**WHO MMR metric calculation:**

[**https://iris.who.int/bitstream/handle/10665/381012/9789240108462-eng.pdf?sequence=1**](https://iris.who.int/bitstream/handle/10665/381012/9789240108462-eng.pdf?sequence=1)

**Evaluation metrics:**

<https://www.ncbi.nlm.nih.gov/books/NBK583962/>

<https://www.mdpi.com/1996-1944/17/20/5127>

<https://www.mdpi.com/2076-3417/15/7/3636>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC6902303/#sec1-7>

<https://www.jmlr.org/papers/volume7/demsar06a/demsar06a.pdf>

<https://www.nature.com/articles/s41598-024-56706-x#Sec10>

<https://arxiv.org/pdf/2304.03990>

<https://www.mdpi.com/2571-9394/5/1/14>

<https://pmc.ncbi.nlm.nih.gov/articles/PMC9949554/#abstract1>

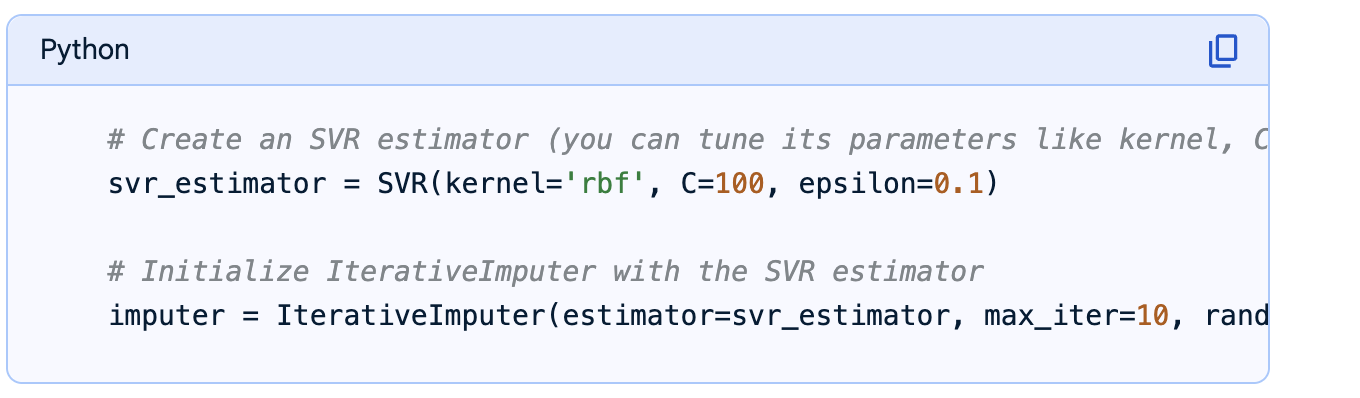
**Machine Learning and imputation software**

**Scikit learn imputation**

[**https://scikit-learn.org/stable/modules/impute.html**](https://scikit-learn.org/stable/modules/impute.html)

**Scikit learn iterative imputation**

<https://scikit-learn.org/stable/auto_examples/impute/plot_iterative_imputer_variants_comparison.html#sphx-glr-auto-examples-impute-plot-iterative-imputer-variants-comparison-py>

**Miss Forest**

[**https://academic.oup.com/bioinformatics/article/28/1/112/219101**](https://academic.oup.com/bioinformatics/article/28/1/112/219101)

**XGBoost**

[**https://xgboost.readthedocs.io/en/stable/tutorials/index.html**](https://xgboost.readthedocs.io/en/stable/tutorials/index.html)

**Lightgbm**

<https://lightgbm.readthedocs.io/en/latest/GPU-Tutorial.html>

**Background research:**

On the consistency of supervised learning with missing values

<https://link.springer.com/article/10.1007/s00362-024-01550-4>

Imputation of missing values in the INFORM Global Risk Index

<https://drmkc.jrc.ec.europa.eu/inform-index/Portals/0/InfoRM/Publications/EUR%2030037%20-%20Imputation%20of%20missing%20values%20in%20the%20INFORM%20Global%20Risk%20Index%20-%20print.pdf>

# **Application of the random forest algorithm to predict skilled birth attendance and identify determinants among reproductive-age women in 27 Sub-Saharan African countries; machine learning analysis**

<https://bmcpublichealth.biomedcentral.com/articles/10.1186/s12889-025-22007-9#Sec2>

**Advanced Machine Learning Techniques for Predictive Analysis of Health Insurance**

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=10695988>

**Machine learning in medicine: a practical introduction to techniques for data pre-processing, hyperparameter tuning, and model comparison**

[**https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01758-8#Sec4**](https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01758-8#Sec4)

**Machine learning for predicting maternal health: Predicting delivery location in a community health worker program in Zanzibar**

[**https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2022.855236/full**](https://www.frontiersin.org/journals/digital-health/articles/10.3389/fdgth.2022.855236/full)

# **Ensemble machine learning framework for predicting maternal health risk during pregnancy**

<https://www.nature.com/articles/s41598-024-71934-x#Abs1>

**Predicting maternal risk level using machine learning models**

<https://pmc.ncbi.nlm.nih.gov/articles/PMC11657143/#Sec5>

**A systematic review of machine learning algorithms for breast cancer detection**

<https://pubmed.ncbi.nlm.nih.gov/40300307/>

# **A Machine Learning Approach for Predicting Maternal Health Risks in Lower-Middle-Income Countries Using Sparse Data and Vital Signs**

<https://www.mdpi.com/1999-5903/17/5/190>

# **Machine learning-enabled maternal risk assessment for women with pre-eclampsia (the PIERS-ML model): a modelling study**

<https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00267-4/fulltext>

**The proportion of missing data should not be used to guide decisions on multiple imputation**

<https://www.jclinepi.com/article/S0895-4356(18)30871-0/pdf>

**Evaluation of multiple imputation with large proportions of missing data: how much is too much?**

<https://pubmed.ncbi.nlm.nih.gov/34568175/>

**Prediction of maternal health risk factors using machine learning algorithms**

[**https://www.sciencedirect.com/science/article/pii/S1877050925016369**](https://www.sciencedirect.com/science/article/pii/S1877050925016369)

**A Robust Machine Learning Predictive Model for Maternal Health Risk**

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9885515>

# **Use of Machine Learning for Early Detection of Maternal Cardiovascular Conditions: Retrospective Study Using Electronic Health Record Data**

<https://cardio.jmir.org/2024/1/e53091/>

**A hybrid boosting ensemble model for predicting maternal mortality and sustaining reproductive**

<https://www.sciencedirect.com/science/article/pii/S2352648322000599>

# **A Machine Learning-Based Risk Prediction Model During Pregnancy in Low-Resource Settings**

<https://www.mdpi.com/2673-9992/25/1/13>

**Child and maternal mortality risk factor analysis using machine learning approaches**

[**https://ieeexplore.ieee.org/document/10131826**](https://ieeexplore.ieee.org/document/10131826)

**Evaluating Advances in Machine Learning Algorithms for Predicting and Preventing Maternal and Foetal Mortality in Nigerian Healthcare: A Systematic Approach**

[**https://hal.science/hal-04994840/document**](https://hal.science/hal-04994840/document)

**Analysis of factors influencing maternal mortality and newborn health – a machine learning approach**

[**https://jmai.amegroups.org/article/view/8590/pdf**](https://jmai.amegroups.org/article/view/8590/pdf)

**Application of random forest model to predict the demand of essential medicines for non-communicable diseases management in public health facilities**

[**https://pmc.ncbi.nlm.nih.gov/articles/PMC9379432/#S2**](https://pmc.ncbi.nlm.nih.gov/articles/PMC9379432/#S2)

# **Prediction of adverse pregnancy outcomes using machine learning techniques: evidence from analysis of electronic medical records data in Rwanda**

[**https://link.springer.com/article/10.1186/s12911-025-02921-z?fromPaywallRec=false**](https://link.springer.com/article/10.1186/s12911-025-02921-z?fromPaywallRec=false)

**Health indicator recording in UK primary care electronic health records: key implications for handling missing data**

[**https://www.dovepress.com/health-indicator-recording-in-uk-primary-care-electronic-health-record-peer-reviewed-fulltext-article-CLEP**](https://www.dovepress.com/health-indicator-recording-in-uk-primary-care-electronic-health-record-peer-reviewed-fulltext-article-CLEP)

# **Data-driven methods for imputing national-level incidence in global burden of disease studies**

[**https://pmc.ncbi.nlm.nih.gov/articles/PMC4431555/**](https://pmc.ncbi.nlm.nih.gov/articles/PMC4431555/)

# **Addressing missing values in routine health information system data: an evaluation of imputation methods using data from the Democratic Republic of the Congo during the COVID-19 pandemic**

<https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-021-00274-z>

# **Enhancing Health and Public Health through Machine Learning: Decision Support for Smarter Choices**

<https://www.mdpi.com/2306-5354/10/7/792>

Use of artificial intelligence for public health surveillance: a case study to develop a machine Learning-algorithm to estimate the incidence of diabetes mellitus in France

<https://link.springer.com/article/10.1186/s13690-021-00687-0#Sec2>

**Extracting social determinants of health from electronic health records using natural language processing: a systematic review**

[**https://academic.oup.com/jamia/article/28/12/2716/6382241?login=true**](https://academic.oup.com/jamia/article/28/12/2716/6382241?login=true)

**Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018**

[**https://www.sciencedirect.com/science/article/pii/S0360835219305893#s0030**](https://www.sciencedirect.com/science/article/pii/S0360835219305893#s0030)

Machine Learning Algorithms in Healthcare: A Literature Survey

<https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9225642>

# **Artificial intelligence and machine learning algorithms for early detection of skin cancer in community and primary care settings: a systematic review**

<https://www.thelancet.com/journals/landig/article/PIIS2589-7500(22)00023-1/fulltext>

# **Supervised Machine Learning Models for Prediction of COVID-19 Infection using Epidemiology Dataset**

[**https://link.springer.com/article/10.1007/s42979-020-00394-7#Sec3**](https://link.springer.com/article/10.1007/s42979-020-00394-7#Sec3)

Deep learning modelling techniques: current progress, applications, advantages, and challenges

<https://link.springer.com/article/10.1007/S10462-023-10466-8#Sec45>

Deep imputation of missing values in time series health data: A review with benchmarking

<https://www.sciencedirect.com/science/article/pii/S1532046423001612#b6>

**Evaluating the state of the art in missing data imputation for clinical data**

[**https://academic.oup.com/bib/article/23/1/bbab489/6457164**](https://academic.oup.com/bib/article/23/1/bbab489/6457164)

# **Handling missing predictor values when validating and applying a prediction model to new patients**

[**https://pmc.ncbi.nlm.nih.gov/articles/PMC7586995/#abstract1**](https://pmc.ncbi.nlm.nih.gov/articles/PMC7586995/#abstract1)

**Dealing with Missing Predictor Values When Applying Clinical Prediction Models**

[**https://academic.oup.com/clinchem/article/55/5/994/5631751**](https://academic.oup.com/clinchem/article/55/5/994/5631751)

**Other pre-processing methods**

# **Encoding-Based Machine Learning Approach for Health Status Classification and Remote Monitoring of Cardiac Patients**

[**https://www.mdpi.com/1999-4893/18/2/94**](https://www.mdpi.com/1999-4893/18/2/94)

# **Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features**

[**https://link.springer.com/article/10.1007/s00180-022-01207-6**](https://link.springer.com/article/10.1007/s00180-022-01207-6)

# **Cross-validation: what does it estimate and how well does it do it?**

[**https://pmc.ncbi.nlm.nih.gov/articles/PMC11412612/#S6**](https://pmc.ncbi.nlm.nih.gov/articles/PMC11412612/#S6)

**Evaluation metrics**

**Evaluation metrics and statistical tests for machine learning**

[**https://www.nature.com/articles/s41598-024-56706-x**](https://www.nature.com/articles/s41598-024-56706-x)

**A comprehensive survey of loss functions and metrics in machine learning**

[**https://link.springer.com/article/10.1007/s10462-025-11198-7#Sec19**](https://link.springer.com/article/10.1007/s10462-025-11198-7#Sec19)

[**https://link.springer.com/article/10.1007/s10462-025-11198-7/tables/5**](https://link.springer.com/article/10.1007/s10462-025-11198-7/tables/5)

[**https://link.springer.com/article/10.1007/s10462-025-11198-7/tables/4**](https://link.springer.com/article/10.1007/s10462-025-11198-7/tables/4)

# **Comprehensive Analysis of Random Forest and XGBoost Performance with SMOTE, ADASYN, and GNUS Under Varying Imbalance Levels**

[**https://www.mdpi.com/2227-7080/13/3/88#:~:text=Additionally%2C%20XGBoost%20demonstrated%20superior%20performance,primary%20models%20for%20this%20research**](https://www.mdpi.com/2227-7080/13/3/88#:~:text=Additionally%2C%20XGBoost%20demonstrated%20superior%20performance,primary%20models%20for%20this%20research)**.**

**Imputation**

**An investigation of the imputation techniques for missing values in ordinal data enhancing clustering and classification analysis validity**

[**https://www.sciencedirect.com/science/article/pii/S2772662223001819**](https://www.sciencedirect.com/science/article/pii/S2772662223001819)

**Handling missing data in clinical research**

<https://www.sciencedirect.com/science/article/pii/S0895435622002189?via%3Dihub>

**Missing data analysis in regression**

[**https://www.tandfonline.com/doi/full/10.1080/08839514.2022.2032925#d1e156**](https://www.tandfonline.com/doi/full/10.1080/08839514.2022.2032925#d1e156)

**The prevention and handling of missing data**

[**https://pmc.ncbi.nlm.nih.gov/articles/PMC3668100/**](https://pmc.ncbi.nlm.nih.gov/articles/PMC3668100/)

**A survey on missing data in machine learning**

[**https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00516-9**](https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00516-9)

**Missing value imputation affects the performance of machine learning: A review and analysis of the literature (2010-2021)**

[**https://www.sciencedirect.com/science/article/pii/S2352914821002653?via%3Dihub**](https://www.sciencedirect.com/science/article/pii/S2352914821002653?via%3Dihub)

**Missing value imputation: A review and analysis of the literature (2006-2017)**

[**https://link.springer.com/article/10.1007/s10462-019-09709-4**](https://link.springer.com/article/10.1007/s10462-019-09709-4)

**Review: A gentle introduction to imputation of missing values**

[**https://www.sciencedirect.com/science/article/pii/S0895435606001971**](https://www.sciencedirect.com/science/article/pii/S0895435606001971)

**Euclidean distance estimation in incomplete datasets**

[**https://www.sciencedirect.com/science/article/pii/S0925231217304320**](https://www.sciencedirect.com/science/article/pii/S0925231217304320)

**Missing covariate data in medical research: To impute is better than to ignore**

[**https://www.sciencedirect.com/science/article/pii/S0895435610000193**](https://www.sciencedirect.com/science/article/pii/S0895435610000193)

# **How much missing data is too much to impute for longitudinal health indicators? A preliminary guideline for the choice of the extent of missing proportion to impute with multiple imputation by chained equations**

[**https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-025-00364-2**](https://pophealthmetrics.biomedcentral.com/articles/10.1186/s12963-025-00364-2)

**A review of ensemble learning and data augmentation models for class imbalanced problems: Combination, implementation and evaluation**

[**https://www.sciencedirect.com/science/article/pii/S0957417423032803**](https://www.sciencedirect.com/science/article/pii/S0957417423032803)

**Multiple imputation by chained equations: what is it and how does it work?**

[**https://onlinelibrary.wiley.com/doi/epdf/10.1002/mpr.329**](https://onlinelibrary.wiley.com/doi/epdf/10.1002/mpr.329)

# **Multiple imputation for analysis of incomplete data in distributed health data networks**

[**https://www.nature.com/articles/s41467-020-19270-2**](https://www.nature.com/articles/s41467-020-19270-2)

**Missing data imputation via the expectation-maximization algorithm can improve principal component analysis aimed at deriving biomarker profiles and dietary patterns**

[**https://pubmed.ncbi.nlm.nih.gov/32035304/**](https://pubmed.ncbi.nlm.nih.gov/32035304/)

**Imputation of multivariate time series data-performance benchmarks for multiple imputation and spectral techniques**

[**https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=2996611**](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2996611)