

Documents

Ismail, A.R.^a, Abidin, N.Z.^a, Maen, M.K.^b

Systematic Review on Missing Data Imputation Techniques with Machine Learning Algorithms for Healthcare
(2022) *Journal of Robotics and Control (JRC)*, 3 (2), pp. 143-152.

DOI: 10.18196/jrc.v3i2.13133

^a Department of Computer Science, Kulliyah of Information and Communication Technology, International Islamic University Malaysia, P.O Box 10, Kuala Lumpur, 50728, Malaysia

^b Department of Information Technology, ITC, hus 1, 2 och 4, Lägerhyddsvägen 2, Box 337, UPPSALA, 751 05, Sweden

Abstract

Missing data is one of the most common issues encountered in data cleaning process especially when dealing with medical dataset. A real collected dataset is prone to be incomplete, inconsistent, noisy and redundant due to potential reasons such as human errors, instrumental failures, and adverse death. Therefore, to accurately deal with incomplete data, a sophisticated algorithm is proposed to impute those missing values. Many machine learning algorithms have been applied to impute missing data with plausible values. However, among all machine learning imputation algorithms, KNN algorithm has been widely adopted as an imputation for missing data due to its robustness and simplicity and it is also a promising method to outperform other machine learning methods. This paper provides a comprehensive review of different imputation techniques used to replace the missing data. The goal of the review paper is to bring specific attention to potential improvements to existing methods and provide readers with a better grasps of imputation technique trends. © 2022 by the authors.

Author Keywords

Healthcare; Imputation; Machine Learning; Missing Data; Review

References

- Rahm, E., Do, Hong Hai
Data Cleaning: Problems and Current Approaches
(2000) *IEEE Data Eng. Bull*, pp. 1-11.
January 2000
- Vinutha, M. R., Chandrika, J.
Imputation as a technique for enhancing the quality of medical data
(2021) *Int. J. Curr. Res. Rev*, 13 (5), pp. 91-95.
- Al Khaldy, M., Kambhampati, C.
Performance Analysis of Various Missing Value Imputation Methods on Heart Failure Dataset
(2016) *Proc. SAI Intell. Syst. Conf*,
- Usharani, Y., Sannulal, P.
A novel approach for imputation of missing values for mining medical datasets
(2016) *2015 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2015*,
- Zhu, Y., Duan, X.
Predictive nursing helps improve treatment efficacy, treatment compliance, and quality of life in unstable angina pectoris patients
(2021) *Am. J. Transl. Res*, 13 (4), pp. 3473-3479.

- Wang, A., Lim, H., Cheng, S. Y., Xie, L.
ANTENNA, a Multi-Rank, Multi-Layered Recommender System for Inferring Reliable Drug-Gene-Disease Associations: Repurposing Diazoxide as a Targeted Anti-Cancer Therapy
(2018) *IEEE/ACM Trans. Comput. Biol. Bioinforma*,
- Kamisan, N. A. B., Lee, M. H., Hussin, A. G., Zubairi, Y. Z.
Imputation techniques for incomplete load data based on seasonality and orientation of the missing values
(2020) *Sains Malaysiana*, 49 (5), pp. 1165-1174.
- Yelipe, U. R., Porika, S., Golla, M.
An efficient approach for imputation and classification of medical data values using class-based clustering of medical records
(2018) *Comput. Electr. Eng*, 66, pp. 487-504.
- Usharani, Y., Sammulal, P.
A novel approach for imputation of missing values for mining medical datasets
(2016) *2015 IEEE Int. Conf. Comput. Intell. Comput. Res. ICCIC 2015*,
- Zainal Abidin, N. Z., Ismail, A. R., Emran, N. A.
Performance Analysis of Machine Learning Algorithms for Missing Value Imputation
(2018) *Int. J. Adv. Comput. Sci. Appl*, 9 (6).
- Pauzi, N. A. M., Wah, Y. B., Deni, S. M., Rahim, S. K. N. A., Suhartono
Comparison of single and mice imputation methods for missing values: A simulation study
(2021) *Pertanika J. Sci. Technol*, 29 (2), pp. 979-998.
- Bargelloni, L.
Data imputation and machine learning improve association analysis and genomic prediction for resistance to fish photobacteriosis in the gilthead sea bream
(2021) *Aquac. Reports*, 20, p. 100661.
- Erlyn, I., Rachmawan, W.
Optimization of Missing Value Imputation using Reinforcement Programming
(2015) *Int. Electron. Symp. (IES), IEEE*, pp. 128-133.
- Platias, C., Petasis, G.
A comparison of machine learning methods for data imputation
(2020) *PervasiveHealth Pervasive Comput. Technol. Healthc*, pp. 150-159.
- Thomas, E.
T. and Rajabi, "A systematic review of machine learning-based missing value imputation techniques
(2021) *Data Technol. Appl*, 55 (4), pp. 558-585.
- Ayilara, O. F., Zhang, L., Sajobi, T. T., Sawatzky, R., Bohm, E., Lix, L. M.
Impact of missing data on bias and precision when estimating change in patient-reported outcomes from a clinical registry
(2019) *Health Qual. Life Outcomes*, 17 (1), pp. 1-9.
- Rostami, S., Kleszcz, A., Dimanov, D., Katos, V.
A machine learning approach to dataset imputation for software vulnerabilities
(2020) *Commun. Comput. Inf. Sci*, 1284, pp. 25-36.

CCIS September

- Lo, A. W., Siah, K. W., Wong, C. H.
Machine Learning with Statistical Imputation for Predicting Drug Approval
(2019) *Harvard Data Sci. Rev.*, (1).
- Bhattacharjee, A., Bayzid, M. S.
Machine learning based imputation techniques for estimating phylogenetic trees from incomplete distance matrices
(2020) *BMC Genomics*, 21 (1), pp. 1-14.
- Su, Y. C., Wu, C. Y., Yang, C. H., Li, B. S., Moi, S. H., Da Lin, Y.
Machine learning data imputation and prediction of foraging group size in a Kleptoparasitic spider
(2021) *Mathematics*, 9 (4), pp. 1-16.
- Solomon, N., Lokhnygina, Y., Halabi, S.
Comparison of regression imputation methods of baseline covariates that predict survival outcomes
(2021) *J. Clin. Transl. Sci.*, 5 (1).
- Khayati, M., Lerner, A., Tymchenko, Z., Cudre´Mauroux, P.
Mind the gap: An experimental evaluation of imputation of missing values techniques in time series
(2020) *Proc. VLDB Endow*, 13 (5), pp. 768-782.
- Huang, L., Wang, C., Rosenberg, N. A.
The Relationship between Imputation Error and Statistical Power in Genetic Association Studies in Diverse Populations
(2009) *Am. J. Hum. Genet.*, 85 (5), pp. 692-698.
- Hughes, R. A., Heron, J., Sterne, J. A. C., Tilling, K.
Accounting for missing data in statistical analyses: Multiple imputation is not always the answer
(2019) *Int. J. Epidemiol.*, 48 (4), pp. 1294-1304.
- Petrazzini, B. O., Naya, H., Lopez-Bello, F., Vazquez, G., Spangenberg, L.
Evaluation of different approaches for missing data imputation on features associated to genomic data
(2021) *BioData Min*, 14 (1), pp. 1-13.
- Siafis, S.
Imputing the number of responders from the mean and standard deviation of CGI-improvement in clinical trials investigating medications for autism spectrum disorder
(2021) *Brain Sci*, 11 (7).
- Hardouin, J. B., Conroy, R., Sébille, V.
Imputation by the mean score should be avoided when validating a Patient Reported Outcomes questionnaire by a Rasch model in presence of informative missing data
(2011) *BMC Med. Res. Methodol.*, 11, pp. 1-13.
- Malan, L., Smuts, C. M., Baumgartner, J., Ricci, C.
Missing data imputation via the expectation-maximization algorithm can improve principal component analysis aimed at deriving biomarker profiles and dietary

patterns

(2020) *Nutr. Res*, 75, pp. 67-76.

- Ghomrawi, H. M. K., Mandl, L. A., Rutledge, J., Alexiades, M. M., Mazumdar, M.
Is there a role for expectation maximization imputation in addressing missing data in research using WOMAC questionnaire? Comparison to the Standard mean approach and a tutorial
(2011) *BMC Musculoskelet. Disord*, 12 (1), p. 109.
- Liu, Y., Gopalakrishnan, V.
An overview and evaluation of recent machine learning imputation methods using cardiac imaging data
(2017) *Data*, 2 (1).
- Javadi, S., Bahrampour, A., Saber, M. M., Garrusi, B., Baneshi, M. R.
Evaluation of Four Multiple Imputation Methods for Handling Missing Binary Outcome Data in the Presence of an Interaction between a Dummy and a Continuous Variable
(2021) *J. Probab. Stat*, 2021, pp. 1-14.
- Chowdhury, S. I., Islam, M. K., Khan
Imputation of Missing Healthcare Data
(2017) *Comput. Inf. Technol. (ICCIT), 2017 20th Int. Conf*, pp. 1-6.
M. H
- Kahale, L. A.
Potential impact of missing outcome data on treatment effects in systematic reviews: Imputation study
(2020) *BMJ*, 370, pp. 1-10.
- Köse, T., Özgür, S., Coşgun, E., Keskinöğlü, A., Keskinöğlü, P., Mrozek, D.
Effect of Missing Data Imputation on Deep Learning Prediction Performance for Vesicoureteral Reflux and Recurrent Urinary Tract Infection Clinical Study
(2020) *Biomed Res. Int*, 2020.
- Wijesuriya, R., Moreno-Betancur, M., Carlin, J. B., Lee, K. J.
Evaluation of approaches for multiple imputation of three-level data
(2020) *BMC Med. Res. Methodol*, 20 (1), pp. 1-15.
- Hegde, H., Shimpi, N., Panny, A., Glurich, I., Christie, P., Acharya, A.
MICE vs PPCA: Missing data imputation in healthcare
(2019) *Informatics Med. Unlocked*, 17 (1), pp. 100275-48.
November, [37] I. Jordanov, N. Petrov, and A. Petrozziello, "Classifiers Accuracy Improvement Based on Missing Data Imputation, *J. Artif. Intell. Soft Comput. Res.*, 8 31 2018
- Kokla, M., Virtanen, J., Kolehmainen, M., Paananen, J., Hanhineva, K.
Random forest-based imputation outperforms other methods for imputing LC-MS metabolomics data: A comparative study
(2019) *BMC Bioinformatics*, 20 (1), pp. 1-11.
- Etoeharnowo, T., Van Os, M. H. J. A.
A Random Forest Approach for Dealing with Missingness: a Case Study in Primary Care Data

(2020) *Leidin Inst. Adv. Comput. Sci*,

- Hong, S., Lynn, H. S.
Accuracy of random-forest-based imputation of missing data in the presence of non-normality, non-linearity, and interaction
(2020) *BMC Med. Res. Methodol*, 20 (1), pp. 1-12.
- Mostafa, S. M.
Imputing missing values using cumulative linear regression
(2019) *CAAI Trans. Intell. Technol*, 4 (3), pp. 182-200.
- Sivapriya, T. R.
Imputation And Classification Of Missing Data Using Least Square Support Vector Machines – A New Approach In Dementia Diagnosis
(2012) *IJARAI - Int. J. Adv. Res. Artif. Intell*, 1 (4), pp. 29-34.
- Rahman, M., Davis, D. N.
Machine Learning Based Missing Value Imputation Method for Clinical Datasets
(2013) *IAENG Trans. Eng. Technol. Springer Netherlands*, 247.
January
- Mandel, S. P.
A Comparison of Six Methods for Missing Data Imputation
(2015) *J. Biom. Biostat*, 6 (1), pp. 1-6.
J
- Seffens, W.
Machine Learning Data Imputation and Classification in a Multicohort Hypertension Clinical Study
(2015) *Bioinform. Biol. Insights*, 9s3, pp. 43-54.
- Razzaghi, T., Roderick, O., Safro, I., Marko, N.
(2015) *Fast Imbalanced Classification of Healthcare Data with Missing Values*, pp. 1-13.
- Huang, M. W., Lin, W. C., Tsai, C. F.
Outlier Removal in Model-Based Missing Value Imputation for Medical Datasets
(2018) *J. Healthc. Eng*, 2018.
- Bektaş, J., Ibrikçi, T., Özcan, İ. T.
The impact of imputation procedures with machine learning methods on the performance of classifiers: An application to coronary artery disease data including missing values
(2018) *Biomed. Res*, 29 (13), pp. 2780-2785.
- Cheng, C., Huang, H.
(2019) *A Distance-Threshold kNN Method for Imputing Medical Data Missing Values*, 7 (1), pp. 13-17.
- Huang, S. F., Cheng, C. H.
A safe-region imputation method for handling medical data with missing values
(2020) *Symmetry MDPI*, 12 (11), pp. 1-19.
- Rockel, T., Joenssen, D. W., Bankhofer, U.
Decision Trees for the Imputation of Categorical Data
(2017) *Kit Sci. Publ*, 2 (1), pp. 1-15.

- Ameta, A., Jain, K.
Data Mining Techniques for the Prediction of Kidney Diseases and Treatment: A Review
(2017) *Int. J. Eng. Comput. Sci*, 6 (2), pp. 20376-20378.
- Nair, S., De La Vara, J. L., Sabetzadeh, M., Briand, L.
An extended systematic literature review on provision of evidence for safety certification
(2014) *Inf. Softw. Technol*, 56 (7), pp. 689-717.
[54] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering, Engineering, 2, 1051, 2007
- Bose, S., Das, C., Gangopadhyay, T., Chattopadhyay, S.
A modified local least squares-based missing value estimation method in microarray gene expression data
(2013) *Int. Conf. Adv. Comput. Netw. Secur*, pp. 18-23.
- Choi, Y. Y., Shon, H., Byon, Y. J., Kim, D. K., Kang, S.
Enhanced application of principal component analysis in machine learning for imputation of missing traffic data
(2019) *Appl. Sci*, 9 (10), pp. 1-15.
- Su, X., Greiner, R., Khoshgoftaar, T. M., Napolitano, A.
Using classifier-based nominal imputation to improve machine learning
(2011) *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 6634, pp. 124-135.
LNAI PART 1
- Atem, F. D., Sampene, E., Greene, T. J.
Improved conditional imputation for linear regression with a randomly censored predictor
(2019) *Stat. Methods Med. Res*, 28 (2), pp. 432-444.
- Sundararajan, A., Sarwat, A. I.
Evaluation of Missing Data Imputation Methods for an Enhanced Distributed PV Generation Prediction
(2019) *Proc. Futur. Technol. Conf*, pp. 590-609.
2AD
- Alber, M.
Integrating machine learning and multiscale modeling—perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences
(2019) *npj Digit. Med*, 2 (1).
- Signorini, M. G., Pini, N., Malovini, A., Bellazzi, R., Magenes, G.
Integrating machine learning techniques and physiology based heart rate features for antepartum fetal monitoring
(2020) *Comput. Methods Programs Biomed*, 185.
- Nikfalazar, S., Yeh, C. H., Bedingfield, S., Khorshidi, H. A.
Missing data imputation using decision trees and fuzzy clustering with iterative learning
(2020) *Knowl. Inf. Syst*, 62 (6), pp. 2419-2437.

- Keerin, P., Kurutach, W.
An Improvement of Missing Value Imputation in DNA Microarray Data Using Cluster-based LLS Method
(2013) *2013 13th Int. Symp. Commun. Inf. Technol.*, pp. 559-564.
- Dzulkalnine, M. F., Sallehuddin, R.
Missing data imputation with fuzzy feature selection for diabetes dataset
(2019) *SN Appl. Sci.*, 1 (4), pp. 1-12.
- Chai, L. E.
Investigating the effects of imputation methods for modelling gene networks using a dynamic Bayesian network from gene expression data
(2014) *Malaysian J. Med. Sci.*, 21 (2), pp. 20-27.
- Wasito, I., Mirkin, B.
Nearest neighbours in least-squares data imputation algorithms with different missing patterns
(2006) *Comput. Stat. Data Anal.*, 50 (4), pp. 926-949.
- Saha, S., Bandopadhyay, S., Ghosh, A., Dey, K. N.
An improved fuzzy based approach to impute missing values in DNA microarray gene expression data with collaborative filtering
2016 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2016, 2016, pp. 911-916.
- Garcia-Laencina, Pedro J., Vidal, A. R. F., Sancho-Gomez, J.-L.
A Robust Approach For Classifying Unknown Data in Medical Diagnosis Problems
(2008) *2008 World Automation Congress*,
- Tavazzi, E., Daberdaku, S., Vasta, R., Calvo, A., Chiò, A., Di Camillo, B.
Exploiting mutual information for the imputation of static and dynamic mixed-type clinical data with an adaptive k-nearest neighbours approach
(2020) *BMC Med. Inform. Decis. Mak.*, 20, pp. 1-23.
Suppl 5
- García-Laencina, P. J., Sancho-Gómez, J. L., Figueiras-Vidal, A. R., Verleysen, M.
K-nearest neighbours based on mutual information for incomplete data classification
(2014) *ESANN 2008 Proceedings, 16th Eur. Symp. Artif. Neural Networks - Adv. Comput. Intell. Learn.*, pp. 37-42.
May 2008
- García-Laencina, P. J., Sancho-Gómez, J. L., Figueiras-Vidal, A. R., Verleysen, M.
K nearest neighbours with mutual information for simultaneous classification and missing data imputation
(2009) *Neurocomputing*, 72 (7–9), pp. 1483-1493.
- Kumutha, V., Palaniammal, S.
An Enhanced Approach on Handling Missing Values Using Bagging k-NN Imputation
(2013) *2013 Int. Conf. Comput. Commun. Informatics*, pp. 1-8.
- Mehrabani-Zeinabad, K., Doostfateme, M., Ayatollahi, S. M. T.
An efficient and effective model to handle missing data in classification
(2020) *Biomed Res. Int.*, 2020.

- Tran, C. T., Zhang, M., Andrae, P., Xue, B.
Multiple Imputation and Ensemble Learning for Classification with Incomplete Data
(2017) *Intell. Evol. Syst*, 187, p. 2009.
December
- Nanni, L., Lumini, A., Brahnam, S.
A classifier ensemble approach for the missing feature problem
(2012) *Artif. Intell. Med*, 55 (1), pp. 37-50.
- An, D., Little, R. J. A., McNally, J. W.
A multiple imputation approach to disclosure limitation for high-age individuals in longitudinal studies
(2010) *Stat. Med*, 29 (17), pp. 1769-1778.
- Schuetz, C. G.
Using neuroimaging to predict relapse to smoking: role of possible moderators and mediators
(2008) *Int. J. Methods Psychiatr. Res*, 17 (1), pp. S78-S82.
Suppl 1
- Conroy, B., Eshelman, L., Potes, C., Xu-Wilson, M.
A dynamic ensemble approach to robust classification in the presence of missing data
(2016) *Mach. Learn*, 102 (3), pp. 443-463.
- Khan, S. S., Ahmad, A., Mihailidis, A.
Bootstrapping and Multiple Imputation Ensemble Approaches for Missing Data
(2016) *Cornell Univ., no. Mi*, pp. 1-17.
- De Silva, H., Perera, A. S.
(2016) *Missing Data Imputation using Evolutionary k-Nearest Neighbor Algorithm for Gene Expression Data*, pp. 141-146.
- Kamiura, N., Ohtsuka, A., Tanii, H., Isokawa, T., Matsui, N.
On Detection of Hematopoietic Tumors Using Self Organizing Maps and Genetic Algorithms
(2005) *2005 IEEE International Conference on Systems, Man and Cybernetics*,
- Priya, M., Kumar, P. R.
Intelligent Approaches for Prognosticating Atherosclerotic and Non-Atherosclerotic Individuals
(2014) *2014 Int. Conf. Commun. Signal Process*, pp. 691-695.
- Almasinejad, P., Golabpour, A., Reza, M., Meybodi, M., Mirzaie, K., Khosravi, A.
A Dynamic Model for Imputing Missing Medical Data: A Multiobjective Particle Swarm Optimization Algorithm
J. Healthc. Eng, 2021, p. 2021.
- Wang, X., Li, W., Sun, Y., Milanovic, S., Kon, M., Castrillon-Candas, J. E.
(2021) *Multilevel Stochastic Optimization for Imputation in Massive Medical Data Records*, arXiv
- Ismail, A. R., Aziz, N. A., Ralib, A. M., Abidin, N. Z., Bashath, S. S.
A particle swarm optimization levy flight algorithm for imputation of missing creatinine dataset

(2021) *Int. J. Adv. Intell. Informatics*, 7 (2), pp. 225-236.

- Anderson, C. L., Liang, Deborah K.
An imputation-regularized optimization algorithm for high dimensional missing data problems and beyond
(2017) *Physiol. Behav.*, 176 (5), pp. 139-148.
J. W. and
- Ke, C.
Prognostics of surgical site infections using dynamic health data
(2017) *J. Biomed. Inform.*, 65, pp. 22-33.
- Rao, V. S. H., Member, S., Kumar, M. N.
Novel Approaches for Predicting Risk Factors of Atherosclerosis
(2013) *IEEE J. Biomed. Heal. Informatics*, 17 (1), pp. 183-189.
- Najib, M., Samat, N. A.
FCMPSO: An Imputation for Missing Data Features in Heart Disease Classification
(2017) *IOP Conf. Ser. Mater. Sci. Eng.*,
- Nekouie, A., Moattar, M. H.
Missing value imputation for breast cancer diagnosis data using tensor factorization improved by enhanced reduced adaptive particle swarm optimization
(2018) *J. King Saud Univ. - Comput. Inf. Sci.*,
- Suresh, M., Taib, R., Zhao, Y., Jin, W.
Sharpening the BLADE: Missing Data Imputation Using Supervised Machine Learning
(2019) *Data Integr. Partnersh. Aust.*, pp. 215-227.
July
- Khan, S. I., Hoque, A. S. M. L.
SICE: an improved missing data imputation technique
(2020) *J. Big Data*, 7 (1).
- Guo, Y., Liu, Z., Krishnswamy, P., Ramasamy, S.
(2019) *Bayesian Recurrent Framework for Missing Data Imputation and Prediction with Clinical Time Series*,
arXiv
- Skivington, K.
A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance
(2018) *BMJ Open*, 2021, p. n2061.
- Zhao, Q., Lee, D.
(2007) *HICCUP: Hierarchical Clustering Based Value Imputation using Heterogeneous Gene Expression Microarray Datasets*,
- Bertsimas, D., Orfanoudaki, A., Pawlowski, C.
(2020) *Imputation of clinical covariates in time series*, 110 (1).
Springer US

- Kumar, R., Chen, T., Hardt, M., Beymer, D., Brannon, K., Syeda-Mahmood, T.
Multiple Kernel Completion and its application to cardiac disease discrimination
(2013) *Proc. - Int. Symp. Biomed. Imaging*, pp. 764-767.
- Le, T., Altman, T., Gardiner, K. J.
Probability-based Imputation Method for Fuzzy Cluster Analysis of Gene Expression Microarray Data
(2012) *2012 Ninth Int. Conf. Inf. Technol. - New Gener*, pp. 42-47.
- Liu, M., Gao, Y., Yap, P. T., Shen, D.
Multi-Hypergraph Learning for Incomplete Multimodality Data
(2018) *IEEE J. Biomed. Heal. Informatics*, 22 (4), pp. 1197-1208.
- Valarmathie, P., Dinakaran, K.
An efficient technique for missing value imputation in microarray gene expression data
(2014) *Icccs*, 114, pp. 073-080.
- Wu, X., Akbarzadeh Khorshidi, H., Aickelin, U., Edib, Z., Peate, M.
Imputation techniques on missing values in breast cancer treatment and fertility data
(2019) *Heal. Inf. Sci. Syst*, 7 (1), pp. 1-8.
- Prates, M. O.
Spatial extreme learning machines: An application on prediction of disease counts
(2018) *Stat. Methods Med. Res*, p. 96228021876798.
- Jin, L.
A comparative study of evaluating missing value imputation methods in label-free proteomics
(2021) *Sci. Rep*, 11 (1), pp. 1-11.
- Zitnik, M.
Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities
(2020) *HHS Public Access*, (50), pp. 71-91.
- Ghandeharioun, A.
Objective Assessment of Depressive Symptoms with Machine Learning and Wearable Sensors Data
(2017) *2017 Seventh Int. Conf. Affect. Comput. Intell. Interact*, pp. 325-332.
- Velasco-Gallego, C., Lazakis, I.
Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study
(2020) *Ocean Eng*, 218, p. 108261.
July
- Nur, N. D., Md Tahir, N., Iffah Abd Latiff, Z., Huzaimy Jusoh, M., Akimasa, Y.
Missing data imputation of MAGDAS-9's ground electromagnetism with supervised machine learning and conventional statistical analysis models
(2022) *Alexandria Eng. J*, 61 (1), pp. 937-947.
M. A. H

- Alade, O. A., Sallehuddin, R., Selamat, A.
Empirical Performance Evaluation of Imputation Techniques using Medical Dataset
(2019) *IOP Conf. Ser. Mater. Sci. Eng*, 551 (1), pp. 0-5.
- Kshirsagar, M., Carbonell, J., Klein-Seetharaman, J.
Techniques to cope with missing data in host-pathogen protein interaction prediction
(2012) *Bioinformatics*, 28 (18), pp. 466-472.
- Kontopantelis, E., White, I. R., Sperrin, M., Buchan, I.
Outcome-sensitive multiple imputation: A simulation study
(2017) *BMC Med. Res. Methodol*, 17 (1), pp. 1-13.
- Ghoneim, V. F., Solouma, N. H., Kadah, Y. M.
Evaluation of missing values imputation methods in cDNA microarrays based on classification accuracy
(2011) *2011 1st Middle East Conf. Biomed. Eng. MECBME 2011*, pp. 367-370.
- Chang, C., Deng, Y., Jiang, X., Long, Q.
Multiple imputation for analysis of incomplete data in distributed health data networks
(2020) *Nat. Commun*, 11 (1), pp. 1-11.
- Ghoneim, V. F., Solouma, N. H., Kadah, Y. M.
The Impact of Missing Values Imputation Methods in cDNA Microarrays on Downstream Data Analysis
(2011) *28th NATIONAL RADIO SCIENCE CONFERENCE 28th NATIONAL RADIO SCIENCE CONFERENCE*,
Nrsc
- Guo, C. Y., Yang, Y. C., Chen, Y. H.
The Optimal Machine Learning-Based Missing Data Imputation for the Cox Proportional Hazard Model
(2021) *Front. Public Heal*, 9, pp. 1-8.
July
- Colubri, A., Silver, T., Fradet, T., Retzepi, K., Fry, B., Sabeti, P.
Transforming Clinical Data into Actionable Prognosis Models: Machine-Learning Framework and Field-Deployable App to Predict Outcome of Ebola Patients
(2016) *PLoS Negl. Trop. Dis*, 10 (3), pp. 1-17.
- Mitra, M., Samanta, R. K.
A Study on UCI Hepatitis Disease Dataset Using Soft Computing
(2017) *AMSE Journals IIETA*, 78, pp. 467-477.
- Suh, M.-K., Woodbridge, J., Lan, M., Bui, A., Evangelista, L. S., Sarrafzadeh, M.
Missing Data Imputation for Remote CHF Patient Monitoring Systems
(2012) *Conf. Proc. IEEE Eng. Med. Biol. Soc. NIH Public Access*, pp. 3184-3187.
- Hondula, D. M.
A respiratory alert model for the Shenandoah Valley, Virginia, USA
(2013) *Int. J. Biometeorol*, 57 (1), pp. 91-105.

- Lin, W. C., Tsai, C. F.
Missing value imputation: a review and analysis of the literature (2006–2017)
(2020) *Artif. Intell. Rev*, 53 (2), pp. 1487-1509.
- Phung, S., Kumar, A., Kim, J.
A deep learning technique for imputing missing healthcare data
(2019) *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6513-6516.
- Jerez, J. M.
Missing data imputation using statistical and machine learning methods in a real breast cancer problem
(2010) *Artif. Intell. Med*, 50 (2), pp. 105-115.
- Grace-Martin, K.
Limitations of Common Solutions to Missing Data
(2001) *Cornell Statistical Consulting Unit*,
November
- Saunders, J. A., Morrow-howell, N., Spitznagel, E., Doré, P., Proctor, E. K., Pescarino, R.
Imputing Missing Data: A Comparison of Methods for Social Work Researchers
(2006) *Social Work Research*, pp. 19-31.
- Huque, M. H., Carlin, J. B., Simpson, J. A., Lee, K. J.
A comparison of multiple imputation methods for missing data in longitudinal studies 01 Mathematical Sciences
(2018) *BMC Med. Res. Methodol*, 18 (1), pp. 1-16.
- Waljee, A. K.
Comparison of imputation methods for missing laboratory data in medicine
(2013) *BMJ Open*, 3 (8), pp. 1-8.
- Soley-bori, M.
Dealing with missing data: Key assumptions and methods for applied analysis
(2013) *PM931 Dir. Study Heal. Policy Manag*, (4), p. 20.
- Haji-maghsoudi, S., Haghdoost, A., Rastegari, A., Baneshi, M. R.
Influence of Pattern of Missing Data on Performance of Imputation Methods: An Example Using National Data on Drug Injection in Prisons
(2013) *Int. J. Heal. Policy Manag*, 1 (1), pp. 69-77.
- Payrovnaziri, S. N., Xing, A., Salman, S., Liu, X., Bian, J., He, Z.
The Impact of Imputation on the Interpretations of Prediction Models: A Case Study on Mortality Prediction for Patients with Acute Myocardial Infarction
(2021) *AMIA ... Annu. Symp. proceedings. AMIA Symp*, 2021, pp. 465-474.
- Rubin, D. B., Little, R. J. A.
(2002) *Statistical Analysis with Missing Data, Second Edi*,
New York: A John Wiley & Sons, Inc., Publication
- Kang, H.
The prevention and handling of the missing data
(2013) *Korean J. Anesthesiol*, 64 (5), pp. 402-406.

- Leurent, B., Gomes, M., Cro, S., Wiles, N., Carpenter, J. R.
Reference-based multiple imputation for missing data sensitivity analyses in trial-based cost-effectiveness analysis
(2020) *Heal. Econ. (United Kingdom)*, 29 (2), pp. 171-184.
- Pedersen, A. B.
Missing data and multiple imputation in clinical epidemiological research
(2017) *Clin. Epidemiol*, 9, pp. 157-166.
- Jacques-Emmanuel Galimard, M. R.-R., Chevret, Sylvie, Protopopescu, Camelia
A multiple imputation approach for MNAR mechanisms compatible with Heckman's model
(2016) *Wiley Online Libr*,
- Pampaka, M., Hutcheson, G., Williams, J.
Handling missing data: analysis of a challenging data set using multiple imputation
(2016) *Int. J. Res. Method Educ*, 7288 (39), pp. 19-37.
1
- Ramli, M. N. N., Yahaya, A. S., Ramli, N. A., Yusof, N. F. F. ., Abdullah, M. M. A.
Roles of Imputation Methods for Filling the Missing Values: A Review
(2013) *International Conference of Advanced Materials Engineering and Technology (ICAMET 2013)*,
November
- Kenward, M. G.
The handling of missing data in clinical trials
(2013) *Clin. Investig. (Lond)*, 3 (3), pp. 241-250.
- Howell, D. C.
(2000) *The Treatment of Missing Data*, pp. 1-44.
- Mera-Gaona, M., Neumann, U., Vargas-Canas, R., López, D. M.
Evaluating the impact of multivariate imputation by MICE in feature selection
(2021) *PLoS One*, 16 (7), pp. 1-28.
July
- Wang, G., Deng, Z., Choi, K.-S.
Tackling missing data in community health studies using additive LS-SVM classifier
(2016) *IEEE J. Biomed. Heal. Informatics*, 22 (2), pp. 1-1.
- Sehgal, M. S. B., Gondal, I., Dooley, L.
K-Ranked Covariance Based Missing Values Estimation for Microarray Data Classification
(2001) *Fourth International Conference on Hybrid Intelligent Systems (HIS'04)*, pp. 8-13.
- Sehgal, M. S. B., Gondal, I., Dooley, L. S.
Gene expression Collateral missing value imputation: a new robust missing value estimation algorithm for microarray data
(2005) *Bioinformatics*, 21 (10), pp. 2417-2423.
- Beretta, L., Santaniello, A.
Nearest neighbor imputation algorithms: a critical evaluation
(2016) *BMC Med. Inform. Decis. Mak*, 16.
Suppl 3

- Acuna, E., Rodriguez, C.
The treatment of missing values and its effect in the classifier accuracy
(2004) *Proceedings of the Meeting of the International Federation of Classification Societies (IFCS)*, (1995), pp. 1-9.
- Chen, W. C. W.
Combining Fourier and Lagged k-Nearest Neighbor Imputation for Biomedical Time Series Data
(2016) *J. Biomed. Informatics*, 33 (2), pp. 557-573.
- Huang, J.
Cross-validation based K nearest neighbor imputation for software quality datasets: An empirical study
(2017) *J. Syst. Softw*, 132, pp. 226-252.
- Samant, R., Rao, S.
Effects of Missing Data Imputation on Classifier Accuracy
(2013) *Int. J. Eng. Res. Technol*, 2 (11), pp. 264-266.
- Rahman, M. G., Islam, M. Z.
A decision tree-based missing value imputation technique for data pre-processing
(2010) *Conf. Res. Pract. Inf. Technol. Ser*, 121, pp. 41-50.
- Jönsson, Per, Wohlin, Claes
An Evaluation of k -Nearest Neighbour Imputation Using Likert Data
(2004) *10th Int. Symp. Softw. Metrics, 2004. Proc*,
- Askarian, M., Escudero, G., Graells, M., Zarghami, R., Jalali-Farahani, F., Mostoufi, N.
Fault diagnosis of chemical processes with incomplete observations: A comparative study
(2016) *Comput. Chem. Eng*, 84, pp. 104-116.
- Rahman, M. G., Islam, M. Z.
Missing value imputation using decision trees and decision forests by splitting and merging records: Two novel techniques
(2013) *Knowledge-Based Syst*, 53, pp. 51-65.

Correspondence Address

Ismail A.R.; Department of Computer Science, P.O Box 10, Malaysia; email: amelia@iium.edu.my

Publisher: Department of Electrical Engineering, Universitas Muhammadiyah Yogyakarta

ISSN: 27155056

Language of Original Document: English

Abbreviated Source Title: J. Robot. Control.

2-s2.0-85131243782

Document Type: Review

Publication Stage: Final

Source: Scopus