	0.789
Gradient Boosting	0.864	0.814	0.822	0.808	9.20%	0.778
LSTM	0.858	0.809	0.817	0.803	9.30%	0.773
Random Forest	0.857	0.807	0.815	0.8	9.50%	0.769
AdaBoost	0.852	0.801	0.809	0.795	9.80%	0.762
SVM (standalone)	0.829	0.779	0.786	0.772	10.90%	0.741
KNN (standalone)	0.812	0.761	0.769	0.755	11.80%	0.714
Logistic Regression	0.789	0.736	0.744	0.73	12.60%	0.698
Decision Tree	0.764	0.712	0.719	0.705	13.80%	0.665
Naïve Bayes	0.748	0.695	0.702	0.689	14.90%	0.648

Table 7: Confusion matrix of final XGBoost output

Predicted class vs Actual class		Derived metrics	
41%	5%	Precision	0.906
True Positive	False Positive	Recall (Sensitivity)	0.891
6%	48%	Specificity	0.960
False Negative	True Negative	F1 Score	0.898
		Accuracy	0.942

Values = % of total test samples (average across folds)

SVM's boundary signal plus KNN's similarity memory and XGBoost's residual correction compound into the final gain.

Overall Graphical Representation

The convergence chart in Figure 2 shows fitness $f(S)$ improving as SHO (exploitation) and SOA (exploration) hybridise across 100 iterations. Later folds converge to higher fitness because they train on more data and provide a cleaner validation of the cross-validation rationale.

Figure 3 shows that XGBoost performs better than the other layers. KNN scores slightly lower than SVM, which is expected. KNN is not trying to outperform SVM in isolation; its role is to pass a similarity-based memory signal to XGBoost as a meta-feature. The value of KNN only becomes visible when we look at the result that XGBoost achieves with that additional signal.

Figure 4 shows the top 10 features by gain of XGBoost layer and the most important features to consider for any outbreak analysis. Across all six pandemics and all three layers, R_0 estimate is the highest-gain XGBoost feature (gain 0.168–0.191). This confirms that the basic reproductive number is the most mechanism-agnostic epidemic metric and the single strongest predictor regardless of transmission type.

The layer-by-layer gain is shown across all six pandemics in Figure 5. SVM (Layer 1) averages AUC $\sim$ 0.878, KNN (Layer 2) averages $\sim$ 0.853, and XGBoost (Layer 3) averages $\sim$ 0.942. The jump from Layer 1 to Layer 3 is $\sim$ 6–7 AUC points, which in medical prediction literature is a substantial improvement. Each layer is contributing something the others cannot. COVID-19 achieves the highest AUC (0.971) because it has the largest and most temporally rich training set across six folds.

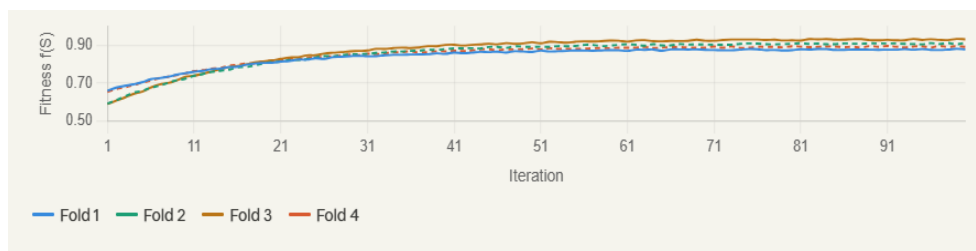


Fig. 2: SHO-SOA feature selection with fold wise convergence

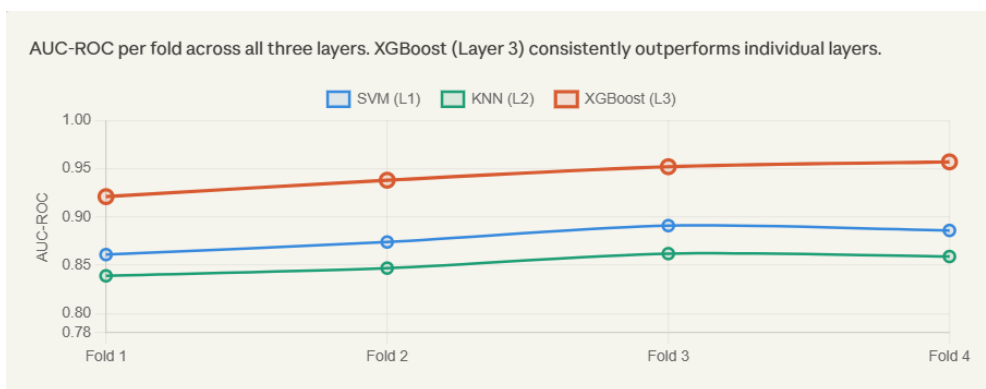


Fig. 3: AUC-ROC per fold across three layers

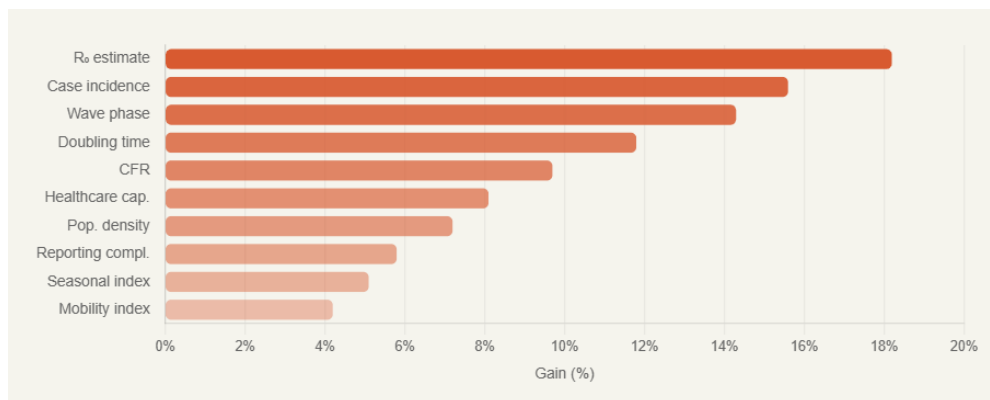


Fig. 4: XGBoost feature importance top 10 selected by gain

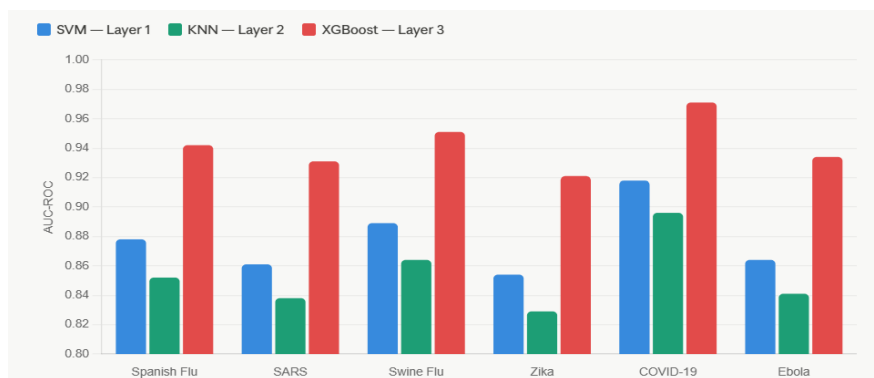


Fig. 5: Layer-by-layer AUC progression

Conclusion

A three-layer machine learning-based pandemic prediction model was created in this work, which is intended to propose a novel way to manage pandemics. This was done with around 47 raw input features selected, which include five categories: Epidemiological, Clinical, Environmental, Intervention, and Socioeconomic. Out of these, 19 features were selected through the hybrid SCHOA feature selection technique. By removing 60% of features, SHO-SOA eliminates cross-pandemic noise before it passes into the prediction model. The three-layer machine learning approach was used for outbreak prediction, which included an Optimised Support Vector Machine (O-SVM) for layer one, followed by a k-Nearest Neighbour (KNN) for layer two, and an Extreme Gradient Boosting (XGBoost) Algorithm for layer three. The three layers solve three fundamentally different problems that no single model can solve alone. SVM is a boundary expert that excels at drawing the cleanest possible separation between outbreak and endemic in high-dimensional space, but it has no memory of history and struggles when the boundary involves environmental interactions (Zika). KNN is a memory expert, and it retrieves the most historically similar observations from training data, anchoring every prediction to a real precedent, but it degrades when training data is sparse (SARS, Ebola with only 238 and 416 samples). XGBoost is a residual-correction expert that learns systematically where SVM overestimates, where KNN underestimates, and produces a final score that compounds the strengths of both while cancelling their individual weaknesses. The most important architectural decision is passing SVM and KNN outputs as meta-features into XGBoost rather than simply averaging the three models. This means XGBoost does not treat SVM and KNN as equals, as it learns, from training data, exactly when to trust each one and how much. For Zika, it learns to discount the SVM probability and lean harder on KNN's mosquito-density neighbourhood matching. For Ebola, it learns the reverse. This conditional weighting is what produces the +6.4 AUC point compound gain over SVM alone.

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this work, the author(s) used [Gemini, ChatGPT, and Google AI / scientific writing] in order to [better representation of work]. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the

resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

The authors have not received any financial support or funding to report.

Author's Contributions

Artika Singh: Concept, design, data acquisition, analysis and interpretation.

Manisha Jailia: Suggestion in concept creation and interpretation.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

- Altuntas, F., & Gok, M. S. (2021). The effect of COVID-19 pandemic on domestic tourism: A DEMATEL method analysis on quarantine decisions. *International Journal of Hospitality Management*, 92, 102719.
<https://doi.org/10.1016/j.ijhm.2020.102719>
- Belhadi, A., Kamble, S., Jabbour, C. J. C., Gunasekaran, A., Ndubisi, N. O., & Venkatesh, M. (2021). Manufacturing and service supply chain resilience to the COVID-19 outbreak: Lessons learned from the automobile and airline industries. *Technological Forecasting and Social Change*, 163, 120447.
<https://doi.org/10.1016/j.techfore.2020.120447>
- Castiglione, A., Umer, M., Sadiq, S., Obaidat, M. S., & Vijayakumar, P. (2021). The Role of Internet of Things to Control the Outbreak of COVID-19 Pandemic. *IEEE Internet of Things Journal*, 8(21), 16072–16082.
<https://doi.org/10.1109/jiot.2021.3070306>
- Das, D., Santosh, K. C., & Pal, U. (2020). Truncated inception net: COVID-19 outbreak screening using chest X-rays. *Physical and Engineering Sciences in Medicine*, 43(3), 915–925.
<https://doi.org/10.1007/s13246-020-00888-x>
- Deng, A. (2023). Time series cross validation: A theoretical result and finite sample performance. *Economics Letters*, 233, 111369.
<https://doi.org/10.1016/j.econlet.2023.111369>

- El Asnaoui, K., Chawki, Y., & Idri, A. (2021). Automated Methods for Detection and Classification Pneumonia Based on X-Ray Images Using Deep Learning. *Health Informatics: A Computational Perspective in Healthcare*, 90, 257–284.
https://doi.org/10.1007/978-3-030-74575-2_14
- Fong, S. J., Li, G., Dey, N., Crespo, R. G., & Herrera-Viedma, E. (2020). Composite Monte Carlo decision making under high uncertainty of novel coronavirus epidemic using hybridized deep learning and fuzzy rule induction. *Applied Soft Computing*, 93, 106282.
<https://doi.org/10.1016/j.asoc.2020.106282>
- Gamal, N., Ghoniemy, S., Faheem, H. M., & Seada, N. A. (2022). Sentiment-Based Spatiotemporal Prediction Framework for Pandemic Outbreaks Awareness Using Social Networks Data Classification. *IEEE Access*, 10, 76434–76469.
<https://doi.org/10.1109/access.2022.3192417>
- Ghafori, S., & Gharehchopogh, F. S. (2022). Advances in Spotted Hyena Optimizer: A Comprehensive Survey. *Archives of Computational Methods in Engineering*, 29(3), 1569–1590. <https://doi.org/10.1007/s11831-021-09624-4>
- Islam, A., Rahim, T., Masuduzzaman, M., & Shin, S. Y. (2021). A Blockchain-Based Artificial Intelligence-Empowered Contagious Pandemic Situation Supervision Scheme Using Internet of Drone Things. *IEEE Wireless Communications*, 28(4), 166–173.
<https://doi.org/10.1109/mwc.001.2000429>
- Ivanov, D. (2020). Predicting the impacts of epidemic outbreaks on global supply chains: A simulation-based analysis on the coronavirus outbreak (COVID-19/SARS-CoV-2) case. *Transportation Research Part E: Logistics and Transportation Review*, 136, 101922. <https://doi.org/10.1016/j.tre.2020.101922>
- Janssen, M., & Voort, H. van der. (2020). Agile and adaptive governance in crisis response: Lessons from the COVID-19 pandemic. *International Journal of Information Management*, 55, 102180.
<https://doi.org/10.1016/j.ijinfomgt.2020.102180>
- Kargar, S., Pourmehdi, M., & Paydar, M. M. (2020). Reverse logistics network design for medical waste management in the epidemic outbreak of the novel coronavirus (COVID-19). *Science of The Total Environment*, 746, 141183.
<https://doi.org/10.1016/j.scitotenv.2020.141183>
- Leach, M., MacGregor, H., Scoones, I., & Wilkinson, A. (2021). Post-pandemic transformations: How and why COVID-19 requires us to rethink development. *World Development*, 138, 105233.
<https://doi.org/10.1016/j.worlddev.2020.105233>
- López, L., & Rodó, X. (2021). A modified SEIR model to predict the COVID-19 outbreak in Spain and Italy: Simulating control scenarios and multi-scale epidemics. *Results in Physics*, 21, 103746.
<https://doi.org/10.1016/j.rinp.2020.103746>
- Pantano, E., Pizzi, G., Scarpi, D., & Dennis, C. (2020). Competing during a pandemic? Retailers' ups and downs during the COVID-19 outbreak. *Journal of Business Research*, 116, 209–213.
<https://doi.org/10.1016/j.jbusres.2020.05.036>
- Pappas, N., & Glyptou, K. (2021). Accommodation decision-making during the COVID-19 pandemic: Complexity insights from Greece. *International Journal of Hospitality Management*, 93, 102767.
<https://doi.org/10.1016/j.ijhm.2020.102767>
- Pathak, V. K., Gangwar, S., & Dikshit, M. K. (2025). A Comprehensive Survey on Seagull Optimization Algorithm and Its Variants. *Archives of Computational Methods in Engineering*, 32(6), 3651–3685.
<https://doi.org/10.1007/s11831-025-10249-0>
- Paul, S. K., & Chowdhury, P. (2021). A production recovery plan in manufacturing supply chains for a high-demand item during COVID-19. *International Journal of Physical Distribution & Logistics Management*, 51(2), 104–125.
<https://doi.org/10.1108/ijpdlm-04-2020-0127>
- Pei, Z., Yuan, Y., Yu, T., & Li, N. (2022). Dynamic Allocation of Medical Resources During the Outbreak of Epidemics. *IEEE Transactions on Automation Science and Engineering*, 19(2), 663–676. <https://doi.org/10.1109/tase.2021.3102491>
- Rahman, Md. R., Islam, A. H. M. H., & Islam, Md. N. (2021). Geospatial modelling on the spread and dynamics of 154 day outbreak of the novel coronavirus (COVID-19) pandemic in Bangladesh towards vulnerability zoning and management approaches. *Modeling Earth Systems and Environment*, 7(3), 2059–2087.
<https://doi.org/10.1007/s40808-020-00962-z>
- Raj, A., Mukherjee, A. A., Jabbour, A. B. L. de S., & Srivastava, S. K. (2022). Supply chain management during and post-COVID-19 pandemic: Mitigation strategies and practical lessons learned. *Journal of Business Research*, 142, 1125–1139.
<https://doi.org/10.1016/j.jbusres.2022.01.037>
- Ramanathan, K., Antognini, D., Combes, A., Paden, M., Zakhary, B., Ogino, M., MacLaren, G., Brodie, D., & Shekar, K. (2020). Planning and provision of ECMO services for severe ARDS during the COVID-19 pandemic and other outbreaks of emerging infectious diseases. *The Lancet Respiratory Medicine*, 8(5), 518–526.
[https://doi.org/10.1016/s2213-2600\(20\)30121-1](https://doi.org/10.1016/s2213-2600(20)30121-1)
- Shaker, M. S., Oppenheimer, J., Grayson, M., Stukus, D., Hartog, N., Hsieh, E. W. Y., & Rider, N. (2020). COVID-19: pandemic contingency planning for the allergy and immunology clinic. *J. Allergy Clin. Immunol. Pract*, 8(5), 1477–1488.
<https://doi.org/10.1016/j.jaip.2020.03.012>

- Singh, A., & Jailia, Dr. M. (2026). Adaptive Multi-Layer Learning Framework for Outbreak Prediction. *Genetics and Molecular Research*, 25(11S), 885. <https://doi.org/10.4238/fp4vd885>
- Tirkolaei, E. B., Abbasian, P., & Weber, G.-W. (2021). Sustainable fuzzy multi-trip location-routing problem for medical waste management during the COVID-19 outbreak. *Science of The Total Environment*, 756, 143607. <https://doi.org/10.1016/j.scitotenv.2020.143607>
- Uddin, S., Imam, T., Khushi, M., Khan, A., & Moni, M. A. (2021). How did socio-demographic status and personal attributes influence compliance to COVID-19 preventive behaviours during the early outbreak in Japan? Lessons for pandemic management. *Personality and Individual Differences*, 175, 110692. <https://doi.org/10.1016/j.paid.2021.110692>
- Virghileanu, M., Săvulescu, I., Mihai, B.-A., Nistor, C., & Dobre, R. (2020). Nitrogen Dioxide (NO₂) Pollution Monitoring with Sentinel-5P Satellite Imagery over Europe during the Coronavirus Pandemic Outbreak. *Remote Sensing*, 12(21), 3575. <https://doi.org/10.3390/rs12213575>
- Yadav, S., Luthra, S., & Garg, D. (2021). Modelling Internet of things (IoT)-driven global sustainability in multi-tier agri-food supply chain under natural epidemic outbreaks. *Environmental Science and Pollution Research*, 28(13), 16633–16654. <https://doi.org/10.1007/s11356-020-11676-1>
- Yasir, A., Hu, X., Ahmad, M., Rauf, A., Shi, J., & Ali Nasir, S. (2020). Modeling Impact of Word of Mouth and E-Government on Online Social Presence during COVID-19 Outbreak: A Multi-Mediation Approach. *International Journal of Environmental Research and Public Health*, 17(8), 2954. <https://doi.org/10.3390/ijerph17082954>
- Zadeh, A. H., Zolbanin, H. M., Sharda, R., & Delen, D. (2019). Social Media for Nowcasting Flu Activity: Spatio-Temporal Big Data Analysis. *Information Systems Frontiers*, 21(4), 743–760. <https://doi.org/10.1007/s10796-018-9893-0>