

Uncertainty Quantification in Demographic and Health Forecasting Using Probabilistic Artificial Intelligence

Ridab Abdalla¹, Altaiyb Mohammed²

¹Sudan University of Science and Technology, College of Science, Statistics Department
Email: rdb_adlan84[at]hotmail.com

²Sudan University of Science and Technology, College of Science, Statistics Department
Email: ibnomer8090[at]gmail.com

Abstract: *Reliable demographic and health forecasting requires accurate predictions and appropriate quantification of uncertainty. This study compares classical statistical models, supervised machine learning, deterministic deep learning, and probabilistic deep learning for forecasting life expectancy across 54 African countries from 2000 to 2023. The dataset contains 1,296 country-year observations and 25 candidate socioeconomic and public health predictors. Models are evaluated over one-, three-, and five-year forecasting horizons using point prediction accuracy, prediction interval coverage, and uncertainty measures. Results indicate that classical ARIMA achieves strong short-term accuracy, with a mean absolute error of 0.652 at the one-year horizon, while XGBoost provides the lowest five-year prediction error of 0.844. Probabilistic deep learning models using Monte Carlo Dropout demonstrate favorable uncertainty calibration, with the MC-Dropout multilayer perceptron achieving 96.3% prediction interval coverage at the five-year horizon. Its estimated epistemic variance share increases from 33.5% to 42.8% between the one- and five-year horizons. These findings highlight the importance of evaluating both predictive accuracy and uncertainty calibration when developing demographic forecasting systems. Probabilistic approaches can complement conventional forecasting models by providing additional information for risk-aware public health planning.*

Keywords: Demographic Forecasting; Life Expectancy; Probabilistic Artificial Intelligence; Bayesian Deep Learning; Uncertainty Quantification; Monte Carlo Dropout

1. Introduction

Demographic forecasting entails estimating future population characteristics, including size, age structure, fertility, mortality, and net migration [1,2]. Healthcare forecasting focuses on projecting future healthcare events, service demand, resource allocations, and population health outcomes using historical, epidemiological, and clinical data[3,4]. Together, these fields support evidence-based decision-making, health system optimization, and long-term socio-economic planning.

Table 1: Key Performance and Monitoring Indicators by Domain

Domain	Key Indicators PDF
Demographics	Population Size, Total Fertility Rate (TFR), Mortality Rate, Life Expectancy, Migration Rate
Healthcare	Incidence Rate, Prevalence Rate, Healthy Life Expectancy (HALE), Healthcare Service Utilization

Changing global age structures and epidemiological dynamics create intricate challenges for both high-income and developing nations, particularly regarding healthcare financing sustainability and service accessibility. While traditional statistical models provided foundational frameworks for population science, their reliance on rigid distributional assumptions and predefined linear relationships limits their applicability in complex, dynamic contexts [2,5].

The integration of Artificial Intelligence (AI), Machine Learning (ML), Deep Learning (DL), and probabilistic modeling has established a new paradigm in predictive health analytics[6–8]. By explicitly representing and propagating uncertainty through probability distributions rather than deterministic single-point metrics, probabilistic AI enables policymakers to evaluate risk profiles and build resilient healthcare delivery systems[9–11].

2. Literature Survey

2.1 Conceptual Framework

Recent developments in demographic and healthcare forecasting have strengthened the role of Bayesian methods and probabilistic models in representing uncertainty. [12] reviewed probabilistic population projection methods across short to long term horizons. The authors highlighted Bayesian approaches to projecting population dynamics and their implications for planning. [13] the main contribution was a neural network-based methodology and forecasting tool designed to provide more detailed age-specific demographic projections. The authors highlighted the limited application of adaptive neural networks in demographic forecasting. [14] introduced a flexible Bayesian approach for modeling shocks in demographic and health indicators such as life expectancy. The method is demonstrated in a model for male period life expectancy; it finds that including shocks in the model results in sharper uncertainty intervals without sacrificing empirical coverage or

prediction error. [15] used Bayesian optimization and Deep learning methods to predict the number of confirmed COVID-19 cases in Saudi Arabia. [16] developed a hybrid model that combines point forecasts from domain-specific models with probabilistic distributions from statistical and machine learning approaches, enabling human input to fine-tune and enhance the system-generated forecasts. [17] applied Bayesian neural network for lung-cancer risk stratification from chest CT, combining imaging and clinical metadata. [18] provided a comprehensive analysis of current Uncertainty Quantification UQ in healthcare, offering an informed framework that highlights how different methods can be integrated into each stage of the ML (processing training and evaluation). [19] developed a novel statistical approach to evaluate the accuracy of probabilistic population forecasts for births, deaths, and net migration. The authors propose using deviance as a scoring rule to measure the fit between predictive distributions and realized values. [20] introduced a new method for the probabilistic projection of subnational population. Small-area projections can support local service planning and could provide demographic inputs for regional health forecasting.

Overall, the literature suggests substantial methodological overlap between demographic and health forecasting, especially in Bayesian modeling and probabilistic prediction

2.2 problem definition

Traditional deterministic approaches often present a single point estimator or a limited set of scenarios, potentially failing to capture the full range of uncertainty surrounding future outcomes. Probabilistic forecasting represents future values as probability distributions and prediction intervals, providing a more comprehensive account of possible outcomes.

2.3 Objective definition

This study compares conventional statistical, machine-learning, deep-learning, and Bayesian deep-learning models to forecast life expectancy in African countries over one-, three-, and five-year horizons. Specifically, it evaluates point-forecasting accuracy, the calibration and sharpness of probabilistic intervals, and the extent to which epistemic uncertainty contributes to overall predictive uncertainty across these time horizons.

3. Data Source and Missing Data

The study analyzed annual data from 54 African countries (2000-2023) using life expectancy at birth as a proxy and 25 demographic, health, nutrition, environmental, health system and socio-economic predictors as inputs. The missing predictor values were addressed by multiple imputation using training data at the source of each forecast in order to avoid the leakage of information. The performance of the models over one-, three-, and five-year time horizons was assessed using point-probability and probabilistic metrics, with Monte Carlo decay used to quantify the predictive uncertainty and time validation

used to ensure the robustness and reproducibility of the methodology.

4. Methodology

4.1 Traditional Forecasting Method

Traditional demographic and health forecasting methods prioritize theoretical interpretability and mathematical simplicity [5,21]. These methodologies fall into several major categories:

- 1) **Trend Extrapolation Methods:** Extrapolate past temporal patterns directly into the future under the assumption that historical trends will persist unchanged. Common models include linear trend models, exponential growth models, and polynomial regressions.
- 2) **Cohort Component Method:** The standard demographic projection model that tracks population changes across age and sex cohorts by evaluating three structural drivers: fertility, mortality, and migration.
- 3) **Time Series Analysis:** Models sequential temporal data where future outcomes depend on historical observations. Typical implementations include Moving Average (MA), Autoregressive (AR), Autoregressive Integrated Moving Average (ARIMA), Exponential Smoothing, and Holt-Winters methods.
- 4) **Life Table Methods:** Systematic statistical methods summarizing mortality patterns, survival probabilities, and life expectancies across demographic cohorts to support public health planning and survival analysis.
- 5) **Regression Forecasting:** Parametric models (e.g., Linear, Multiple, Poisson, and Logistic regressions) designed to model mathematical relationships between continuous or categorical response variables and predictor covariates.
- 6) **Epidemiological Models:** Differential-equation-based compartmental models designed to simulate disease transmission dynamics:
 - **SIR Model:** Divides population N into Susceptible (S), Infected (I), and Recovered (R) groups.
 - **SEIR Model:** Incorporates an Exposed (E) latent phase group before infectiousness.
- 7) **Delphi Method & Scenario-Based Forecasting:** Expert-driven consensus techniques and multi-scenario analyses (e.g., high, medium, and low growth trajectories) designed to evaluate qualitative or policy-driven future projections.

4.2 Artificial Intelligence and Machine Learning Approaches

4.2.1 Supervised Machine Learning (SML)

Supervised machine learning algorithms learn a predictive mapping function f that transforms an input feature vector $x \in \mathbb{R}^D$ into a target outcome label $y \in Y$ using a labeled training dataset $D = \{(x_i, y_i)\}_{i=1}^N$. Mathematically, [22–26] SML can be formulated as:

$$y = f(x_i, \theta) + \varepsilon \quad (1)$$

Where θ represents the model parameters estimated during optimization and ε represents unexplained random error.

SML encompasses two core predictive tasks:

- **Classification:** Maps input features to discrete categorical classes (e.g., binary disease diagnosis, risk stratification, migration pattern grouping).
- **Regression:** Predicts continuous output metrics (e.g., mortality rates, future life expectancy, hospital admission counts, healthcare expenditures).

Common SML models include Linear and Logistic Regressions, Decision Trees, Support Vector Machines (SVM), Random Forests, K-Nearest Neighbors (K-NN), Naïve Bayes, Gradient Boosting (e.g., XGBoost, LightGBM, CatBoost), and Gaussian Process Regressions [27,28].

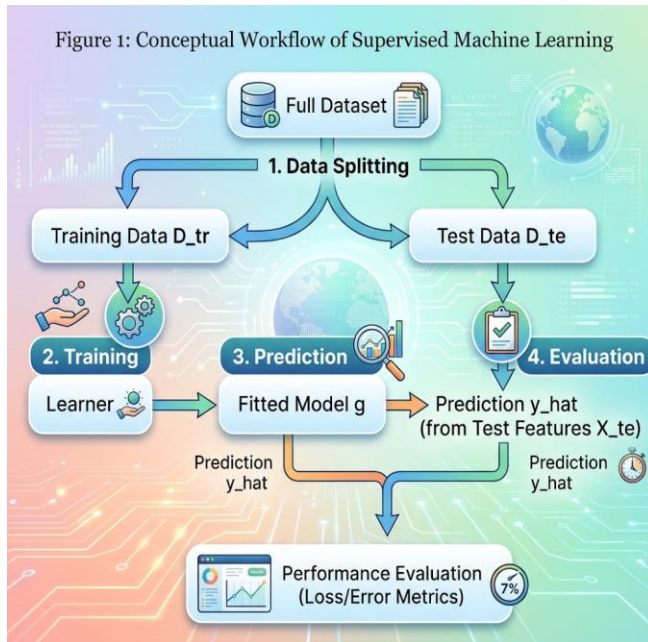


Figure 1: Conceptual Workflow of Supervised Machine Learning
Generation by AI

4.2.2 Unsupervised Machine Learning (UL)

Unsupervised learning analyzes unlabeled datasets to discover underlying structural distributions, hidden groupings, or feature associations without pre-assigned target responses [29]. Key techniques include:

- **Clustering:** Groups data points based on spatial or statistical similarity measures (e.g., K-Means, Hierarchical Clustering, DBSCAN, Gaussian Mixture Models) for patient outcome stratification.
- **Association Rule Learning:** Identifies co-occurrence patterns among variables using algorithms like Apriori, FP-Growth, and ECLAT.
- **Anomaly Detection:** Identifies atypical observations that deviate from normative structural profiles (e.g., Isolation Forest, One-Class SVM) to detect epidemiological outbreaks or fraudulent utilization.
- **Dimensionality Reduction:** Compresses high-dimensional feature spaces into lower-dimensional representations while retaining maximal variance (e.g., Principal Component Analysis [PCA], Autoencoders, t-SNE, UMAP).

4.2.3 Semi-Supervised & Reinforcement Learning

- **Semi-Supervised Learning (SSL):** Integrates small sets of labeled data with abundant unlabeled records. SSL uses generative models $p(x, y) = p(y)p(x|y)$ or deep pseudo-labeling architectures to reduce dependence on costly manual annotations in clinical and demographic datasets.
- **Reinforcement Learning (RL):** Formalizes sequential decision-making under dynamic environmental conditions. An autonomous agent interacts with an environment by selecting actions a_t at state s_t , receiving scalar feedback rewards r_t to maximize long-term cumulative utility [30].

4.3. Deep Learning and Probabilistic Deep Learning Architecture

4.3.1 Neural Network Fundamentals

Deep Learning (DL) utilizes multi-layered Artificial Neural Networks (ANNs) to automatically extract hierarchical feature representations from high-dimensional, unstructured data [6,31,33].

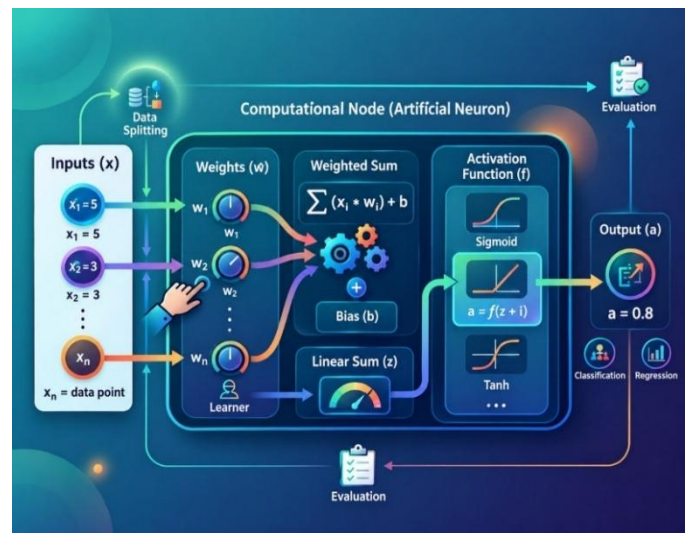


Figure 2: Computational Node in an Artificial Neural Network
Generation by AI

A single computational unit (perceptron) computes a weighted linear combination of input features, adds a scalar bias parameter b , and transforms the pre-activation state via a non-linear activation function $f(z)$:

$$z = \sum_{j=1}^n x_j w_j + b \Rightarrow a = f(z) \quad (2)$$

Common activation functions include (Dubey et al., 2022; Heaton, 2018):

- **Linear Identity:** $f(z) = z$ (used in continuous output layers)
- **Sigmoid:** $f(z) = \frac{1}{1+e^{-z}}$ (maps outputs to (0, 1) range)
- **Hyperbolic Tangent (Tanh):** $f(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$ (maps outputs to (-1, 1) range)
- **Rectified Linear Unit (ReLU):** $f(z) = \max(0, z)$ (mitigates vanishing gradient problems)

- **Softmax:** $f_k(z) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}}$ (generates multi-class probability distributions)

The network optimizes parameter weights W and biases b by executing backpropagation via stochastic gradient descent algorithms to minimize a designated loss function $L(y, \hat{y})$:

$$\text{Regression (MSE): } L(y, \hat{y}) = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (3)$$

$$\text{Classification (cross-entropy): } L(y, \hat{y}_i) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^K y_{ij} \log(\hat{y}_{ik}) \quad (4)$$

4.3.2 Major Deep Learning Architecture

- 1) **Multilayer Perceptron (MLP):** Standard feed-forward networks composed of fully connected linear layers interleaved with non-linear activation functions, suited for tabular demographic and clinical predictor variables.
- 2) **Convolutional Neural Networks (CNN):** Architectures leveraging spatial convolution operators to extract local contextual representations. For a 1D sequence, the convolution output z_i and activation h_i are defined as:

$$z_i = \sum_{k=0}^{k-1} w_k x_{i+k} + b \Rightarrow h_i = f(z_i) \quad (5)$$

- 3) **Recurrent Neural Networks (RNN) & LSTM:** Designed to handle sequential temporal trajectories by maintaining a hidden state vector h_t that carries contextual memory across time steps:

$$h_t = \phi(W_{xh}x_t + W_{hh}h_{t-1} + b_h) \quad (6)$$

$$\hat{y}_t = g(W_{hy}h_t + b_y) \quad (7)$$

- 4) **Long Short-Term Memory (LSTM)** variants overcome vanishing/exploding gradient constraints, making them effective for long-term health trajectory and population forecasting [33,34].

4.4. Modeling Uncertainty in Demographic and Health Systems

4.4.1 Taxonomies of Uncertainty

Uncertainty refers to incomplete knowledge regarding physical mechanisms, system behaviors, or future dynamic states [9,23,35]. It is broadly classified into two categories:

- 1) **Aleatoric Uncertainty (Stochastic / Irreducible):** Inherent variance and randomness observed within data generation processes. It cannot be eliminated by adding more training data.
- 2) **Epistemic Uncertainty (Systematic / Reducible):** Uncertainty stemming from inadequate model knowledge, sparse observational data, or model misspecifications. Epistemic uncertainty can be reduced by collecting more data or expanding model capacity

4.4.2 Mathematical Formulation of System Uncertainty Propagation

Let a computational model U predict system output u given input feature vector x and internal parameter set α_u , subject to structural model error ε_u : [11,35]

$$u = U(x; \alpha_u) + \varepsilon_u \quad (8)$$

When evaluated within a validation domain against observational system data y , the unified validation response model v is represented as:

$$v = V(u, y; \alpha_v) + \varepsilon_v \quad (9)$$

The total combined set of system uncertainties U_{total} is expressed as the union of four structural components:

$$U_{total} = U_{param} \cup U_{input} \cup U_{mod} \cup U_{obs}$$

Where:

- U_{param} : Uncertainty in parameter estimation.
- U_{input} : Uncertainty in primary input feature values.
- U_{mod} : Inadequacy or structural approximations in the mathematical model.
- U_{obs} : Measurement and observational instrument errors.

Uncertainty propagation translates these cumulative input uncertainties into output predictive distributions:

$$\{U_{param} \cup U_{input} \cup U_{mod} \cup U_{obs}\} \rightarrow U_{pred}$$

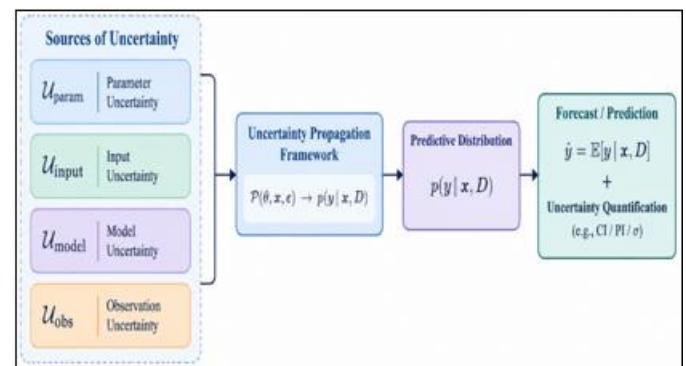


Figure 3: Systemic Propagation of Uncertainties to Predictive Outputs.

Generation by AI

4.4.3 Uncertainty Quantification (UQ) Framework

Uncertainty Quantification (UQ) provides a mathematical framework for characterizing, propagating, and reducing uncertainties in scientific models. Rather than evaluating models strictly by single-point accuracy, UQ quantifies prediction interval coverage and distributional bounds [9,10,14].

A standard UQ workflow follows four sequential steps:

- 1) **Identification:** Isolating all internal and external sources of uncertainty within the system.
- 2) **Characterization:** Mathematically parameterizing uncertainties using empirical probability density functions or interval bounds.
- 3) **Propagation:** Forward-propagating input uncertainties through the computational pipeline to the output metrics.
- 4) **Impact Analysis:** Assessing output sensitivity to refine decisions, prioritize data collection, and support robust policy formulation under uncertainty.

5. Empirical Analysis and Results

5.1 Analytical Panel Dataset and Data Quality

The empirical evaluation uses a pan-African panel dataset covering 54 countries over 24 years (2000–2023), producing $N = 1,296$ country-year observations across 25 candidate public health and socio-economic predictors. Data verification confirmed zero duplicate records and zero missing observations in the primary target variable, life expectancy at birth.

Table 2: Summary and Quality Characteristics of the Pan-African Panel Dataset

Characteristic	Result
African Countries	54
Study Period	2000–2023 (24 years)
Country-Year Observations	1,296
Candidate Predictors	25
Development Subsample (2000–2018)	1,026 observations
Final-Test Subsample (2019–2023)	270 observations
Mean Predictor Availability	95.35%

Descriptive analysis indicates substantial cross-regional and income-based heterogeneity in life expectancy across Africa. The overall mean life expectancy is 60.23 years ($SD = 7.67$). Subregionally, Northern Africa exhibits the highest mean (70.52 years), whereas Southern Africa exhibits the lowest (55.83 years). By income group, High-Income African nations average 73.39 years compared to 57.08 years in Low-Income nations.

Bivariate correlation analysis identified strong structural drivers of life expectancy: Young-age dependency ($r = -0.671$, $p < 0.001$), Basic sanitation access ($r = 0.657$, $p < 0.001$), Stunting under five ($r = -0.650$, $p < 0.001$), Adolescent fertility ($r = -0.645$, $p < 0.001$), and Basic drinking water access ($r = 0.638$, $p < 0.001$).

Table 3: Predictive Point Accuracy and Calibration Metrics Across Model Families and Forecast Horizons (1, 3, and 5 Years)

Forecast Horizon	Model Family	Best Point Model	Locked Final MAE	Best Distributional / Interval Model	PICP95	MPIW95
1 Year (H1)	Classical	ARIMA	0.652	ARIMA	90.7%	2.086
	Machine Learning	SVR-RBF	0.670	SVR-RBF / Elastic Net	94.4%	3.428
	Deep Learning	Deep MLP	1.202	GRU	96.3%	3.905
	Bayesian DL	MC-Dropout LSTM	0.884	MC-Dropout MLP	98.1%	6.300
3 Years (H3)	Classical	Naive	0.966	Naive	88.9%	6.816
	Machine Learning	Elastic Net	1.322	Elastic Net	92.6%	6.012
	Deep Learning	Deep MLP	1.678	Deep MLP / LSTM	92.6%	8.430
	Bayesian DL	MC-Dropout GRU	1.583	MC-Dropout MLP	85.2%	6.389
5 Years (H5)	Classical	ETS	1.043	ARIMA	85.2%	5.327
	Machine Learning	XGBoost	0.844	XGBoost	100.0%	10.673
	Deep Learning	Deep MLP	1.487	Deep MLP / GRU / LSTM	100.0%	13.520
	Bayesian DL	MC-Dropout MLP	1.114	MC-Dropout MLP	96.3%	6.094

In rolling one-year evaluations across the out-of-sample period (2019–2023), MC-Dropout MLP achieved the best continuous ranked probability score and interval sharpness, attaining the lowest Winkler score ($Winkler95 = 11.514$) with 97.4% empirical coverage.

5.2 Empirical Performance Across Model Families

Models were evaluated across three horizons: Horizon 1 (H1: 1 year ahead), Horizon 3 (H3: 3 years ahead), and Horizon 5 (H5: 5 years ahead).

Classical Baseline Models

Classical models demonstrated strong short-term point accuracy but suffered severe prediction interval degradation over longer horizons. At H1, ARIMA achieved $MAE = 0.652$ and $RMSE = 2.372$. However, at H3, ARIMA's 95% Prediction Interval Coverage Probability (PICP95) collapsed to 25.9% ($MPIW95 = 3.688$), and Exponential Smoothing (ETS) dropped to 22.2%.

Machine Learning Models

Supervised ML models demonstrated robust point estimation. SVR-RBF achieved $MAE = 0.670$ at H1. At H5, XGBoost achieved the best overall point performance across all model families ($MAE = 0.844$, $RMSE = 1.200$). Conformal prediction intervals for ML models yielded 100.0% PICP95 coverage at H5, though accompanied by wide interval widths ($MPIW95 = 10.673$ for XGBoost).

Deep Learning Models

Deterministic DL models showed moderate point performance. Deep MLP achieved $MAE = 1.202$ at H1, 1.678 at H3, and 1.487 at H5. Sequential models (GRU and LSTM) performed competitively at H1 ($MAE = 1.228$ and 1.303, respectively).

Probabilistic / Bayesian Deep Learning Models

Bayesian DL architectures utilizing Monte Carlo (MC) Dropout maintained high distributional calibration and narrow prediction intervals. MC-Dropout LSTM achieved $MAE = 0.884$ ($CRPS = 0.738$) at H1. At H5, MC-Dropout MLP achieved $MAE = 1.114$, $CRPS = 0.823$, and $PICP95 = 96.3\%$ with a compact mean interval width ($MPIW95 = 6.094$).

5.3 Empirical Predictive Uncertainty Decomposition

Variance decomposition separating total predictive uncertainty into epistemic (model parameter) and aleatoric (data noise) components revealed a consistent structural shift across temporal horizons:

Table 4: Empirical Variance Decomposition of Epistemic and Aleatoric Uncertainty Across Forecast Horizons

Horizon	Model Architecture	Epistemic SD (σ_{epi})	Aleatoric SD (σ_{ale})	Total SD (σ_{tot})	Epistemic Variance Share (%)
1 Year (H1)	MC-Dropout MLP	0.906	1.262	1.573	33.5%
	MC-Dropout GRU	0.503	0.841	1.003	33.5%
	MC-Dropout LSTM	0.534	0.786	0.960	33.6%
3 Years (H3)	MC-Dropout MLP	0.958	1.233	1.584	35.4%
	MC-Dropout GRU	0.526	0.706	0.888	37.6%
	MC-Dropout LSTM	0.556	0.794	0.979	34.2%
5 Years (H5)	MC-Dropout MLP	0.983	1.121	1.505	42.8%
	MC-Dropout GRU	0.562	0.777	0.969	36.8%
	MC-Dropout LSTM	0.570	0.823	1.015	35.0%

5.4 Discussion

5.4.1 Synthesis of Empirical Findings: Accuracy vs. Calibration

Our empirical analysis across 54 African nations provides key insights into the trade-offs between point prediction accuracy and uncertainty calibration in public health planning. A critical finding is the **Point-Calibration Paradox**: models optimized solely for point accuracy (such as classical ARIMA or non-parametric XGBoost) often produce unreliable uncertainty bounds over multi-year horizons.

- Short-Horizon Dynamics (1-Year):** Classical ARIMA (MAE = 0.652) and Support Vector Regression (MAE = 0.670) outperform complex deep neural network architectures. In stable, short-term settings, linear temporal autoregression effectively captures local continuity.
- Medium-to-Long Horizon Failure of Classical Models (3–5 Years):** As the forecast horizon extends to 3 and 5 years, classical prediction intervals collapse. ARIMA's 95% interval coverage drops to 25.9% at Horizon 3, reflecting severe overconfidence. Classical parametric formulas fail to account for cumulative parameter drift and non-linear structural transitions in macro-health indicators.
- Long-Term Superiority of Gradient Boosting:** At Horizon 5, XGBoost achieves the lowest point prediction error (MAE = 0.844), demonstrating the capacity of tree ensembles to capture non-linear socio-economic predictors (e.g., sanitation, dependency ratios, stunting) over longer spans.
- Distributional Superiority of Probabilistic AI:** While deterministic deep learning architectures (MLP, GRU, LSTM) struggle with point metrics relative to XGBoost, probabilistic variants using Monte Carlo Dropout provide well-calibrated prediction intervals. MC-Dropout MLP achieves near-optimal interval coverage at Horizon 5 (PICP95 = 96.3%) with narrow interval widths (MPIW95 = 6.094) and the lowest rolling Winkler score (11.514).

5.4.2 Epistemic vs. Aleatoric Uncertainty Dynamics

The mathematical decomposition of variance in Section 5.3 (table 4) clarifies how uncertainty evolves over time. Aleatoric uncertainty (data noise, measurement errors in health registries) remains relatively stable across horizons $\sigma_{ale} \approx 0.78 - 1.26$. In contrast, epistemic uncertainty (model parameter ignorance) increases noticeably over longer forecast horizons.

For MC-Dropout MLP, the epistemic variance share expands from 33.5% at Horizon 1 to **42.8% at Horizon 5**. This increase highlights that long-term demographic forecasting risk is primarily driven by model parameter uncertainty under shifting macroeconomic conditions rather than baseline measurement noise. Deterministic models ignore this epistemic expansion, resulting in overconfident predictions. Probabilistic AI explicitly captures epistemic variance through weight distribution sampling, preventing misleading forecasts in public health planning.

5.4.3 Methodological Implications for Public Health Policy

For health ministry planners, resource allocators, and international agencies (e.g., WHO, United Nations Population Division), relying on single-point forecasts introduces systemic risk. A deterministic forecast predicting a single life expectancy value can lead to misallocated hospital infrastructure, inaccurate pension fund estimates, and inadequate epidemic preparedness if underlying socio-economic drivers fluctuate.

Probabilistic AI addresses these challenges by transforming health forecasting from point estimation into distributional risk evaluation.

- Optimized Capacity Planning:** Probabilistic distributions allow policy-makers to plan resource allocations against conservative lower-bound estimates (e.g., 5th percentile outcomes).
- Targeted Interventions:** High epistemic uncertainty identifies specific subregions or indicators requiring improved data collection and monitoring.
- Policy Resilience:** Evaluating full prediction intervals enables health systems to design flexible policies resilient to unexpected socio-economic shocks.

5.4.4 Comparative Analysis with Existing Literature

Our results align with recent methodological literature emphasizing uncertainty quantification in deep learning. While classical studies highlighted the interpretability of ETS and ARIMA, our findings demonstrate their vulnerability to prediction interval degradation in non-linear settings. Furthermore, our empirical results support recent findings by the many of studies, demonstrating that Bayesian approximations (such as MC Dropout) effectively isolate epistemic uncertainty and improve predictive reliability in dynamic domain applications.

6. Conclusion

This study compared classical statistical forecasting, supervised machine learning, deterministic deep learning, and probabilistic deep learning for life expectancy forecasting across 54 African countries. The findings demonstrate that different forecasting approaches offer distinct advantages depending on the prediction horizon and evaluation criterion. Classical ARIMA achieved strong short-term predictive accuracy, while XGBoost produced the lowest point prediction error at the five-year horizon. Probabilistic deep learning using Monte Carlo Dropout provided an additional advantage through uncertainty estimation and favorable prediction interval calibration. The increasing epistemic variance share observed for the MC-Dropout multilayer perceptron suggests that model uncertainty becomes more influential over longer forecasting horizons. These results emphasize the importance of considering both point accuracy and predictive uncertainty when evaluating demographic forecasting systems. Integrating probabilistic uncertainty estimates with accurate forecasting methods may support more informed resource allocation, improved risk assessment, and resilient public health planning. Future research could extend this evaluation to additional health indicators and assess the robustness of probabilistic models under demographic and socioeconomic changes.

References

- [1] United Nations, Department of Economic and Social Affairs, Population Division. World Population Prospects: Methodology of the United Nations population estimates and projections, 2024. UN DESA/POP/2024/DC/NO. 10. United Nations, New York.
https://population.un.org/wpp/assets/Files/WPP2024_Methodology-Report_Final.pdf.
- [2] Vanella P, Deschermeier P, Wilke CB. An Overview of Population Projections—Methodological Concepts, International Data Availability, and Use Cases. *Forecasting*. 2020;2(3):346-363. doi:10.3390/forecast2030019
- [3] Celi LA, Majumder MS, Ordóñez P, Osorio JS, Paik KE, Somai M, eds. *Leveraging Data Science for Global Health*. Springer International Publishing; 2020. doi:10.1007/978-3-030-47994-7
- [4] World health statistics 2024: monitoring health for the SDGs, sustainable development goals. Accessed October 15, 2026.
<https://www.who.int/publications/i/item/9789240094703>
- [5] Hyndman, R. J., & Athanasopoulos, G. *Forecasting: Principles and Practice* (3rd ed.). OTexts, Melbourne, Australia, 2021. <https://otexts.com/fpp3/>.
- [6] Bishop CM, Bishop H. *Deep Learning: Foundations and Concepts*. Springer International Publishing; 2024. doi:10.1007/978-3-031-45468-4
- [7] Heaton J. Ian Goodfellow, Yoshua Bengio, and Aaron Courville: *Deep learning: The MIT Press*, 2016, 800 pp, ISBN: 0262035618. *Genet Program Evolvable Mach*. 2018;19(1-2):305-307. doi:10.1007/s10710-017-9314-z
- [8] Murphy KP. *Probabilistic Machine Learning: An Introduction*. The MIT Press; 2022.
- [9] Hüllermeier E, Waegeman W. Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach Learn*. 2021;110(3):457-506. doi:10.1007/s10994-021-05946-3
- [10] Ståhl N, Falkman G, Karlsson A, Mathiason G. Evaluation of Uncertainty Quantification in Deep Learning. In: Lesot MJ, Vieira S, Reformat MZ, et al., eds. *Information Processing and Management of Uncertainty in Knowledge-Based Systems*. Springer International Publishing; 2020:556-568. doi:10.1007/978-3-030-50146-4_41
- [11] Tyralis H, Papacharalampous G. A review of predictive uncertainty estimation with machine learning. *Artif Intell Rev*. 2024;57(4):94. doi:10.1007/s10462-023-10698-8
- [12] Raftery AE, Ševčíková H. Probabilistic population forecasting: Short to very long-term. *International Journal of Forecasting*. 2023;39(1):73-97. doi:10.1016/j.ijforecast.2021.09.001
- [13] Mezhev, S., & Krayushkin, M. Artificial intelligence in forecasting demographic processes. Proceedings of the 12th International Conference on Applied Innovations in IT, 2024, 12(1), 181–187. <https://doi.org/10.25673/115696>.
- [14] Susmann H, Alkema L. Flexibly Modeling Shocks to Demographic and Health Indicators with Bayesian Shrinkage Priors. *arXiv*. Preprint posted online October 11, 2024;arXiv:2410.09217. doi:10.48550/arXiv.2410.09217
- [15] Alhhazmi A, Alferidi A, Almutawif YA, Makhdoom H, Albasri HM, Sami BS. Artificial intelligence in healthcare: combining deep learning and Bayesian optimization to forecast COVID-19 confirmed cases. *Front Artif Intell*. 2024;6:1327355. doi:10.3389/frai.2023.1327355
- [16] Hewage HC, Rostami-Tabar B, Syntetos A, Liberatore F, Milano GR. A novel hybrid approach to contraceptive demand forecasting: integrating point predictions with probabilistic distributions. *International Journal of Production Research*. Published online July 17, 2025:1-35. doi:10.1080/00207543.2025.2526169
- [17] Amasiadi N, Aslani-Gkatzamanidou M, Theodorakopoulos L, et al. AI-Driven Bayesian Deep Learning for Lung Cancer Prediction: Precision Decision Support in Big Data Health Informatics. *BioMedInformatics*. 2025;5(3):39. doi:10.3390/biomedinformatics5030039
- [18] López, L. J. L., Elsharief, S., Al Jorf, D., Darwish, F., Ma, C., & Shamout, F. E. Uncertainty quantification for machine learning in healthcare: A survey. Proceedings of Machine Learning Research, 2025, 287, 862–907. <https://proceedings.mlr.press/v287/lopez25a.html>.
- [19] Wilson WW, Haiges R, Christe K. Evaluating probabilistic population forecasts. *SSRN*. Preprint posted online 2023. doi:10.2139/ssrn.4417034
- [20] Goes, J., & Engelhardt, H. Probabilistic population forecasts for small regions. *Demographic Research*, 2026, 54, 719–762. <https://doi.org/10.4054/DemRes.2026.54.2>

- [21] Ahlbom A. *Biostatistics for Epidemiologists*. CRC Press; 2017.
- [22] Berry MW, Mohamed A, Yap BW, eds. *Supervised and Unsupervised Learning for Data Science*. Springer International Publishing; 2020. doi:10.1007/978-3-030-22475-2
- [23] Bammer G, Smithson M. *Uncertainty and Risk: Multidisciplinary Perspectives*. Routledge; 2012.
- [24] Hastie T, Friedman J, Tibshirani R. *The Elements of Statistical Learning*. Springer New York; 2001. doi:10.1007/978-0-387-21606-5
- [25] Jung A. Machine Learning: The Basics. *arXiv*. Preprint posted online January 29, 2022;arXiv:1805.05052. doi:10.48550/arXiv.1805.05052
- [26] Sohil F, Sohali MU, Shabbir J. An introduction to statistical learning with applications in R: by Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani, New York, Springer Science and Business Media, 2013, \$41.98, eISBN: 978-1-4614-7137-7. *Statistical Theory and Related Fields*. 2022;6(1):87-87. doi:10.1080/24754269.2021.1980261
- [27] James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. Vol 103. Springer New York; 2013. doi:10.1007/978-1-4614-7138-7
- [28] Raschka S, Liu Y (Hayden), Mirjalili V. *Machine Learning with PyTorch and Scikit-Learn: Develop Machine Learning and Deep Learning Models with Python*. Packt Publishing Ltd; 2022.
- [29] Naeem S, Ali A, Anam S, Ahmed MM. An Unsupervised Machine Learning Algorithms: Comprehensive Review. *IJCDS*. 2023;13(1):911-921. doi:10.12785/ijcds/130172
- [30] Russell SJ, Norvig P. *Artificial Intelligence: A Modern Approach*. Prentice Hall; 1995.
- [31] Bisong E. *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress; 2019. doi:10.1007/978-1-4842-4470-8
- [32] Kelleher JD. *Deep Learning*. The MIT Press, 2019. <https://doi.org/10.7551/mitpress/11171.001.0001>.
- [33] Kim KG. Book Review: Deep Learning. *Healthc Inform Res*. 2016;22(4):351. doi:10.4258/hir.2016.22.4.351
- [34] Chollet F. *Deep Learning with Python*. Second edition. Manning Publications; 2021.
- [35] Sullivan TJ. *Introduction to Uncertainty Quantification*. Springer; 2015.