Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

DenVis-S: A Sustainable Visual Analytics Framework for Intelligent Disease Surveillance, Risk Assessment, and Public Health Resource Planning

Farrukh Arslan

Farrukh_arslan@uet.edu.pk

University of Engineering and Technology Lahore

Aqsa Shabbir

Lahore College For Women University

Saba Ramzan

University of Lahore

Research Article

Keywords: disease surveillance, deep learning, classical forecasting, ARIMA, ETS, seasonal-naive, linear autoregression, LSTM, GRU, temporal convolutional network, visual analytics, spatio-temporal analysis, public health resource planning, sustainable healthcare, emergency surveillance

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11114287/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

DenVis-S: A Sustainable Visual Analytics Framework for Intelligent Disease Surveillance, Risk Assessment, and Public Health Resource Planning

Farrukh Arslan[1], Aqsa Shabbir[2], Saba Ramzan[3]

[1] University of Engineering and Technology, Lahore, Pakistan

[2]Lahore College For Women University, Lahore, Pakistan

[3]University of Lahore, Lahore, Pakistan

*Correspondence: Farrukh Arslan, Farrukh_arslan@uet.edu.pk

Abstract: Defining disease surveillance in a resource-poor health system will be a crucial aspect that will ensure accurate identification of abnormal activity, spatial prioritisation and future-looking intelligence to enable efficient use of scarce public health resources. In this paper, we present DenVis-S, a sustainability-based visual analytics framework that combines data preprocessing, transparent syndromic case identification, spatio-temporal visual analytics and short-term prediction. It turns heterogeneous emergency surveillance data into a reliable surveillance signal and tests LSTM, GRU and TCN alongside persistence, seasonal-naive, ARIMA, ETS and linear-autoregressive predictions of one day ahead. We found that GRU performed poorly on the test days of August (204.60 presentations) and sMAPE (37.39%), whereas linear autoregression had the lowest RMSE (335.60) and ARIMA had the best MAE (207.11) and RMSE (336.17). Our study confirms that deep learning offers competitive short-term surveillance intelligence, but not always better than the classic forecasting methods on the short 4 months series. The framework is designed to facilitate risk-based interpretation rather than diagnostic verification, so observed and forecast signals can be used to inform human review, preparedness and targeted public-health plans. The sustainability contribution is framed as better targeting, coordination, and prioritization of surveillance and preparedness activities instead of a direct reduction in resource consumption. DenVis-S therefore provides a modular path to connect smart surveillance, comparative prediction, visual analytics and resource-informed public health decision support.

Keywords: disease surveillance; deep learning; classical forecasting; ARIMA; ETS; seasonal-naive; linear autoregression; LSTM; GRU; temporal convolutional network; visual analytics; spatio-temporal analysis; public health resource planning; sustainable healthcare; emergency surveillance

1. Introduction

Effective disease surveillance requires timely detection of abnormal activity, spatial prioritization and forward-looking intelligence to inform public health decisions. Traditional surveillance models may be hindered by long reporting times, fragmented data, heterogeneous data formats and limited analytic capabilities. These challenges are especially important when health systems are confronted with changing disease patterns and are working with limited personnel, infrastructure and intervention resources [1–3]. In dengue-endemic settings, these constraints are especially relevant as dengue surveillance can be based on very different time and location scales.

From a sustainability perspective, intelligent surveillance needs to be about more than detecting disease-related signals. Public-health systems need to have surveillance personnel, clinical capacity, laboratory services, communication channels, field teams, transportation, and intervention budgets in place. An approach that identifies new signals earlier and allows for clear prioritization can increase targeting and coordination of these resources and help us prioritize them. Digital platforms can be a key element of sustainable healthcare if they encourage information sharing, coordination, and evidence-based interaction between healthcare providers [4], but sustainability claims need to be tied to real system results and digitalization alone is not a reliable way to show that we can win money or have an economic and environmental return.

Deep learning is a flexible computational framework for extracting temporal dependencies from surveillance data. Recurrent architectures such as LSTM and GRU can model sequential dependencies while temporal convolutional architectures can capture patterns over multiple receptive-field scales. Recent disease prediction research has made use of deep learning, spatial, environmental and other contextual information to gain an understanding of the risk in a more precise manner [5–7]. We aim to address an operational problem: how heterogeneous emergency-surveillance data can be turned into a transparent, reproducible surveillance signal and combined with short-horizon deep learning to help achieve risk planning. We do not intend our framework to replace laboratory validation or expert evaluation. The purpose of such a study is to find temporal and spatial signals that require a better human evaluation.

DenVis-S is built around a simple chain of evidence: surveillance records are quality-checked and standardized; a transparent condition-compatible presentation rule is applied; temporal and locality-level patterns are visualized; deep learning models predict the volume of surveillance in the short term; and information from these approaches is used to make decisions about surveillance and resource prioritization. This is in line with early warning algorithms that organize surveillance data, identify alarm signals and enable timely response and investigation [8]. Interactive WebGIS surveillance can also support this early warning chain by providing spatially organized information for response and investigation [9,10]. The LSTM, GRU, and temporal-convolutional architectures of DenVis-S are based on sequence modelling frameworks [11–13].

Our Contributions

The key contribution of this study is to develop DenVis-S, a sustainability-oriented visual analytics model based on syndromic surveillance, spatio-temporal analysis, comparative forecasting and human-centred public health decision support in a single workflow. Unlike disease prediction or spatial visualization based methods that only focus on disease prediction and only on data visualization,

DenVis-S bridges the gap between heterogeneous emergency surveillance data and a transparent surveillance proxy and time-series prediction and locality-level workload interpretation. The framework is designed to be a decision-support system (as opposed to an outbreak prediction or outbreak detection system).

More specifically, the contributions of this study are fourfold:

1. Transparent surveillance-target construction: DenVis-S is a reproducible condition-compatible emergency-presentation (DCEP) rule that makes heterogeneous emergency data an interpretable syndromic surveillance signal. As in the present works, the target is explicitly considered as a surveillance proxy instead of an actual dengue incidence, in order to avoid the diagnostic limitations of presentation-based surveillance.
2. Integrated spatio-temporal visual analytics: The framework uses temporal surveillance patterns and locality-level concentration analysis for the context-based interpretation of surveillance workload change. Since validated population denominators and geographic coordinates were not available, the spatial component is framed as concentration of surveillance workload and not as population-adjusted disease risk.
3. Controlled comparative forecasting: DenVis-S compares LSTM, GRU, and TCN architectures with persistence, seasonal-naive, ARIMA, ETS and linear-autoregressive baselines in a similar chronological train-validation-test framework. This comparison avoids assuming that deep learning is universally superior and allows nonlinear sequence models to be compared against existing statistical forecasts.
4. Sustainability-oriented decision support: Instead of equating predictive accuracy to sustainability, DenVis-S takes surveillance data and forecasts and converts them into a human-in-the-loop planning approach where surveillance intensity, investigation, communication and preparedness resources may be prioritized according to the latest time and space signals. The research thus frames sustainability

around better targeting, coordination and prioritization and does so without establishing financial, environmental or personnel savings.

So DenVis-S is system-level integration rather than a new deep-learning architecture. The framework offers a modular pathway from heterogeneous emergency surveillance data to interpretable predictive intelligence and resource-informed public health planning, while also explicitly restricting the claims that can be made from a short, single-site observational dataset. This is important because our evaluation covers one hospital, one city and four months, so the empirical findings should be viewed as a case study evaluation rather than broad operational generalizability.

2. Literature Review

Timely surveillance is essential for early warning, because actionable signals need to be detected, interpreted and communicated before public-health responses can be optimally targeted [8]. Structured early-warning systems can organise diverse observations into indicators that can be used to assess risks and respond. Interactive WebGIS and geographic surveillance can help to build additional links in the early-warning chain, by arranging surveillance information in the space to be explored and reacted to [9,10]. The temporal modelling of DenVis-S is based on the same recurrent and temporal-convolutional sequence-model formulation [11-13]. Early reports of dengue prevalence and surveillance progress are now being reported from WHO as well [14]. Digital health implementation may face organizational resistance and resource allocation challenges and therefore there is an imperative for evaluation beyond predictive performance [15]. The epidemiological literature further demonstrates the scale and regional variety of dengue transmission. Global burden estimates, vector distribution studies and climate-sensitive transmission research can provide context for timing-sensitive surveillance and forecasting [16–23].

Emergency and syndromic surveillance can also complement laboratory or notifiable disease detection as a signal of changing symptom burden. In other words, these signals cannot be as specific as laboratory-confirmed diagnosis and should be taken as an initial check-in, not a definitive diagnosis of infection. Systematic reviews and recommendations to surveillance are based on the use of early emergency and clinical presentation data as a complementary signal before laboratory diagnosis, and the need for clear and systematic construction and detection of the syndrome [24-32]. Spatial scan statistics have been used to identify geographic concentration, but DenVis-S uses locality frequency data instead of formal scan statistics because the population denominators and validated coordinates are not available [32]. With respect to dengue prediction, the literature has shown considerable variance in predictors, time aggregation, validation, and evaluation. Climate-informed and hybrid deep learning studies indicate the potential of recurrent, convolutional, and ensemble predictions [33-36]. Studies from Vietnam, China, and other countries illustrate that climate variables, hybrid architectures, and ensemble neural models are used in climate models and zero-inflated modelling have been shown to affect dengue prediction performance [37-41].

Emergency hospital data can complement laboratory or notifiable-disease surveillance because emergency departments can observe changes in symptom burden before formal confirmation becomes available. However, symptom-based surveillance has lower diagnostic specificity. In the present study, we consider emergency presentations as a syndromic signal and not as a substitute for laboratory confirmed surveillance. The clinical and epidemiological literature also supports the view that dengue surveillance is a layered evidence problem where clinical presentation, vector ecology, environmental conditions and diagnostic confirmation provide different levels of information [42–44].

Geospatial analytics is widely used to assess spatial concentration, understand heterogeneity, and communicate surveillance patterns with decision makers. Spatial decision-support and WebGIS approaches integrate geographic information with surveillance data to support targeted investigation and response [9,10]. Visual analytics can complement the predictive modelling by enabling users to visualize observations, trends, uncertainty, and geographic concentration in combination. Visual analytics is considered as an interpretation layer in DenVis-S: the model is able to develop a prediction, while the temporal and locality-level data provide context to determine if a predicted increase is isolated, persistent, or concentrated in a geographically common source location. This view is consistent with GeoAI and visual analytics research showing how spatially explicit AI and interactive dashboards, and human-centered visual reasoning can contribute to public health intelligence and decision making [45-53].

Deep learning has become increasingly popular in temporal and spatio-temporal disease prediction. Reviews have shown that LSTM, CNN, attention-based recurrent networks, transfer learning and other related models can effectively predict diseases in infectious disease settings [7]. For example, studies on LSTM-based and geospatially enriched models show the benefit of learning temporal dependencies while incorporating contextual predictors [5,6]. The core recurrent and temporal-convolutional sequence models are the basis for the LSTM, GRU and TCN models used in our study [11–13]. DenVis-S is built on these same temporal architectures and not on a universal basis of these. The framework is based on controlled comparison, short-horizon forecasting, transparent target construction and integrating the predictions with visual interpretation.

DenVis-S uses LSTM, GRU and TCN as complementary temporal architectures but not universal model superiority. A short-horizon forecasting design is appropriate when the temporal coverage is limited, and longer observation periods are needed to evaluate annual seasonality and broader generalization. Sustainable healthcare needs to account for social, economic, organizational and environmental consequences rather than just predictability. While digital platforms can assist sustainability by facilitating information exchange and coordination, sustainability should be assessed at the health-system level and not assumed from the technology alone [4,10]. Recent sustainability research also emphasizes usability, scalability, resource implications, equity and longer-term system effects [10].

Therefore, this paper defines the sustainability relevance of DenVis-S in a narrow sense: if short-term predictive surveillance can identify where and when emergency demand is increasing, public health managers may be better prepared to prioritize surveillance intensity, vector-control investigation, communication activities and preparedness resources. Therefore, the study does not assess actual cost, energy, carbon or personnel savings from those decisions.

The existing work on disease forecasting uses deep learning for temporal and spatial pattern modelling, but GIS-based surveillance platforms provide complementary spatial intelligence [5-7,9]. The remaining opportunity is to marry transparent surveillance-target construction, short-horizon deep learning, visual analytics and sustainability-driven decision support in a single tool-based workflow. DenVis-S's main gap is system-level integration: different types of deep temporal models are used to identify a surveillance proxy and confirm the diagnosis and forecasting from a different deep temporal architecture and then to turn predictions into a cautious resource-planning process. At the other end of the spectrum, GIS-based surveillance platforms can provide useful spatial intelligence but may not provide a deep learning forecasting layer [9]. DenVis-S combines all these approaches and transforms routine hospital emergency information into a syndromic surveillance signal, coupled with short horizon deep learning, and presents the results from a sustainability perspective. The main methodological gap is not the development of yet another neural architecture but the construction of a reproducible pipeline that explicitly distinguishes a syndromic proxy from confirmed dengue, evaluates several deep temporal architectures under the same data constraints, and translates forecasts into cautious resource planning. Finally, the sustainability approach of DenVis-S is consistent with the recent digital-health research which treats sustainability as a system property over many levels: efficiency, accessibility, organizational integration, environmental impact, long-term health-system benefits, etc. and not as a result of AI. Also, recent reviews highlight that the environmental advantage of digital health is not a result of the use of AI but of measurement rather than assumption. As such, DenVis-S considers its sustainability contribution as a potential improvement in targeting, coordination, and prioritization of surveillance and preparedness activities, but leaves the financial, energy, carbon, or personnel savings to future evaluations.

From an analytical point of view, the use of multiple forecasting architectures should incorporate classical and naive baselines as well as validation procedures and transparent error reporting. Forecasting approaches focus on benchmark models, time series evaluation, and careful separation of training and test data [54,55]. We therefore compare persistence, seasonal naive, ARIMA, ETS, and linear autoregression to deep learning models based on chronological train-validation-test separation. Adam-based optimization, decoupled weight decay, and loss mechanisms for neural models are well established [56,57], while sophisticated loss algorithms such as the Huber loss can help to reduce the impact of exceptionally large residuals [58]. Geospatial intelligence and visual analytics are complementary tools for translating unsynthetic data from surveillance into data that can be interpreted in spatial and temporal space and time. GeoAI study has shown the growing use of spatially explicit artificial intelligence in health and geography knowledge discovery [45–47]. Interactive dashboards and web-based geographic interfaces have also been useful in conveying rapidly evolving public-health data to decision makers [48,49]. Visual analytics research has emphasized the integration of computational analysis and human reasoning so that users can observe trends, patterns, anomalies, and context instead of using only the model results [50–53]. The epidemiological and clinical literature also support treating dengue surveillance as a layered evidence problem where population burden, clinical presentation, environmental conditions, vector ecology, and diagnostic proof are different [42–44]. DenVis-S intends to position its DCEP target as a syndromic surveillance proxy and its outputs as risk-oriented intelligence, not disease-related disease events.

Recent studies of dengue prediction consistently show significant heterogeneity in predictor variables, temporal aggregation, validation methods, and evaluation criteria [33–36]. This heterogeneity is important for model comparison, since differences in the resolution of the data, forecasting horizon, external covariates, and validation procedures are the key factors that influence the performance of the models, rather than the architecture of the model itself. Climate-informed deep learning has emerged as a promising model for multi-location prediction in the context of Vietnam and China, and hybrid recurrent-convolutional and ensemble neural approaches have built longer forecasting horizons and model flexibility [37–40]. More recently, zero-inflated data and spatio-temporal resolution are also the important variables for dengue prediction performance [41].

The literature on syndromic and emergency department surveillance can support routine clinical presentation data as an early, complementary signal before laboratory confirmation is available [24-30]. Recent clinical studies have shown that emergency department syndromic systems provide timely public health intelligence, and that cluster detection studies support the need for clearly defined temporal, spatial, and spatio-temporal detection methods [25,26]. Traditional surveillance literature also supports the belief that syndromic signals are useful and not a threat to definitive diagnosis [27–29]. Analysis performance can also be influenced by the choice of aberration detection approach and syndrome construction [30,31]. Local scan statistics can be used to identify geographic concentration, but DenVis-S was not able to use them for the numerical scan statistics as the population denominators and validated coordinates could not be obtained [32].

The larger epidemiological literature has established the extent and geographic heterogeneity of dengue transmission and provides the context for surveillance-oriented modelling. Global burden estimates and spatial-risk studies show a great deal of geographical variation in dengue burden and vector suitability, as well as climate-sensitive transmission and changing vector distributions, which emphasize the need for timely surveillance and prediction [16-23].

The comparative evidence summarized in Table 1 clarifies how DenVis-S differs from recent forecasting, geospatial, and WebGIS approaches.

Recent dengue surveillance and forecasting studies demonstrate substantial progress in geospatial risk mapping, deep-learning prediction, and WebGIS-based decision support. However, these contributions tend to emphasize one or two analytical dimensions rather than integrating transparent syndromic surveillance, short-horizon deep learning, visual analytics, and sustainability-oriented resource prioritization within a single workflow. Table 1 summarizes the principal methodological differences and clarifies the research gap addressed by DenVis-S.

Table 1. Comparison of DenVis-S with recent dengue surveillance, geospatial analytics, and deep-learning forecasting studies, highlighting the system-level research gap addressed by the proposed framework.

Study	Data / Context	Analytical approach	Primary contribution	Key limitation / boundary	Gap addressed by DenVis-S
Li et al. (2022) [5]	Weekly dengue cases; Brazil; 2013–2020; environmental and geospatial predictors	Google Earth Engine predictors with RF, LSTM and LSTM-attention; 1–4 week forecasting	Combines big geospatial/environmental predictors with deep learning for multi-week dengue forecasting	Requires richer environmental/geospatial predictors and focuses primarily on forecasting rather than emergency-surveillance workflow integration	DenVis-S emphasizes a lighter surveillance-input pathway, transparent syndromic target construction, visual interpretation, and decision-support integration.
Majeed et al. (2023) [6]	Dengue prediction in Malaysia with spatial information	LSTM with spatial attention	Demonstrates that spatially informed deep learning can improve dengue prediction	Focuses on predictive modelling; does not provide an integrated visual analytics and resource-prioritization workflow	DenVis-S links temporal DL forecasts with locality-level surveillance concentration and human-in-the-loop prioritization.
Li and Dong (2022) [7]	Urban dengue forecasting literature using big geospatial data	Review of geospatial and data-driven dengue risk forecasting	Synthesizes GIS, environmental, remote-sensing and data-driven approaches	Highlights the value of heterogeneous contextual predictors but does not itself provide an operational integrated surveillance artifact	DenVis-S operationalizes a modular surveillance-to-decision pipeline and explicitly separates surveillance workload from population-adjusted disease risk.
Javaid et al. (2023) [9]	Vector-borne disease surveillance; real-time WebGIS context	WebGIS-based real-time surveillance and response	Demonstrates the value of interactive geographic surveillance for vector-borne disease response	Strong spatial and operational orientation, but the framework does not focus on the specific LSTM/GRU/TCN	DenVis-S adds controlled deep-learning comparison, short-horizon workload forecasting, and sustainability-

				comparative forecasting design used here	oriented resource prioritization to visual surveillance.
DenVis-S (this study)	Emergency-surveillance presentations transformed into a transparent syndromic signal; short-horizon daily forecasting	DCEP construction + locality concentration + LSTM/GRU/TCN + persistence + seasonal-naive + ARIMA + ETS + linear AR + visual analytics + human-in-the-loop decision support	Integrates surveillance-target construction, temporal DL forecasting, spatial workload interpretation, and sustainability-oriented planning in one reproducible workflow	Short temporal coverage, no laboratory-confirmed labels, coarse address proxies, and negative R^2 limit generalizability and prohibit claims of autonomous outbreak detection	Addresses the system-level integration gap while explicitly bounding interpretation and separating workload concentration from population-adjusted incidence.

Research gap. The comparison indicates that recent studies have established the value of deep learning for dengue forecasting [5,6], geospatial and environmental data for risk modelling [5,7], and interactive WebGIS for surveillance and response [9]. The remaining gap is primarily architectural and operational: comparatively fewer approaches connect a transparent surveillance proxy, controlled short-horizon deep-learning comparison, spatial workload interpretation, visual analytics, and sustainability-oriented human decision support in a single reproducible workflow. DenVis-S addresses this gap as a system-level design contribution rather than claiming a new neural architecture. The framework also makes explicit that syndromic surveillance signals are not equivalent to confirmed disease and that resource-prioritization outputs require human validation.

3. Materials and Methods

DenVis-S follows a design-and-evaluation workflow in which data quality, target construction, modelling, visualization, and sustainability interpretation are treated as connected stages. The framework begins with hospital emergency records and ends with a planning-oriented visual representation of observed and predicted DCEP activity. Its modular structure allows laboratory-confirmed diagnoses, climate variables, vector indices, population denominators, mobility indicators,

or additional healthcare facilities to be incorporated in future versions without changing the core forecasting logic.

The overall processing pathway is illustrated in Figure 1.

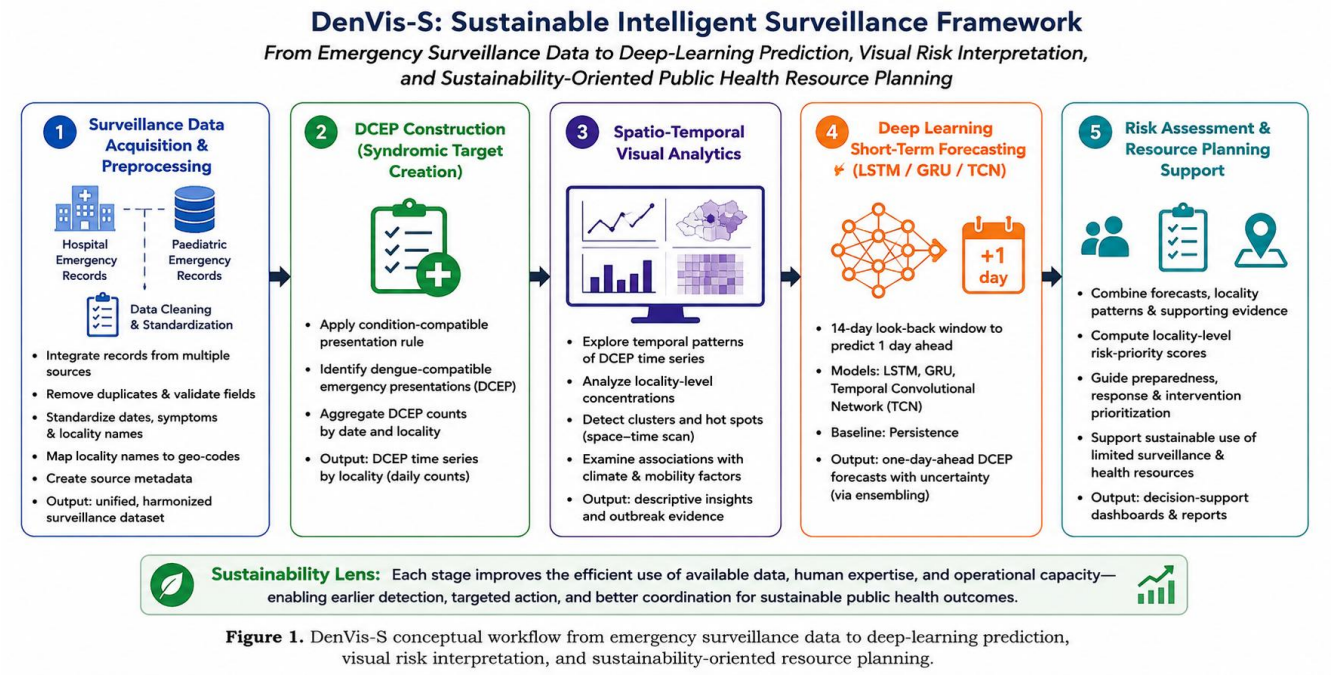


Figure 1. DenVis-S conceptual workflow from emergency surveillance data to deep-learning prediction, visual risk interpretation, and sustainability-oriented resource planning.

Figure 1. DenVis-S conceptual workflow from emergency surveillance data to deep-learning prediction, visual risk interpretation, and sustainability-oriented resource planning.

Data representation and provenance are explicitly part of the framework. Input surveillance records could contain temporal, demographic, symptom or complaint, and locality information. The framework retains source provenance while harmonizing field names, formats, and representations before target construction and time aggregation. Record counts, collection periods, and source institution are only reported in the experimental design and results sections.

Our framework can handle all emergency surveillance records of heterogeneous data that can be stored in various files that are not organized in a single file. Data fields can include dates, demographics, locality information, and one or more complaint or symptom fields. The records are combined to form a common analysis before target construction and temporal aggregation. The data source, observation

period, number of records, and field composition are reported in our experimental section. The data composition and number of records are summarized in Table 2.

Table 2. Composition of the emergency-surveillance dataset used for DCEP target construction.

Dataset component	Records	Key information
Main emergency	133,398	Date; age; sex; address; three complaint fields
Paediatric emergency	18,564	Day/date; age; sex; address; complaint
Total	151,962	Combined emergency surveillance records
Explicit 'dengue' diagnosis text	0	No record contained an explicit dengue label
DCEP records	62,537	Fever/pyrexia + ≥1 predefined associated symptom

Data quality assessment includes structural header detection, validation of temporal fields, duplicate pattern identification, missingness assessment, and preservation of source provenance. Structural anomalies are reviewed before modelling and observation gaps are distinguished from genuine zero activity wherever possible. All of these steps are repeated in the experimental train, validation, and test partitions.

Preprocessing consists of structural validation, field harmonization, date normalization, complaint-text or symptom-field normalization, whitespace normalization, and preservation of source provenance. The procedure is designed to minimize avoidable variation in heterogeneous surveillance inputs before applying the condition-compatible presentation rule.

A transparent syndromic target is defined from the presentation characteristics. Let F_i indicate if record i has a fever-related or other target condition compatible symptom and A_i indicate whether record i contains at least one associated symptom or presentation feature. Then the binary target is created using the rule shown in Equation (1). As source records have no laboratory-confirmed dengue

indicator, we use a transparent syndromic case definition. A record was identified as DCEP if its complaint text contained a fever-related term and at least one associated symptom. The fever group consisted of 'fever', 'pyrexia', and high-grade fever variants. The associated group included arthralgia, myoarthralgia, myalgia, body aches/body pain, headache, chills, vomiting, nausea, bleeding, joint pain, or muscle pain.

The target definition is conservative: a condition-compatible emergency presentation is a surveillance proxy and should not be interpreted as a laboratory-confirmed diagnosis. The details of what symptoms and what is performed based on data and the study objective are derived from an experimental implementation. $DCEP_{\mathrm{i}} = \{ 1, if\ F_{\mathrm{i}} = 1\ and\ A_{\mathrm{i}} = 1;\ 0, otherwise \}$ (1)

Where F_i indicates the presence of a target-condition-compatible signal, and A_i indicates the presence of at least one predefined associated symptom or presentation feature in record i. The operational target is used as a surveillance proxy and should not be interpreted as confirmed disease diagnosis.

Temporal aggregation converts record-level observations into a daily or otherwise operational time-series signal. For each time step t, the primary target Y_t denotes the number of condition-compatible presentations. Additional descriptive features may include total emergency volume, demographic summaries, locality frequencies, and daily symptom or presentation-feature counts.

Daily aggregation was used because the source records provide daily emergency activity and the study objective is short-horizon forecasting. For each day t, the primary target Y_t is the number of DCEP records. Additional descriptive features included total emergency presentations, unique address strings, male/female counts, and daily symptom counts for fever, arthralgia, body aches/pain, headache, chills, vomiting, nausea, and bleeding.

Locality information is used as a spatial proxy rather than as a claim of precise population-adjusted geographic risk. Standardized locality strings, administrative areas, or validated coordinates can be used to identify recurring source concentrations. When population denominators or catchment

information are unavailable, locality counts should be interpreted as surveillance-workload concentrations rather than incidence rates.$Y_t = \Sigma_{i=1}^{N} I(DCEP_i = 1 \; and \; date_i = t)$ (2)

Address information was used as a spatial proxy rather than a fully geocoded coordinate system. Standardized address-string frequencies were used to identify recurring source localities. This approach allows spatially oriented visual analysis while avoiding an unsupported claim of precise geographic risk. Population denominators were not available; therefore, locality counts are interpreted as surveillance workload concentrations rather than incidence rates.

$$C_{l,t} = \Sigma_{i=1}^{N} I(DCEP_i = 1, locality_i = l, date_i = t) \quad (3)$$

$$\tilde{x}_{t,j} = (x_{t,j} - \mu_j^{train}) / \sigma_j^{train} \quad (4)$$

The forecasting problem is defined as one-day-ahead prediction of DCEP volume using a 14-day look-back window for the deep-learning models. The input sequence contains the previous 14 days of standardized epidemiological features. Three deep-learning architectures were evaluated: LSTM, GRU, and TCN. To establish stronger and more interpretable reference points, four classical forecasting baselines were added: seasonal-naive forecasting with a seven-day seasonal lag, autoregressive integrated moving average (ARIMA), exponential smoothing (ETS), and linear autoregression. Persistence forecasting was retained as the simplest naive benchmark, defined as using the most recent observed DCEP count as the next-day prediction.

$$X_t = [x_{t-13}, \ldots, x_t], \; L = 14 \quad (5)$$

$$\hat{Y}_{t+1} = f\theta(X_t) \quad (6)$$

3.5.1 LSTM

LSTM was selected because its gated memory mechanism is designed to retain relevant temporal information while reducing the vanishing-gradient problem associated with conventional recurrent networks. The implemented model used two recurrent layers with 64 hidden units, followed by a 32-unit dense layer and a single linear output.

$$i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i) \quad (7)$$

$$f_t = \sigma(W f x_t + U f h_{t-1} + b f) \quad (8)$$

$$o_t = \sigma(W o x_t + U o h_{t-1} + b o) \quad (9)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot tanh(W c x_t + U c h_{t-1} + b c) \quad (10)$$

$$h_t = o_t \odot tanh(c_t) \quad (11)$$

3.5.2 GRU

GRU was evaluated as a lower-complexity recurrent alternative. The model used two GRU layers with 64 hidden units, followed by the same 32-unit dense output head used for LSTM. Using matched output capacity provides a more controlled comparison between recurrent architectures.

$$z_t = \sigma(W z x_t + U z h_{t-1} + b z) \quad (12)$$

$$r_t = \sigma(W r x_t + U r h_{t-1} + b r) \quad (13)$$

$$\tilde{h}_t = tanh(W h x_t + U h(r_t \odot h_{t-1}) + b h) \quad (14)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (15)$$

3.5.3 Temporal Convolutional Network

TCN was included as a non-recurrent deep temporal architecture. The implementation used causal dilated one-dimensional convolutions with dilation factors of 1, 2, and 4, followed by a dense

regression head. TCN provides a useful comparison because it learns temporal receptive fields through convolution rather than recurrent state transitions.

$$y_t = \Sigma_{k=0}^{K-1} w_k x_{t-}d \cdot k, \quad d \in \{1, 2, 4\} \tag{16}$$

$$H^l = ReLU(W^l * d\, H^{l-1} + b^l) + H^{l-1} \tag{17}$$

3.5.4 Classical Forecasting Baselines

The classical baselines were included to determine whether the observed forecasting gains could be obtained with conventional time-series methods rather than deep sequence architectures. Seasonal-naive forecasting uses the value from the same weekday in the previous week. ARIMA models the differenced series through autoregressive and moving-average terms. ETS uses exponential smoothing of the level without imposing an explicit seasonal component because the available observation period is only four months. Linear autoregression predicts the current value from a fixed number of recent observations. Model configurations were selected on the July validation period and then evaluated on the August test period using one-step-ahead rolling forecasts. The selected configurations were ARIMA(2,0,1), level-only ETS, seasonal-naive with period 7, and linear autoregression with three lags. This design prevents the deep-learning models from being compared only against an unusually weak persistence reference and follows established forecasting practice [54,55].

$$\varphi(B)(1 - B)\hat{}d\, y_t = c + \theta(B)\varepsilon_t$$

$$\hat{y}_{t+1} = y_{t+1-7}$$

$$\ell_t = \alpha y_t + (1 - \alpha)\ell_{t-1}, \quad \hat{y}_{t+h}|_t = \ell_t$$

$$\hat{y}_t = c + \Sigma_{j=1}^{p} \varphi_j y_{t-j} + \varepsilon_t$$

3.5.5 Training and Validation

The chronological partitioning strategy separates model fitting, validation, and final testing in time to avoid information leakage. Features and the target are standardized using statistics derived only from

the training period. The look-back window is fixed for the comparative experiment, while optimization uses AdamW, Smooth L1 loss, gradient clipping, mini-batches, and early stopping based on validation performance. Dataset-specific dates, partition sizes, and source-data gaps are reported in the experimental-design section.

$$L\delta(e) = \{ ½e^2, if\ |e| \leq \delta;\ \delta(|e| - ½\delta), otherwise \} \quad (18)$$

$$g'_k = g_k\, min(1, c / ||g_k||) \quad (19)$$

$$\theta_{k+1} = \theta_k - \eta\, \hat{m}_k/(\sqrt{\hat{v}_k} + \varepsilon) - \eta\lambda\theta_k \quad (20)$$

$$\theta * = arg\, min_k\, Lval(\theta_k) \quad (21)$$

Forecasting performance was assessed using MAE, RMSE, coefficient of determination (R^2), and symmetric mean absolute percentage error (sMAPE). sMAPE was preferred over conventional MAPE because the daily series contains low counts and a missing source day, making division-based MAPE unstable.

$$MAE = (1/n)\, \Sigma_{i=1}^{n} |y_i - \hat{y}_i| \quad (22)$$

$$RMSE = \sqrt{[(1/n)\, \Sigma_{i=1}^{n} (y_i - \hat{y}_i)^2]} \quad (23)$$

$$R^2 = 1 - [\Sigma_i (y_i - \hat{y}_i)^2 / \Sigma_i (y_i - \bar{y})^2] \quad (24)$$

$$sMAPE = (100/n)\, \Sigma_{i=1}^{n} |y_i - \hat{y}_i| / [(|y_i| + |\hat{y}_i|)/2] \quad (25)$$

The study does not equate predictive accuracy with sustainability. Instead, the sustainability interpretation follows a decision-support pathway: forecasted DCEP activity can indicate periods when surveillance intensity or preparedness may need to increase, while locality-level concentrations can

identify where additional investigation or targeted communication may be prioritized. Potential sustainability benefits are therefore framed as improved targeting and coordination. No direct claim is made about financial savings, energy reduction, carbon reduction, or causal improvements in health outcomes because those quantities were not measured.

$$P_{l,t} = wB\, B_t + wF\, F_t + wL\, L_{l,t} \quad (26)$$

$$wB + wF + wL = 1 \quad (27)$$

4. Experimental Design

4.1 Mathematical and Experimental Robustness

A future multi-year evaluation should vary look-back length L, forecast horizon h, and model architecture while preserving chronological separation. Relative sensitivity can be reported as:

$$S_M = [(M - M_0) / M_0] \times 100\% \quad (28)$$

Although the underlying dataset contains 151,962 patient records, the forecasting task is based on a short sequence of daily observations. Patient records within the same day are therefore not independent forecasting samples. This distinction limits the statistical strength of model comparisons and is why the present results are interpreted as a case-study evaluation.

The experimental design was intentionally constrained by the available data. The dataset provides a large number of individual records but only four months of temporal coverage. Consequently, the primary evaluation unit is the daily DCEP count rather than patient-level diagnosis, and the forecast horizon is limited to one day ahead. This avoids overstating the ability of the models to learn annual seasonality from a short time series.

The chronological train–validation–test configuration and forecasting settings are summarized in Table 3.

Table 3. Chronological experimental partitioning and forecasting configuration used for model training, validation, and testing.

Component	Period/setting	Purpose
Training	1 May–30 June 2011	Model fitting
Validation	1–31 July 2011	Early stopping and model selection

Testing	1–31 August 2011	Final evaluation
Observed test days	30	13 August excluded because of source-data gap
Look-back	14 days	Previous two weeks of standardized features
Target	Daily DCEP count	One-day-ahead forecast
Classical baselines	July validation; August rolling test	ARIMA, ETS, seasonal-naive, linear AR, persistence

Three deep architectures and four classical baselines were evaluated under the same target definition and chronological test period, together with persistence. The deep models use the 14-day standardized-feature look-back, whereas the classical models use the DCEP daily series directly. Classical configurations were selected using July validation and evaluated on the observed August test dates with one-step-ahead rolling forecasts. Because the test period contains only 30 observed days after excluding the 13 August source-data gap, the results are interpreted as a case-study comparison rather than a definitive benchmark of forecasting performance.

5. Results

5.1 Dataset Characteristics

The final dataset contained 151,962 emergency records, of which 133,398 were from the main emergency service and 18,564 from paediatric emergency services. No record contained the explicit term 'dengue' in the complaint fields. The DCEP rule identified 62,537 records, corresponding to 41.15% of all records. All DCEP records originated from the main-emergency dataset under the adopted symptom rule; the paediatric complaint field contained fever-related presentations but did not satisfy the combined fever-plus-associated-symptom rule. The monthly emergency-presentation volume, DCEP volume, and DCEP proportion are summarized in Table 4.

Table 4. Monthly emergency-presentation volume, DCEP volume, and DCEP proportion during the observation period.

Month	All presentations	DCEP	DCEP proportion
May 2011	36,218	14,725	40.66%

June 2011	37,560	15,667	41.71%
July 2011	38,447	14,909	38.78%
August 2011	39,737	17,236	43.38%

Total emergency volume increased from 36,218 records in May to 39,737 in August. DCEP volume increased from 14,725 in May to 17,236 in August, with a 15.61% increase from July to August. The largest daily DCEP count was 2,008 on 28 August, followed by 1,296 on 29 June, 1,111 on 29 July, and 1,102 on 28 July. The daily DCEP trajectory is shown in Fig. 2, while the corresponding monthly DCEP volume is visualized in Fig. 3.

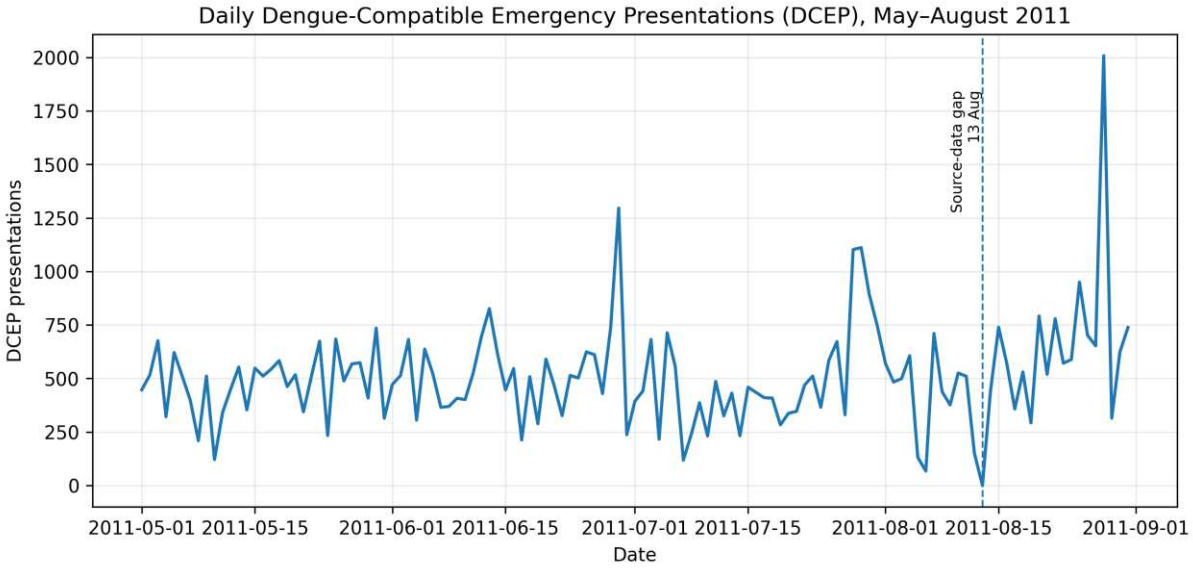


Figure 2. Daily DCEP volume across the May–August 2011 observation period; the dashed marker indicates the source-data gap on 13 August.

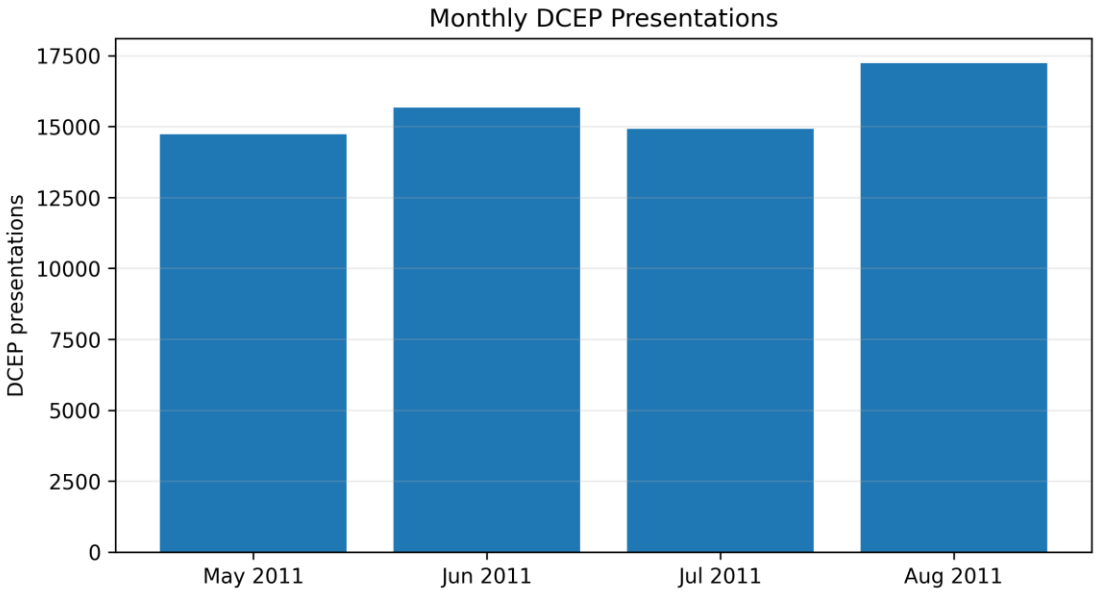


Figure 3. Monthly DCEP volume across the May–August 2011 observation period.

5.2 Locality-Level Surveillance Concentration

Address-string analysis showed strong concentration in a small number of recurring source labels. The most frequent standardized string among DCEP records was GOWALMANDI LAHORE (16,152 records), followed by NISBAT ROAD LAHORE (2,076), SHEKHUPURA (1,433), SHAHDARA LHR (874), and SHAHDARA (811). The ten most frequent standardized address strings represented 23,491 DCEP records, or approximately 37.56% of all DCEP records. These figures should not be interpreted as population-adjusted disease risk because population denominators, geocoded coordinates, and catchment-area information were unavailable.

Table 5 lists the ten most frequent standardized address strings, while Figure 4 visualizes their relative surveillance-workload concentration.

Table 5. Ten most frequent standardized address strings among DCEP records, reported as surveillance-workload concentrations.

Standardized address string	DCEP records
Gowalmandi Lahore	16152
Nisbat Road Lahore	2076
Shekhupura	1433

Shahdara Lhr	874
Shahdara	811
Badami Bagh Lhr	508
Mozang Lahore	437
Ravi Rd Lhr	419
Misrishah Lahore	392
Lohari Gate Lhr	389

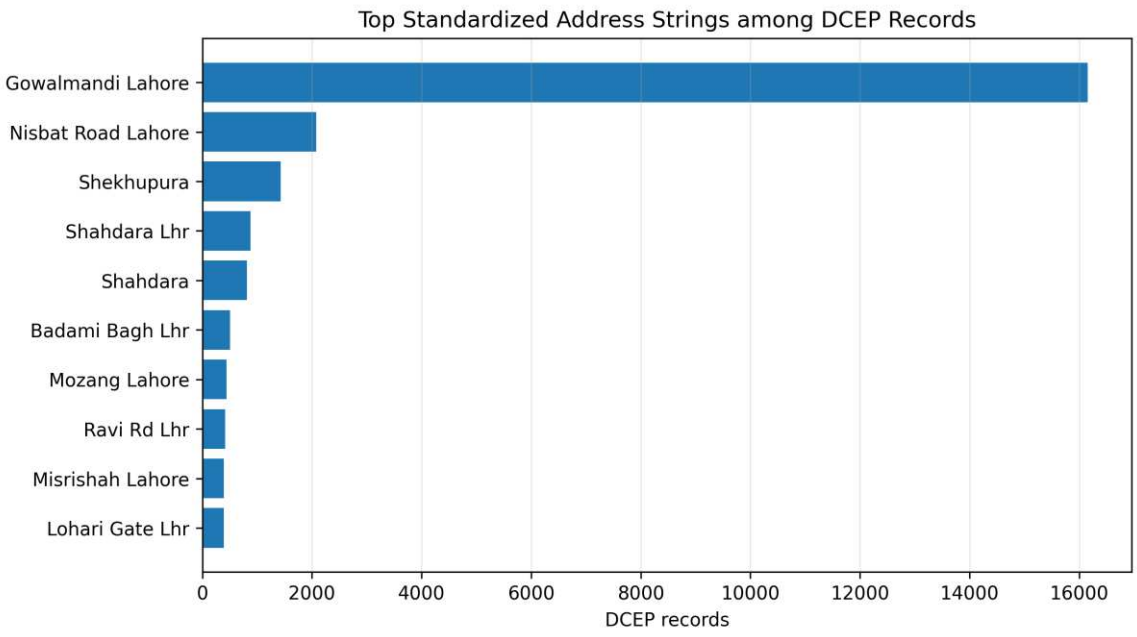


Figure 4. Ten most frequent standardized address strings among DCEP records, representing surveillance-workload concentrations rather than population-normalized incidence rates.

5.3 Forecasting Performance across Classical and Deep-Learning Models

The expanded benchmark shows that both classical and deep-learning models improve substantially over persistence on the observed August test days. GRU achieved the lowest MAE (204.60 presentations) and sMAPE (37.39%), while linear autoregression achieved the lowest RMSE (335.60). ARIMA was closely competitive, with MAE 207.11 and RMSE 336.17, and ETS also improved materially over persistence. Seasonal-naive provided a useful weekly reference but remained weaker than the other classical and deep-learning models. Among the deep models, GRU remained the

strongest on MAE and sMAPE, followed by TCN and LSTM. All R^2 values remained below zero on the short test period, so the results should be interpreted as lower-error workload anticipation rather than strong variance explanation.

The complete comparison is reported in Table 6, with error magnitudes visualized in Figure 5 and observed-versus-GRU forecasts shown in Figure 6.

Table 6. Forecasting performance of persistence, classical time-series baselines, and deep-learning models on the observed August test days.

Model	MAE	RMSE	R^2	sMAPE (%)
Persistence	306.53	472.68	-1.024	52.40
Seasonal-naïve	266.10	427.93	-0.659	46.44
ARIMA (2,0,1)	207.11	336.17	-0.024	37.79
ETS	231.53	352.45	-0.125	41.04
Linear AR (3)	209.16	335.60	-0.020	37.71
LSTM	215.99	348.97	-0.103	39.91
GRU	204.60	342.11	-0.060	37.39
TCN	210.61	345.10	-0.079	38.74

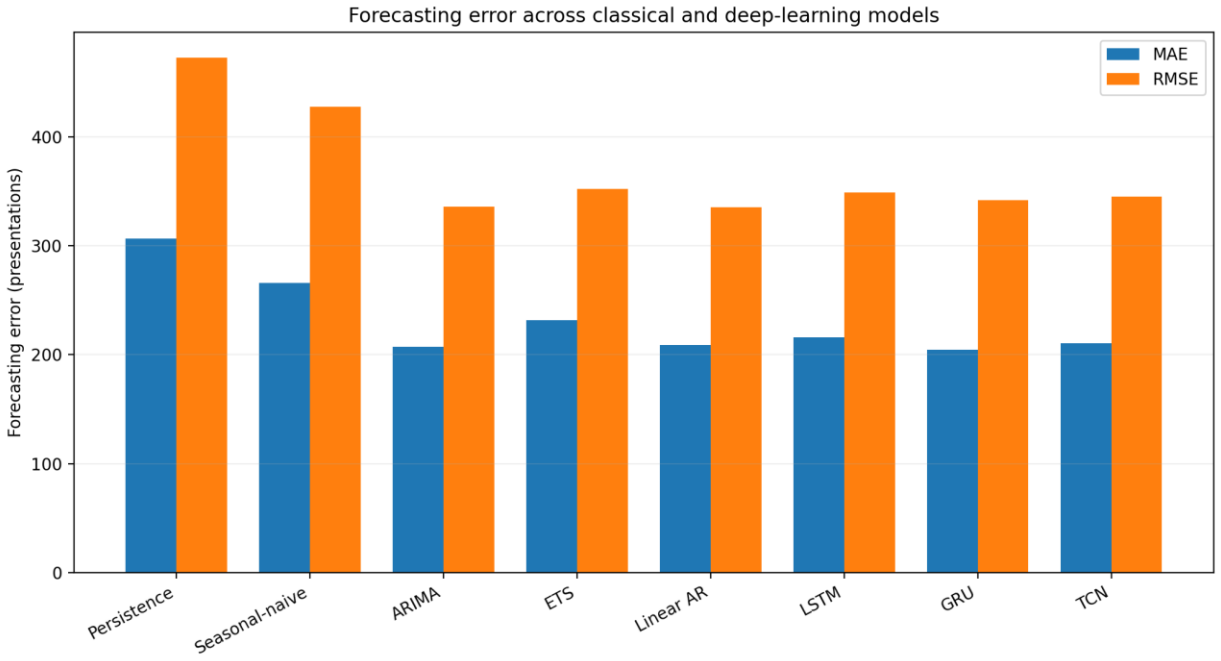


Figure 5. Comparison of MAE and RMSE forecasting errors across persistence, classical time-series baselines, and deep-learning models; lower values indicate better performance.

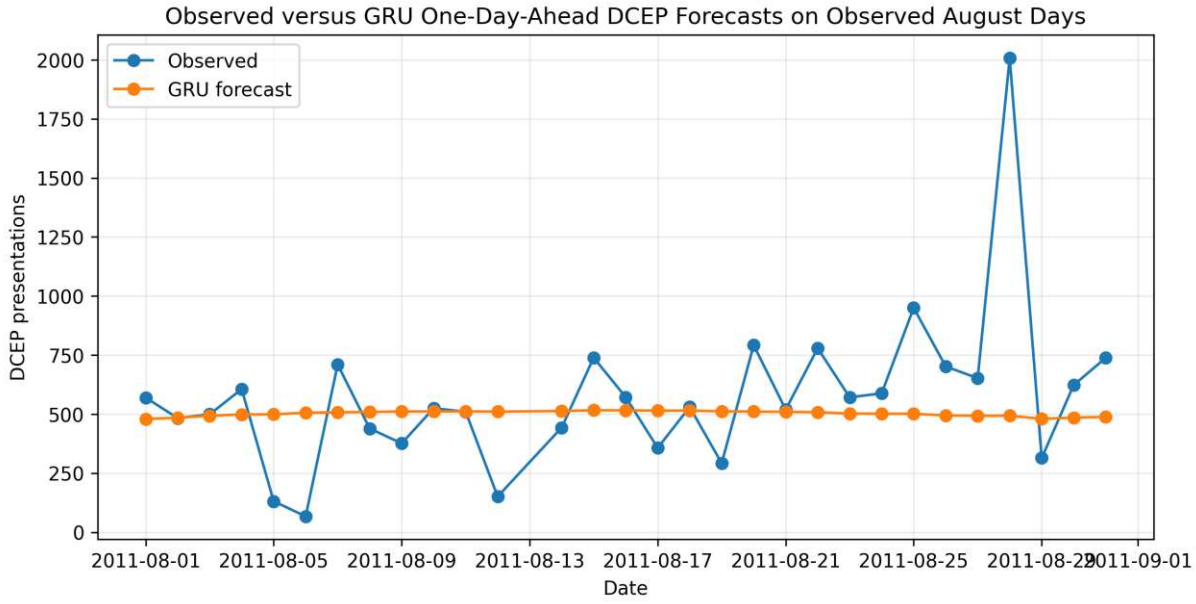


Figure 6. Observed versus GRU one-day-ahead DCEP forecasts for the observed August test days; the source-data gap on 13 August is excluded.

5.3.1 Comparative Error Reduction

The numerical results show that the advantage over persistence is not unique to deep learning. Seasonal-naive reduced MAE by 13.20% and RMSE by 9.47%. ARIMA reduced MAE by 32.44% and RMSE by 28.88%, ETS reduced MAE by 24.47% and RMSE by 25.43%, and linear autoregression reduced MAE by 31.76% and RMSE by 29.00%. Among the deep models, GRU reduced MAE by 33.25%, RMSE by 27.62%, and sMAPE by 28.65%. Thus, the broader empirical finding is that useful temporal structure can be exploited by both conventional and deep forecasting approaches, while the best model depends on the evaluation metric.

The relative improvements are summarized in Table 7 and visualized in Figure 7.

The magnitude of the improvement is meaningful for a short-horizon surveillance workload. The strongest MAE improvement is obtained by GRU, while linear autoregression gives the smallest RMSE. These differences show that the added classical baselines are not merely nominal comparators; they are competitive forecasting alternatives. Nevertheless, remaining errors are substantial relative to daily surveillance volume, so the models should be interpreted as workload-anticipation tools rather than precise daily-count estimators.

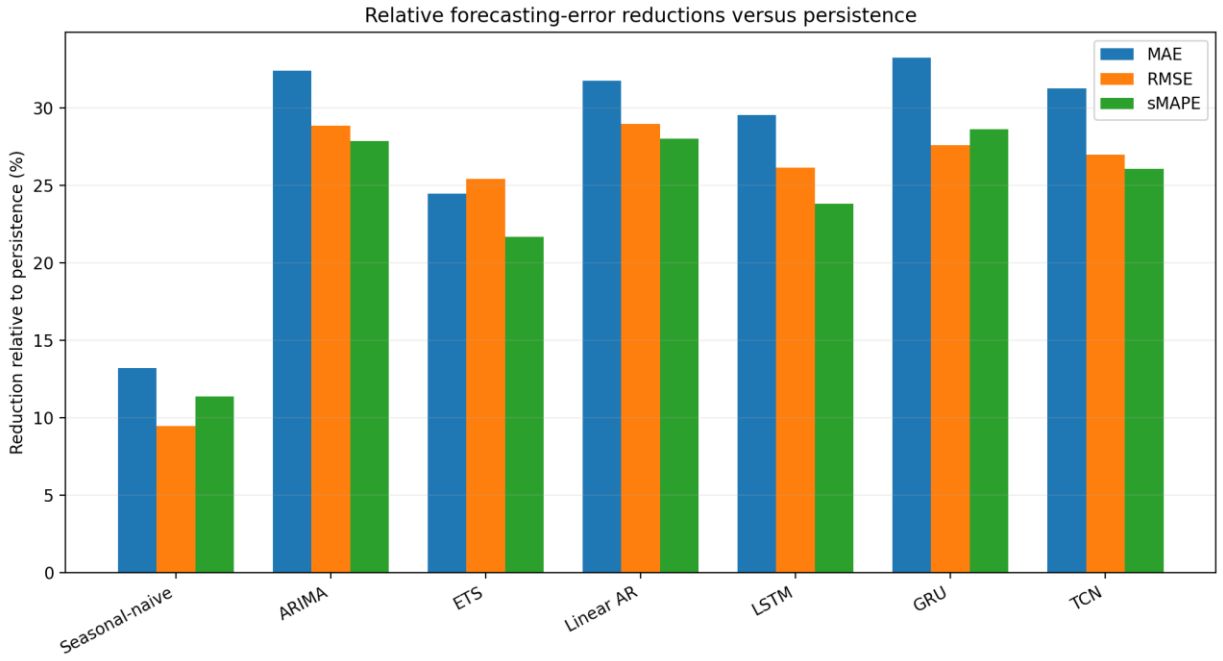


Figure 7. Relative reductions in MAE, RMSE, and sMAPE achieved by classical and deep-learning models relative to persistence.

5.3.2 Normalized Performance and Model Separation

Normalizing each error measure by the corresponding persistence error provides a scale-independent view of model separation. Seasonal-naive requires 86.80% of the persistence MAE and 90.53% of its RMSE, whereas ARIMA requires 67.56% and 71.12%, ETS 75.53% and 74.57%, and linear autoregression 68.24% and 71.00%, respectively. Among the deep models, GRU requires 66.75% of persistence MAE, 72.38% of its RMSE, and 71.35% of its sMAPE. Thus, GRU has the lowest normalized MAE and sMAPE, while linear autoregression has the lowest normalized RMSE, with ARIMA closely behind.

Figure 8 shows that all benchmark models reduce normalized error relative to persistence. ARIMA and linear autoregression have the lowest normalized RMSE ratios, while GRU has the lowest normalized MAE and sMAPE ratios. This metric-dependent ranking reinforces the need to evaluate multiple baselines.

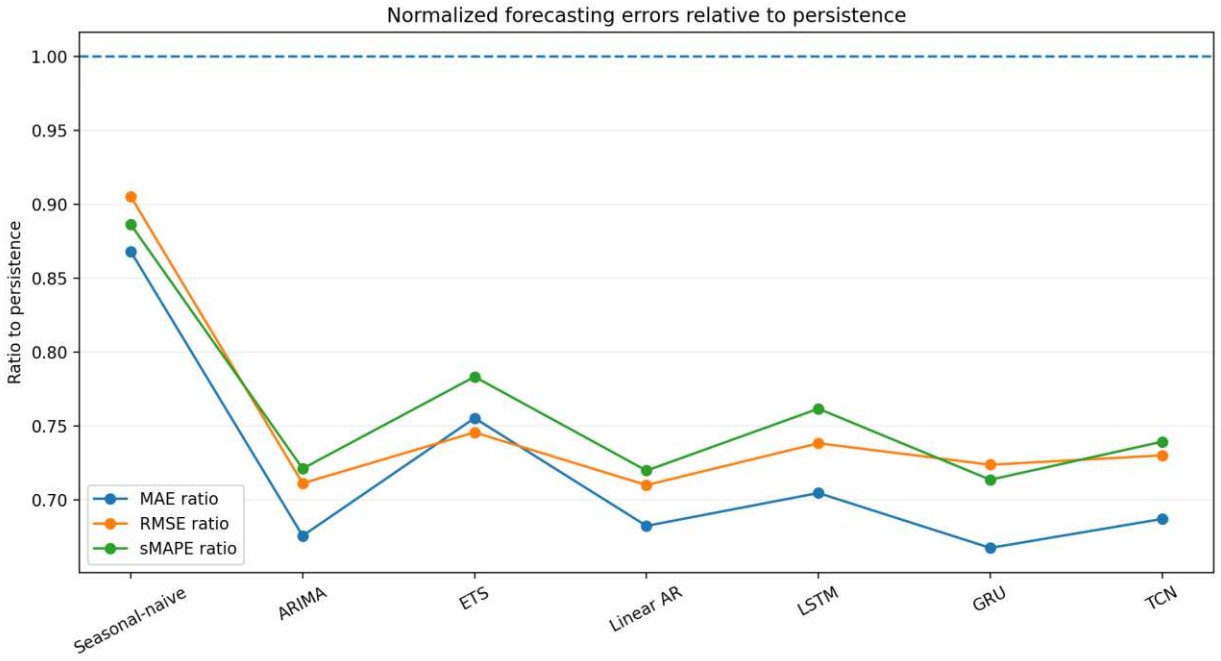


Figure 8. Normalized MAE, RMSE, and sMAPE error ratios relative to persistence across all forecasting models; values below 1.0 indicate improvement.

Table 7. Relative improvement of classical and deep-learning models over the persistence baseline across MAE, RMSE, and sMAPE.

Model	MAE reduction	RMSE reduction	sMAPE reduction	MAE rank
Seasonal-naive	13.20%	9.47%	11.36%	7
ARIMA (2,0,1)	32.44%	28.88%	27.88%	2
ETS	24.47%	25.43%	21.67%	6
Linear AR (3)	31.76%	29.00%	28.03%	3
LSTM	29.54%	26.17%	23.84%	5
GRU	33.25%	27.62%	28.65%	1
TCN	31.29%	26.99%	26.07%	4
Persistence	0.00%	0.00%	0.00%	8

The separation across the whole benchmark is more significant than the separation among deep models alone. GRU has the lowest MAE and sMAPE while linear autoregression has the lowest RMSE and ARIMA is a close second. The deep models are very closely grouped with GRU, TCN, and LSTM with relatively small absolute differences. This ranking is not meant to argue that deep learning is always

better than other types of models but to show that different models are better under different data and time conditions.

5.3.3 MAE–RMSE Relationship and Error Behaviour

The MAE–RMSE relationship provides more insight into the error behaviour. Persistence shows the largest errors, and the classical and deep learning models cluster much tighter. Linear autoregression is the most favourable RMSE model, followed by ARIMA; GRU is the most favourable MAE model. Because RMSE weights larger errors more strongly than MAE, linear autoregression and ARIMA are very competitive for controlling larger forecast misses, whereas GRU is more attractive for average absolute error and percentage-scaled error.

Figure 9 shows the MAE–RMSE trade-off. Linear autoregression and ARIMA are the two most favorable RMSE regions, while GRU has the lowest MAE. Persistence is the least favourable benchmark, and seasonal-naive is an intermediate but weaker reference.

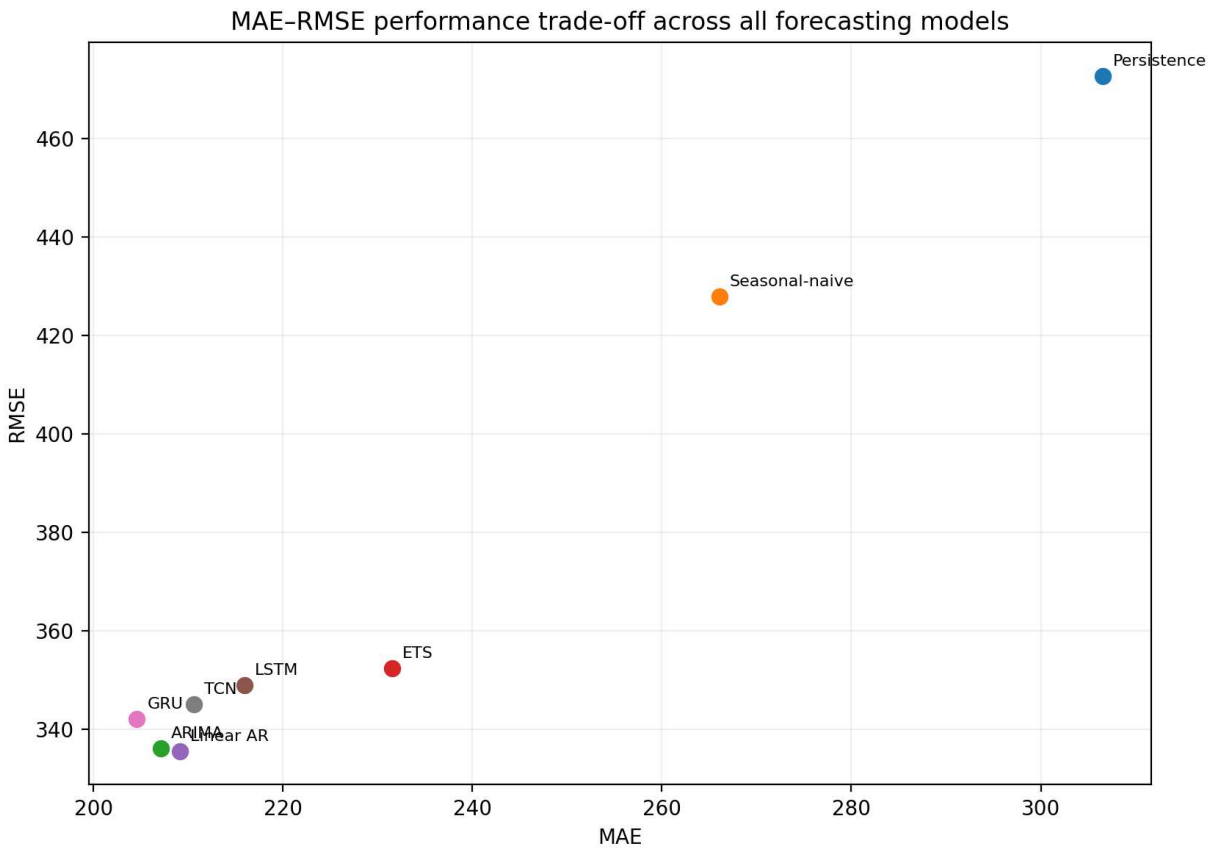


Figure 9. MAE–RMSE performance trade-off across persistence, classical time-series baselines, and deep-learning models.

5.3.4 Interpretation of Negative R^2

The negative R^2 values require particular attention. Linear autoregression (−0.020) and ARIMA (−0.024) are closest to zero, while GRU is −0.060, TCN −0.079, LSTM −0.103, ETS −0.125, seasonal-naive −0.659, and persistence −1.024. Thus, lower forecast error does not imply positive variance-explanation performance on this short test sequence. The appropriate interpretation is that the models, particularly ARIMA, linear autoregression, and GRU, reduce forecasting error relative to simple references but do not establish strong explanatory power or long-horizon generalization.

This distinction is important for public-health interpretation. A model can be useful for anticipating approximate workload while still being unable to reproduce sharp peaks accurately. Figure 6 visually illustrates this issue: the GRU forecast is comparatively smooth, whereas observed DCEP counts contain abrupt excursions, including the large late-August peak. Such smoothing can reduce average error but can also cause underestimation of extreme surveillance-demand days. Consequently, the forecasting layer should be combined with threshold-based monitoring of observed burden and expert review rather than used as an autonomous peak-event detector.

5.3.5 Overall Finding

Taken together, the results provide a stronger and more balanced empirical conclusion. Under the common chronological train–validation–test design and one-day-ahead objective, all evaluated classical and deep-learning models improved substantially over persistence on absolute-error measures. GRU produced the lowest MAE and sMAPE, whereas linear autoregression produced the lowest RMSE and ARIMA was closely competitive. The findings therefore do not support a claim that deep learning is universally superior on this short four-month series. Instead, they demonstrate that DenVis-S benefits from a comparative forecasting layer in which conventional and deep temporal models can be evaluated under identical data constraints.

5.4 Risk-Oriented Surveillance Interpretation

The results support a risk-oriented rather than diagnostic interpretation. The large DCEP peaks on 28 August, 29 June, 29 July, and 28 July identify periods of unusually high emergency surveillance demand under the adopted symptom definition. When combined with recurring locality concentrations, these peaks provide a basis for prioritizing review of emergency preparedness, vector-control intelligence, and community communication. The GRU forecasts can be used as an additional signal for short-term workload anticipation, but the negative R^2 values demonstrate that the forecasts should be treated as supportive intelligence rather than autonomous outbreak declarations.

6. Sustainability-Oriented Decision Support

6.1 Sustainability-Oriented Resource Prioritization

DenVis-S supports a resource-prioritization perspective by combining three complementary information types: current surveillance burden, spatial concentration of source addresses, and short-term predicted demand. These signals could help public-health teams to identify where more review, preparedness, communication, or investigative attention may be warranted whereas in this sense it is consistent with the sustainability principle that digital health innovations should be evaluated in terms of system-level efficiency, coordination, usability, and resource implications [4,10,59–62].

The results of our study are the basis for this concept. We had 151,962 emergency records with 62,537 DCEP records (41.15% of all emergency presentations). The monthly volume of DCEP increased from 14,725 in May to 17,236 in August, and the highest number of DCEP on a single day was 2,008 on 28 August. On the locality level, the top 10 most common standardized address strings had 23,491 DCEP records (about 37.56% of all DCEP records). This shows that the surveillance workload was heterogeneous in time and source locality. However, address-based concentrations are surveillance-workload concentrations rather than population-adjusted incidence because population denominators, geocoded coordinates, and catchment-area information were not available.

The forecasting results add a short-term expectation element to this resource planning model. The classical and deep learning models performed well over persistence, with GRU achieving the lowest MAE and sMAPE and linear autoregression with the lowest RMSE. These results suggest that the forecasting layer can be useful for workload-anticipation information other than just last-observation persistence, but they also indicate that a sustainable decision-support system should not expect a more complex deep learning architecture to be superior to it because of robustness, computational requirements, interpretability, and operational requirements.

The sustainability implication should therefore be framed as better targeting and coordination and not as resource savings. This study does not report financial expense, staff-time savings, energy consumption, carbon reduction, or intervention efficacy. DenVis-S being able to achieve these benefits would need to be evaluated at the system level to look at resource allocation, workload, response time, and operational performance, both before and after the implementation. In the following, we show a logical approach to decision-support through predictive surveillance that can help managers focus limited capacity in regions where workload and surveillance demands are high and the operational process is transparent. Algorithm 1 performs this prioritization by combining the observed burden, predicted activity, and locality concentration while keeping human review.

Algorithm 1. Sustainability-Oriented Resource Prioritization

```
Input: observed DCEP burden Y_t; emergency volume E_t; locality counts C_{l,t}; one-
day-ahead forecast Ŷ_{t+1}; configurable thresholds and weights.
Output: ranked surveillance-priority list for human review.
1: Validate temporal completeness and flag source-data gaps; do not treat an
unobserved day as zero activity.
2: Standardize the current burden, forecast, and locality-count measures against
their selected historical reference ranges.
3: Compute temporal burden score B_t and forecast score F_t from the standardized
current and predicted DCEP activity.
4: For each locality l, compute a workload-concentration score L_{l,t} from
standardized DCEP counts.
```

```
5: Compute priority score P_{l,t} = w_B B_t + w_F F_t + w_L L_{l,t}, where w_B + w_F
+ w_L = 1.
6: Rank localities and periods by P_{l,t}; retain the highest-priority items for
qualified human review.
7: Report observed burden, forecast, locality concentration, and data-quality flags
together with the ranking.
8: Return the ranked list; do not convert P_{l,t} directly into an autonomous
intervention or resource-allocation decision.
```

6.2 Human-in-the-Loop Early-Warning and Decision Support

DenVis-S is designed to support decision-making instead of outbreak detection. The DCEP target is determined based on the emergency presentation characteristics and is a syndromic surveillance proxy rather than a laboratory-confirmed diagnosis. Therefore, a rising DCEP trajectory or forecasted increase in a one-day forecast on the trajectory of the DCEP should be a trigger for a human review, not an automatic diagnosis of an outbreak or an automatic intervention response. This interpretation is consistent with early-warning guidance on signal detection, risk assessment, and response [8].

The forecasting results demonstrate why human oversight is still needed. GRU had the best overall error profile, but its R^2 was −0.060, while TCN and LSTM had −0.079 and −0.103 respectively. Figure 6 also shows that the GRU forecast is relatively smooth and does not display the magnitude of the large late-August peak. Thus, a model can improve the average forecasting error while still underestimating the sudden increase in surveillance demand. For practical purposes, the forecast data needs to be evaluated in conjunction with observed DCEP burden, temporal trends, recurring locality concentrations, and independent clinical, laboratory, or field intelligence.

The proposed human-in-the-loop workflow can be described as: (i) follow the observed DCEP and emergency burden; (ii) prepare the one-day-ahead forecast; (iii) examine the recurring source locations and temporal concentration; (iv) validate the signal using clinical, laboratory, and field data; (v) focus on surveillance and preparedness on high-priority episodes or zones; (vi) incorporate new confirmed observations and context into the next system update. This approach keeps the control of the expert and

allows the computational models to alleviate the burden of constantly reviewing large amounts of surveillance information.

Algorithm 2 formalizes this workflow and explicitly keeps human authority on the final early-warning and preparedness decision.

This design also facilitates the governance of AI-supported public health surveillance. The model is viewed as the analytical component of a larger visual and epidemiological process, not as a black-box authority. As such, it is especially appropriate given the short time period of four months, the lack of laboratory-confirmed dengue labels, the missing source day on 13 August, and the lack of population-adjusted geographic information. DenVis-S is best understood as a process for turning surveillance data into interpreted early warning signals that can inform but not replace expert public health advice. The following algorithm is intended to convey this human-in-the-loop approach; it should be seen as a design specification, not an empirical experiment.

Algorithm 2. Human-in-the-Loop Early-Warning and Decision Support

```
Input: current DCEP series; emergency volume; one-day-ahead forecast; locality
concentrations; clinical/laboratory/field intelligence.
Output: human-reviewed early-warning status and documented action.
1: MONITOR current DCEP and total emergency burden for unusual temporal activity.
2: FORECAST next-day DCEP volume using the selected deep-learning model (GRU in the
present case study).
3: LOCALIZE recurring source concentrations using standardized locality strings.
4: CROSS-CHECK the forecast and observed signal against clinical, laboratory, field,
and data-quality information.
5: If evidence is inconsistent or insufficient, assign REVIEW-ONLY status and request
additional verification.
6: If evidence is corroborated, assign a configurable priority level and identify
surveillance/preparedness actions for expert consideration.
7: Record the human decision, supporting evidence, and subsequent confirmed
observations.
```

```
8: UPDATE the surveillance record and contextual inputs for the next forecasting
cycle; retain human authority over final decisions.
```

6.3 Sustainable Use of Limited Surveillance Capacity and Equity

Address- or locality-based prioritization can lead to inequity if perceived as absolute disease risk. Healthcare-seeking behaviour, access barriers, referral patterns, population density, standardization effects, and hospital catchment effects can affect record concentrations. So future deployment should consider locality concentration as a context-specific prioritization cue, and integrate surveillance counts with population denominators, socioeconomic indicators, healthcare-access measures, and validated geographic boundaries before using the system for high-stakes allocation decisions.

6.4 Proposed Operational Decision Cycle

The proposed operational decision cycle takes the results of DenVis-S and transforms them into a structured, human-supervised public-health workflow. Rather than treating the deep-learning forecast as a decision-making process by itself, the process combines DCEP burden, short-term forecasts, locality-level concentration, and context before the action is taken on surveillance or preparedness. This is especially important because DCEP is a syndromic surveillance proxy instead of laboratory-confirmed dengue and our study does not establish population-adjusted geographic risk. DenVis-S is therefore designed to pay more attention to and verify the observed risk, rather than automatically trigger interventions. The whole sequence is summarized in Table 8 and our human-in-the-loop logic is explained in Algorithm 2.

The first stage, monitoring, continuously monitors the daily DCEP burden and total emergency volume to detect unusual activity in the environment. Monitoring is the foundation on which the future changes are expected and helps us to distinguish between small fluctuations and sustained rises. The second stage, forecasting, provides a one-day-ahead DCEP estimate using the selected deep learning model. In this case, the GRU has the lowest point estimate forecasting errors among all the models we have tested but we consider the forecast as additional information from surveillance and not as an independent

prediction of confirmed disease. The third stage, localization, focuses on recurring source localities and spatial concentration. This step adds geographic context to the temporal signal and can reveal areas that are contributing significantly to the observed surveillance workloads. These concentrations may be indicative of healthcare-seeking behaviour, referral pathways, catchment effects, population density, and access differences that are not representative of population-normalized disease risk. The fourth stage, validation, then integrates the observed and forecasted information with available clinical, laboratory, field, and data quality information. When data are inconsistent, incomplete, or insufficient to be considered, the system should return a review-only status rather than escalate the signal.

The fifth stage, prioritization, is only activated when the surveillance signal is sufficiently corroborated. At this stage, public health professionals can set up arbitrary priority levels for periods or localities to monitor further, if possible, and prepare for or plan for them. This allows us to concentrate the scarce resources to focus on areas when multiple evidence streams indicate greater need, with individuals in charge of making the final decision. The sixth stage, learning, stores the following observations and contextual information and human decisions as well as updates the surveillance record for future forecasting cycles. So this cycle is not a one-time prediction pipeline, but an iterative process.

In the end, the operational value of DenVis-S is to provide a repeatable transition from data to evidence, from evidence to human review, and from validated signals to targeted planning. This structure will ensure continued use of surveillance capacity because it does not allocate attention to all periods and regions at the same time but acknowledges uncertainty and the limitations of syndromic data. Therefore, the cycle should be viewed as a proposed operational design specification rather than a validated intervention-allocation algorithm and prospective evaluation would be performed to see whether it improves the response time, resource efficiency, equity, or public-health implications.

The proposed sequence is summarized in Table 8, showing monitoring, forecasting, localization, cross-checking, and documented human action.

Table 8. Proposed operational decision cycle linking surveillance monitoring, forecasting, localization, human review, and documented action.

Step	Function	Decision-support action
1	Monitor	Review daily DCEP burden and emergency volume
2	Forecast	Generate one-day-ahead DCEP estimates using the selected DL model
3	Localize	Inspect recurring source localities and spatial concentration
4	Validate	Compare model signal with clinical, laboratory, and field intelligence
5	Prioritize	Allocate surveillance and preparedness attention to validated high-priority periods/areas
6	Learn	Feed new confirmed observations and contextual variables back into the system

7. Discussion

The forecasting experiment revealed some complementary findings. First, the deep learning architectures evaluated can reduce the absolute forecasting error compared to persistence, indicating that the multivariate temporal representation might contain more information than the immediate target value. Second, differences between LSTM, GRU, and TCN should be interpreted cautiously since short-horizon forecasting performance depends on the data and evaluation window size. The results are thus more a case for comparative experimentation than a claim of universal architectural superiority. Negative or weak R^2 values are equally important to interpret. They suggest that the models can increase absolute error without explaining a significant fraction of the variance in the short test period. Thus, we should consider deep learning in this context as part of surveillance intelligence rather than as an independent predictive tool. Longer multi-year datasets with confirmed results and contextual information would be required to ensure that our results are generalizable across seasons, locations, and epidemic cycles. The high number of DCEP records in recurring address strings would suggest that emergency surveillance at the hospital is spatially structured. However, the magnitude of GOWALMANDI and other address labels may be related to population density, proximity to the hospital, referral behavior, address standardization, spelling duplications, or catchment effects as well as disease burden. Therefore, the current analysis should be viewed as a tool to find surveillance concentrations instead of as a valid geographic incidence map.

Compared to longer-term forecasting studies that incorporate climate, land-use, demographic, and satellite variables [5–7], DenVis-S has a much narrower information base, focusing on transparent target construction, short-term deep learning, visual analytics, and decision-making. Its contribution is system-level integration rather than higher predictive sophistication or the introduction of a new neural architecture. Recently, the sustainability and digital health literature also argues that technological interventions should be evaluated in relation to system-level sustainability and not only technical performance [4,10].

DenVis-S supports a shift from retrospective reporting to anticipatory surveillance, and this is a good thing for sustainable health systems because they do not have enough preparedness resources and demand is geographically diverse. However, sustainable deployment is still to be done with governance, interoperability, user training, data quality monitoring, model recalibration, and human supervision. The framework should therefore be seen as a decision-support system that is integrated into a wider public health system and should not be substituted for laboratory surveillance or expert judgment.

The framework is structurally generalizable because its preprocessing, symptom-rule, forecasting, and visualization modules can be replaced or extended. On the other hand, its empirical generalizability is very limited: the current evaluation covers only one hospital, one city, and four months of 2011, so it cannot prove performance across other hospitals, years, climatic conditions, or healthcare systems. Multi-site, multi-year, and prospective validation are necessary before operational claims can be made.

8. Limitations and future work

A key limitation of our approach is the number of independent time-series observations. A large number of record-level observations does not necessarily mean a large number of independent time-series observations. Our results should be interpreted not as a clear difference between deep architectures but as an empirical comparison of the two in the evaluation context.

Diagnostic specificity. Syndromic or presentation-based targets may include non-target conditions. Other studies should connect surveillance records with laboratory-confirmed or clinically adjudicated outcomes where available.

Temporal coverage. Short observation periods limit our ability to learn annual seasonality and longer-term epidemic dynamics. Multi-year data are needed to evaluate seasonal generalization and longer forecasting horizons.

Data completeness. Missing or structurally incomplete observation periods can bias time analysis if interpreted as true zero activity. Future systems should automatically record and monitor data-completeness indicators.

8.1 Future Research Agenda

The principal extensions are summarized in Table 9, including richer contextual data, improved model validation, spatial refinement, uncertainty quantification, and prospective usability evaluation.

Table 9. Future research priorities for extending DenVis-S with richer contextual data, stronger validation, and operational deployment.

Area	Priority development
Data integration	Add laboratory confirmation, climate, vector, mobility, population, and land-use variables
Deep learning	Evaluate attention-based LSTM/GRU, ConvLSTM, TCN variants, and transformer models on multi-year data
Spatial modelling	Integrate geocoded coordinates and population-adjusted risk surfaces
Validation	Conduct multi-hospital and multi-year external validation
Sustainability	Measure intervention costs, staff workload, resource utilization, and environmental indicators
Decision support	Evaluate user task performance and public-health decision quality

9. Conclusions

In this paper, we presented DenVis-S, a sustainable visual analytics framework for intelligent disease surveillance, risk assessment and public health resource planning. In this framework we combine transparent surveillance-target construction, spatio-temporal visualization, deep learning forecasting, and human-centred decision-support logic in a modular workflow. The experimental evaluation clearly demonstrates how LSTM, GRU, and TCN models can be compared to a persistence baseline for short-horizon surveillance forecasting. Our results support the use of deep learning as an additional predictive intelligence, but it is important to be cautious when temporal coverage, diagnostic specificity, spatial precision, or independent sample size is not available.

From a sustainability perspective, DenVis-S is a way to target scarce surveillance and preparedness resources to periods and locations where the signals are high. Given that actual economic, environmental and health-system savings need to be measured, those benefits should be treated as hypotheses for future analysis. Future work should incorporate confirmed outcomes, multi-period and multi-site observations, environmental and population covariates, confirmed spatial data and direct measures of resource utilization. In this way, DenVis-S will be able to grow from a prototype to a more robust sustainable public-health intelligence platform.

Author Contributions

Conceptualization, F.A. and A.S.; methodology, F.A., A.S., and S.R.; software, F.A. and A.S.; validation, F.A., A.S., and S.R.; formal analysis, F.A. and S.R.; investigation, A.S. and S.R.; data curation, A.S.; writing—original draft preparation, F.A. and A.S.; writing—review and editing, F.A., A.S., and S.R.; visualization, A.S.; supervision, F.A.; project administration, F.A.; funding acquisition, F.A. All authors have read and agreed to the published version of the manuscript.

Funding

No external funding was received for this study.

Clinical Trial Number

N/A

Ethics and Consent to Participate declarations

N/A

Data Availability Statement

The experimental dataset consists of retrospective emergency records from Mayo Hospital, Lahore, for May–August 2011. The data are not publicly deposited in this study. The dataset is available on request.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. World Health Organization. Dengue and severe dengue. WHO Fact Sheet; 2025.

2. World Health Organization. Vector-borne diseases. WHO Fact Sheet; 2024.

3. World Health Organization. Dengue—Pakistan. WHO Health Topic; 2025.

4. Faggini, M.; Cosimato, S.; Nota, F.D.; Nota, G. Pursuing Sustainability for Healthcare through Digital Platforms. Sustainability 2019, 11, 165. https://doi.org/10.3390/su11010165.

5. Li, Z.; Dong, J.; Yang, L.; et al. Forecasting Weekly Dengue Cases by Integrating Google Earth Engine-Based Risk Predictor Generation and Google Colab-Based Deep Learning Modeling in Fortaleza and the Federal District, Brazil. Int. J. Environ. Res. Public Health 2022, 19, 13555. https://doi.org/10.3390/ijerph192013555.

6. Majeed, M.A.; Shafri, H.Z.M.; Zulkafli, Z.; Wayayok, A. A Deep Learning Approach for Dengue Fever Prediction in Malaysia Using LSTM with Spatial Attention. Int. J. Environ. Res. Public Health 2023, 20, 4130. https://doi.org/10.3390/ijerph20054130.

7. Li, Z.; Dong, J. Big Geospatial Data and Data-Driven Methods for Urban Dengue Risk Forecasting: A Review. Remote Sens. 2022, 14, 5052. https://doi.org/10.3390/rs14195052.

8. World Health Organization. Operational Guide: The Early Warning and Response System (EWARS) for Dengue Outbreaks, 2nd ed.; WHO: Geneva, Switzerland, 2020.

9. Javaid, M.; Sarfraz, M.S.; Aftab, M.U.; Zaman, Q.u.; Rauf, H.T.; Alnowibet, K.A. WebGIS-Based Real-Time Surveillance and Response System for Vector-Borne Infectious Diseases. Int. J. Environ. Res. Public Health 2023, 20, 3740. https://doi.org/10.3390/ijerph20043740.

10. Degavre, F.; Kieffer, S.; Bol, D.; Dekimpe, R.; Desterbecq, C.; Pirson, T.; Sandu, G.; Thibault, S. Searching for Sustainability in Health Systems: Toward a Multidisciplinary Evaluation of Mobile Health Innovations. Sustainability 2022, 14, 5286. https://doi.org/10.3390/su14095286.

11. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Comput. 1997, 9, 1735–1780.

12. Cho, K.; van Merriënboer, B.; Gulcehre, C.; et al. Learning Phrase Representations Using RNN Encoder–Decoder for Statistical Machine Translation. arXiv 2014, arXiv:1406.1078.

13. Bai, S.; Kolter, J.Z.; Koltun, V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv 2018, arXiv:1803.01271.

14. World Health Organization. Dengue: Global Situation, Surveillance and Progress—2024 Update. Weekly Epidemiological Record 2025, 100, 665–678.

15. Lu, W.-C.; Tsai, I.-C.; Wang, K.-C.; et al. Innovation Resistance and Resource Allocation Strategy of Medical Information Digitalization. Sustainability 2021, 13, 7888. https://doi.org/10.3390/su13147888.

16. Bhatt, S.; Gething, P.W.; Brady, O.J.; Messina, J.P.; Farlow, A.W.; Moyes, C.L.; Drake, J.M.; Brownstein, J.S.; Hoen, A.G.; Sankoh, O.; et al. The Global Distribution and Burden of Dengue. Nature 2013, 496, 504–507. https://doi.org/10.1038/nature12060.

17. Stanaway, J.D.; Shepard, D.S.; Undurraga, E.A.; Halasa, Y.A.; Coffeng, L.E.; Brady, O.J.; Hay, S.I.; Bedi, N.; Bensenor, I.M.; Castañeda-Orjuela, C.A.; et al. The Global Burden of Dengue: An Analysis from the Global Burden of Disease Study 2013. Lancet Infect. Dis. 2016, 16, 712–723. https://doi.org/10.1016/S1473-3099(16)00026-8.

18. Messina, J.P.; Brady, O.J.; Golding, N.; Kraemer, M.U.G.; Wint, G.R.W.; Ray, S.E.; Pigott, D.M.; Shearer, F.M.; Johnson, K.; Earl, L.; et al. The Current and Future Global Distribution and Population at Risk of Dengue. Nat. Microbiol. 2019, 4, 1508–1515. https://doi.org/10.1038/s41564-019-0476-8.

19. Ryan, S.J.; Carlson, C.J.; Mordecai, E.A.; Johnson, L.R. Global Expansion and Redistribution of Aedes-Borne Virus Transmission Risk with Climate Change. PLoS Negl. Trop. Dis. 2019, 13, e0007213. https://doi.org/10.1371/journal.pntd.0007213.

20. Kraemer, M.U.G.; Sinka, M.E.; Duda, K.A.; Mylne, A.Q.N.; Shearer, F.M.; Barker, C.M.; Moore, C.G.; Carvalho, R.G.; Coelho, G.E.; Van Bortel, W.; et al. The Global Distribution of the Arbovirus Vectors Aedes aegypti and Aedes albopictus. eLife 2015, 4, e08347. https://doi.org/10.7554/eLife.08347.

21. Wilder-Smith, A.; Ooi, E.E.; Horstick, O.; Wills, B. Dengue. Lancet 2019, 393, 350–363. https://doi.org/10.1016/S0140-6736(18)32560-1.

22. Ebi, K.L.; Nealon, J. Dengue in a Changing Climate. Environ. Res. 2016, 151, 115–123. https://doi.org/10.1016/j.envres.2016.07.026.

23. GBD 2023 Dengue Collaborators. Global Burden of Dengue, 1990–2023. Lancet Infect. Dis. 2025.

24. Hughes, H.E.; Edeghere, O.; O'Brien, S.J.; Vivancos, R.; Elliot, A.J. Emergency Department Syndromic Surveillance Systems: A Systematic Review. BMC Public Health 2020, 20, 1891. https://doi.org/10.1186/s12889-020-09949-y.

25. Yeng, P.K.; Woldaregay, A.Z.; Solvoll, T.; Hartvigsen, G. Cluster Detection Mechanisms for Syndromic Surveillance Systems: Systematic Review and Framework Development. JMIR Public Health Surveill. 2020, 6, e11512. https://doi.org/10.2196/11512.

26. Adedire, O.; Love, N.K.; Hughes, H.E.; Buchan, I.; Vivancos, R.; Elliot, A.J. Early Detection and Monitoring of Gastrointestinal Infections Using Syndromic Surveillance: A Systematic Review. Int. J. Environ. Res. Public Health 2024, 21, 489. https://doi.org/10.3390/ijerph21040489.

27. Henning, K.J. Overview of Syndromic Surveillance—What Is Syndromic Surveillance? MMWR Suppl. 2004, 53, 5–11.

28. Mandl, K.D.; Overhage, J.M.; Wagner, M.M.; Lober, W.B.; Sebastiani, P.; Mostashari, F.; Pavlin, J.A.; Gesteland, P.H.; Treadwell, T.; Koski, E.; et al. Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience. J. Am. Med. Inform. Assoc. 2004, 11, 141–150. https://doi.org/10.1197/jamia.M1356.

29. Buehler, J.W.; Hopkins, R.S.; Overhage, J.M.; Sosin, D.M.; Tong, V. Framework for Evaluating Public Health Surveillance Systems for Early Detection of Outbreaks: Recommendations from the CDC Working Group. MMWR Recomm. Rep. 2004, 53(RR-5), 1–11.

30. Lombardo, J.S.; Burkom, H.; Elbert, E.; Magruder, S.; Sari, J.; Kay, J.; et al. A Systems Overview of the Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE II). J. Urban Health 2003, 80(Suppl. 1), i32–i42.

31. Fricker, R.D., Jr.; Hegler, B.L.; Dunfee, D.A. Comparing Syndromic Surveillance Detection Methods: EARS' versus a CUSUM-Based Methodology. Stat. Med. 2008, 27, 3407–3429. https://doi.org/10.1002/sim.3197.

32. Kulldorff, M. A Spatial Scan Statistic. Commun. Stat. Theory Methods 1997, 26, 1481–1496. https://doi.org/10.1080/03610929708831995.

33. Leung, X.Y.; Islam, R.M.; Adhami, M.; Ilic, D.; McDonald, L.; Palawaththa, S.; Diug, B.; Munshi, S.U.; Karim, M.N. A Systematic Review of Dengue Outbreak Prediction Models: Current Scenario and

Future Directions. PLoS Negl. Trop. Dis. 2023, 17, e0010631. https://doi.org/10.1371/journal.pntd.0010631.

34. Lima, C.L.; da Silva, A.C.G.; Moreno, G.M.M.; Cordeiro da Silva, C.; Musah, A.; Aldosery, A.; Dutra, L.; Ambrizzi, T.; Borges, I.V.G.; Tunali, M.; et al. Temporal and Spatiotemporal Arboviruses Forecasting by Machine Learning: A Systematic Review. Front. Public Health 2022, 10, 900077. https://doi.org/10.3389/fpubh.2022.900077.

35. Sylvestre, E.; Joachim, C.; Cécilia-Joseph, E.; Bouzillé, G.; Campillo-Gimenez, B.; Cuggia, M.; Cabié, A. Data-Driven Methods for Dengue Prediction and Surveillance Using Real-World and Big Data: A Systematic Review. PLoS Negl. Trop. Dis. 2022, 16, e0010056. https://doi.org/10.1371/journal.pntd.0010056.

36. Baharom, M.; Ahmad, N.; Hod, R.; Abdul Manaf, M.R. Dengue Early Warning System as Outbreak Prediction Tool: A Systematic Review. Risk Manag. Healthc. Policy 2022, 15, 871–886. https://doi.org/10.2147/RMHP.S361106.

37. Nguyen, V.-H.; Tuyet Hanh, T.T.; Mulhall, J.; Minh, H.V.; Duong, T.Q.; Chien, N.V.; et al. Deep Learning Models for Forecasting Dengue Fever Based on Climate Data in Vietnam. PLoS Negl. Trop. Dis. 2022, 16, e0010509. https://doi.org/10.1371/journal.pntd.0010509.

38. Zhang, Q.; Sun, K.; Chinazzi, M.; Pastore y Piontti, A.; Dean, N.E.; Rojas, D.P.; et al. Forecast of Dengue Cases in 20 Chinese Cities Based on the Deep Learning Method. Int. J. Environ. Res. Public Health 2020, 17, 453. https://doi.org/10.3390/ijerph17020453.

39. Zhao, X.; Li, K.; Ang, C.K.E.; Cheong, K.H. A Deep Learning Based Hybrid Architecture for Weekly Dengue Incidences Forecasting. Chaos Solitons Fractals 2023, 168, 113170. https://doi.org/10.1016/j.chaos.2023.113170.

40. Panja, M.; Chakraborty, T.; Nadim, S.S.; Ghosh, I.; Kumar, U.; Liu, N. An Ensemble Neural Network Approach to Forecast Dengue Outbreak Based on Climatic Condition. Chaos Solitons Fractals 2023, 167, 113124. https://doi.org/10.1016/j.chaos.2023.113124.

41. Francisco, M.E.; Carvajal, T.M.; Watanabe, K. Hybrid Machine Learning Approach to Zero-Inflated Data Improves Accuracy of Dengue Prediction. PLoS Negl. Trop. Dis. 2024, 18, e0012599. https://doi.org/10.1371/journal.pntd.0012599.

42. van den Hurk, A.F.; Hall-Mendelin, S.; Pyke, A.T.; Smith, G.A.; Mackenzie, J.S. Vector-Borne Diseases in a Changing Climate: The Role of Forecasting and Early Warning Systems. Curr. Opin. Environ. Sustain. 2016, 21, 1–7.

43. Simmons, C.P.; Farrar, J.J.; Nguyen, v.V.; Wills, B. Dengue. N. Engl. J. Med. 2012, 366, 1423–1432. https://doi.org/10.1056/NEJMra1110265.

44. Guzman, M.G.; Harris, E. Dengue. Lancet 2015, 385, 453–465. https://doi.org/10.1016/S0140-6736(14)60572-9.

45. Shaban-Nejad, A.; Michalowski, M.; Buckeridge, D.L. Health Intelligence: How Artificial Intelligence Transforms Public Health Surveillance. Yearb. Med. Inform. 2023, 32, 120–131.

46. Kamel Boulos, M.N.; Peng, G.; VoPham, T. An Overview of GeoAI Applications in Health and Healthcare. Int. J. Health Geogr. 2019, 18, 7. https://doi.org/10.1186/s12942-019-0171-2.

47. Janowicz, K.; Gao, S.; McKenzie, G.; Hu, Y.; Bhaduri, B. GeoAI: Spatially Explicit Artificial Intelligence Techniques for Geographic Knowledge Discovery and Beyond. Int. J. Geogr. Inf. Sci. 2020, 34, 625–636. https://doi.org/10.1080/13658816.2019.1684500.

48. Dong, E.; Du, H.; Gardner, L. An Interactive Web-Based Dashboard to Track COVID-19 in Real Time. Lancet Infect. Dis. 2020, 20, 533–534. https://doi.org/10.1016/S1473-3099(20)30120-1.

49. Boulos, M.N.K.; Geraghty, E.M. Geographical Tracking and Mapping of Coronavirus Disease COVID-19/SARS-CoV-2 Epidemic and Associated Events around the World: How 21st Century GIS Technologies Are Supporting Global Public Health Crisis Management. Int. J. Health Geogr. 2020, 19, 8. https://doi.org/10.1186/s12942-020-00202-8.

50. Keim, D.; Mansmann, F.; Schneidewind, J.; Ziegler, H. Challenges in Visual Data Analysis. In Proceedings of the IEEE International Conference on Information Visualization, Washington, DC, USA, 2006; pp. 9–16.

51. Thomas, J.J.; Cook, K.A., Eds. Illuminating the Path: The Research and Development Agenda for Visual Analytics; IEEE Computer Society: Los Alamitos, CA, USA, 2005.

52. Shneiderman, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In Proceedings of the IEEE Symposium on Visual Languages, Boulder, CO, USA, 1996; pp. 336–343. https://doi.org/10.1109/VL.1996.545307.

53. Ware, C. Information Visualization: Perception for Design, 3rd ed.; Morgan Kaufmann: Waltham, MA, USA, 2012.

54. Hyndman, R.J.; Athanasopoulos, G. Forecasting: Principles and Practice, 3rd ed.; OTexts: Melbourne, Australia, 2021.

55. Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. The M4 Competition: Results, Findings, Conclusion and Way Forward. Int. J. Forecast. 2018, 34, 802–808. https://doi.org/10.1016/j.ijforecast.2018.06.001.

56. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. arXiv 2014, arXiv:1412.6980.

57. Loshchilov, I.; Hutter, F. Decoupled Weight Decay Regularization. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 2019.

58. Huber, P.J. Robust Estimation of a Location Parameter. Ann. Math. Stat. 1964, 35, 73–101. https://doi.org/10.1214/aoms/1177703732.

59. Hameed, K.; Naha, R.; Hameed, F. Digital Transformation for Sustainable Health and Well-Being: A Review and Future Research Directions. Discov. Sustain. 2024, 5, 104. https://doi.org/10.1007/s43621-024-00273-8.

60. Rahat, N. Mapping Sustainability Practices in the Healthcare Sector: A Systematic Literature Review and Future Research Agenda. Int. J. Consum. Stud. 2024, 48, e12997. https://doi.org/10.1111/ijcs.12997.

61. Assessing the Carbon Footprint of Digital Health Interventions: A Scoping Review. J. Am. Med. Inform. Assoc. 2022, 29, 2128–2139. https://doi.org/10.1093/jamia/ocac196.

62. Aligning with the Goals of the Planetary Health Concept Regarding Ecological Sustainability and Digital Health: Scoping Review. J. Med. Internet Res. 2025, 27, e71795.