

Documents

Zamani, A.S.^{a b}, Hashim, A.H.A.^b, Shatat, A.S.A.^c, Akhtar, M.M.^d, Rizwanullah, M.^a, Mohamed, S.S.I.^a

Implementation of machine learning techniques with big data and IoT to create effective prediction models for health informatics

(2024) *Biomedical Signal Processing and Control*, 94, art. no. 106247, .

DOI: 10.1016/j.bspc.2024.106247

^a Department of Computer and Self Development, Preparatory Year Deanship, Prince Sattam bin Abdulaziz University, Saudi Arabia

^b Department of Electrical and Computer Engineering, International Islamic University Malaysia, Malaysia

^c Department of Management Information Systems, College of Administrative Sciences, Applied Science University, Bahrain

^d College of Applied Medical Science, Department of Basic Sciences, Riyadh Elm University (REU), Saudi Arabia

Abstract

As a result of the availability of healthcare data in sheer size, big data analytics has to grow regularly in this industry to ensure new and effective opportunities. This is helpful in providing early prevention, prediction, and detection of disease, thus helping in the enhancement of the overall life quality of the individuals. Likewise, in this paper, a machine learning-based big data analytics model is developed for predicting multi-diseases to provide a better decision support system for various healthcare applications. This developed framework utilizes the MapReduce framework, where the map phase performs feature extraction and the reduce phase performs feature selection for the purpose of handling and processing big data. The required healthcare data is collected from external web sources. In the map phase, the statistical features and the Principal Component Analysis (PCA) features are extracted. In the reduction phase, the optimal features are selected with the aid of the developed Hybrid Flower Pollination Bumblebees Optimization Algorithm (HFPBOA). Then, the Ensemble Learning (EL) model is developed to predict the multi-diseases. Moreover, the parameters present in the EL classifiers are optimized by using the same HFPBOA. The final prediction output is obtained by averaging the weight function between the outputs of the NN, KNN, and fuzzy classifier. Thus, the offered model attains 40.1%, 28.7%, 23.6%, and 10.5% improved than SSA-EL, DOA-EL, BOA-EL, and FA-EL respectively in terms of best value. The effectiveness computed for the developed multi-disease prediction framework is guaranteed by comparing the results among the recently developed prediction approaches. © 2024 Elsevier Ltd

Author Keywords

Big Data; Fuzzy Classifier; Health Informatics; Hybrid Flower Pollination Bumblebees Optimization Algorithm; K-Nearest Neighbour; Machine Learning Techniques; Neural Networks; Prediction Model

Index Keywords

Big data, Classification (of information), Data Analytics, Data handling, Decision support systems, Forecasting, Fuzzy sets, Health care, Machine learning, Medical informatics, Nearest neighbor search, Principal component analysis; Ensemble learning, Fuzzy classifiers, Health informatics, Hybrid flower pollination bumblebee optimization algorithm, K-near neighbor, Machine learning techniques, Nearest-neighbour, Neural-networks, Optimization algorithms, Prediction modelling; Learning algorithms; algorithm, article, big data, bumblebee, classifier, decision support system, diagnosis, feature extraction, feature selection, flower, fuzzy system, health care availability, human, k nearest neighbor, learning, machine learning, medical informatics, nerve cell network, pollination, prediction, prevention, principal component analysis, quality of life

References

- Yu, W., Liu, Y., Dillon, T., Rahayu, W., Mostafa, F.
An integrated framework for health state monitoring in a Smart factory employing IoT and big data techniques
(2022) *IEEE Internet Things J.*, 9 (3), pp. 2443-2454.
- Zheng, M., Bai, S.
Implementation of universal health Management and monitoring system in resource-constrained environment based on internet of things
(2021) *IEEE Access*, 9, pp. 138744-138752.
- Sivakumar, N.R., Karim, F.K.D.
An IoT based big data framework using equidistant heuristic and duplex deep neural network for diabetic disease prediction
(2021) *J. Ambient Intell. Humanized Comput.*,

- Subramaniaswamy, V., Gunasekaran Manogaran, R., Logesh, V.V., Malathi, D.N.C., Senthilselvan, N.
Retracted article: an ontology-driven personalized food recommendation in IoT-based healthcare system
(2019) *J. Supercomput.*, 75, pp. 3184-3216.
- Andrew Onesimu, J., Karthikeyan, J., Sei, Y.
An efficient clustering-based anonymization scheme for privacy-preserving data collection in IoT based healthcare services
(2021) *Peer-to-Peer Networking and Applications*, 14, pp. 1629-1649.
- Zhong, Y., Chen, L., Dan, C., Rezaeiapanah, A.
A systematic survey of data mining and big data analysis in internet of things
(2022) *J. Supercomput.*,
- Kumar, A., Abhishek, K., Nerurkar, P., Khosravi, M.R.
Muhammad Rukunuddin Ghalib & Achyut Shankar, "big data analytics to identify illegal activities on bitcoin blockchain for IoT,"
(2021) *Pers. Ubiquit. Comput.*,
- Sunder Ali Khowaja
Parus khuwaja, Kapal dev & Giuseppe D'Aniello, "VIRFIM: an AI and internet of medical things-driven framework for healthcare using smart sensors,"
(2021) *Neural Comput. & Applic.*,
- Hashem, M., Youssef, A.E.
Teeth infection and fatigue prediction using optimized neural networks and big data analytic tool
(2020) *Clust. Comput.*, 23, pp. 1669-1682.
- Moqurrab, S.A., Tariq, N., Anjum, A., Asheralieva, A., Malik, S.U.R., Malik, H., Pervaiz, H., Gill, S.S.
A deep Learning-based privacy-preserving model for smart healthcare in internet of medical things using fog computing
(2022) *Wirel. Pers. Commun.*, 126, pp. 2379-2401.
- Ji, C., Yupeng, H., Wang, K., Zhan, P., Li, X., Zheng, X.
Identifiable temporal feature selection via horizontal visibility graph Towards Smart medical applications
(2021) *Interdiscip. Sci.: Comput. Life Sci.*, 13, pp. 717-730.
- Shen, Y.-C., Hsia, T.-C., Hsu, C.-H.
Software optimization in ultrasound imaging technique using improved deep belief Learning network on the internet of medical things platform
(2021) *Wirel. Pers. Commun.*,
- Lin, J.-W., Arul, J.M., Kao, J.-T.
A bottom-up tree based storage approach for efficient IoT data analytics in cloud systems
(2021) *Journal of Grid Computing*, 19.
- Mohapatra, A.G., Talukdar, J., Tarini, C.
Mishra, Sameer Anand, Ajay jaiswal, Ashish Khanna & Deepak Gupta, "fiber bragg grating sensors driven structural health monitoring by using multimedia-enabled IoT and big data technology,"
(2022) *Multimed. Tools Appl.*, 81, pp. 34573-34593.
- Adi, E., Anwar, A.
Zubair Baig&SheraliZeadally, "machine learning and data analytics for the IoT,"
(2020) *Neural Comput. & Applic.*, 32, pp. 16205-16233.

- Himeur, Y., Elnour, M.
FodilFadli, Nader meskin, Ioan petri, YacineRezgui, FaycalBensaali & Abbes Amira, “AI-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives,”
(2022) *Artif. Intell. Rev.*,
- Jamil, H., Umer, T.
CelalCeken&Fadi al-Turjman, “decision based model for real-time IoT analysis using big data and machine Learning,”
(2021) *Wirel. Pers. Commun.*, 121, pp. 2947-2959.
- Kumar, A., Sharma, K., Sharma, A.
Genetically optimized fuzzy C-means data clustering of IoMT-based biomarkers for fast affective state recognition in intelligent edge analytics
(2021) *Appl. Soft Comput.*, 109.
- Tan, T.-H., Hsu, C.-C., Chen, C.-J., Hsu, S.-L., Liu, T.-L., Lin, H.-J., Wang, J.-J., Huang, C.-C.
Predicting outcomes in older ED patients with influenza in real time using a big data-driven and machine learning approach to the hospital information system
(2021) *BMC Geriatrics*, 21, p. 280.
- Safa, M., Pandian, A.
Intelligent big data analytics model for efficient Cardiac disease prediction with IoT Devices in WSN using fuzzy rules
(2021) *Wirel. Pers. Commun.*,
- Ed-daoudy and KhalilMaalmi, A.
A new Internet of Things architecture for real-time prediction of various diseases using machine learning on big data environment
(2019) *Journal of Big Data*, 6.
- Manogaran, V., Vijayakumar, R.
Varatharajan, PriyanMalarvizhi Kumar, Revathi Sundarasekar & Ching-Hsien hsu, “machine Learning based big data processing framework for cancer diagnosis using hidden Markov model and GM ClusteringGunasekaran,”
(2018) *Wirel. Pers. Commun.*, 102, pp. 2099-2116.
- Jalal Laassiri, T.
Using big data-machine learning models for diabetes prediction and flight delays analytics
(2020) *Journal of Big Data*, 7.
- LirimAshiku, M.
AI-Amin, Sanjay madria, CihanDagli, “machine Learning models and big data tools for evaluating kidney acceptance,”
(2021) *Procedia Comput. Sci.*, 185, pp. 177-184.
- Pustokhin, D.A., Pustokhina, I.V., Rani, P., Kansal, V., Elhosenye, M., Joshi, G.P., Shankarg, K.
Optimal deep learning approaches and healthcare big data analytics for mobile networks toward 5G
(2021) *Comput. Electr. Eng.*, 95.
- (2005), 1, pp. 341-345.
Myoung Soo Park, Jin Hee Na, and Jin Young Choi, “PCA-based feature extraction using class information,” IEEE International Conference on Systems, Man and Cybernetics, Waikoloa, HI, USA

- Khelifi, F., Jiang, J.
K -NN regression to improve statistical feature Extraction for texture retrieval
(2011) *IEEE Trans. Image Process.*, 20 (1), pp. 293-298.
- Marinakis, Y., Marinaki, M.
Bumble bees mating optimization algorithm for the vehicle routing problem
(2011) *Handbook of Swarm Intelligence*, pp. 347-369.
- Alam, D.F., Yousri, D.A., Eteiba, M.B.
Flower Pollination Algorithm based solar PV parameter estimation
(2015) *Energy Conv. Manag.*, 101, pp. 410-422.
- de Medeiros Martins, A., Neto, A.D.D., de Melo, J.D.
Neural networks applied to classification of data based on Mahalanobis metrics
(2003) *Proc. Int. Joint Conf. Neural Net.*, 4, pp. 3071-3076.
- Angelov, P.P., Zhou, X.
Evolving fuzzy-rule-based classifiers from data streams
(2008) *IEEE Trans. Fuzzy Syst.*, 16 (6), pp. 1462-1475.
- Ghorbani, R., Ghousi, R.
Comparing different resampling methods in predicting students' performance using machine Learning techniques
(2020) *IEEE Access*, 8, pp. 67899-67911.
- Zhu, H., Wang, X., Wang, R.
Fuzzy monotonic K-Nearest neighbor versus monotonic fuzzy K-Nearest neighbor
(2022) *IEEE Trans. Fuzzy Syst.*, 30 (9), pp. 3501-3513.
- Yaghoubi, M., Eslami, M., Noroozi, M., Mohammadi, H., Kamari, O., Palani, S.
Modified salp Swarm optimization for parameter estimation of Solar PV models
(2022) *IEEE Access*, 10, pp. 110181-110194.
- Pooja Singh; Marcello Carvalho Reis; Victor Hugo C. Albuquerque
Design of Artificial Intelligence Enabled Dingo Optimizer for energy Management in 6G communication networks
(2023) *AI-Enabled 6G Networks and Applications*,
Wiley
- Woźniak, M., Wiczorek, M., Siłka, J.
BiLSTM deep neural network model for imbalanced medical data of IoT systems
(2023) *Futur. Gener. Comput. Syst.*, 141, pp. 489-499.
- Vara Siddardha Reddy, M., Sai Prasad, R., Sai Jagan, R., Selvi, M.
Artificial intelligence for IoT-based Healthcare System, 2023
(2023) *International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-5.

Correspondence Address

Zamani A.S.; Department of Computer and Self Development, Saudi Arabia; email: a.zamani@psau.edu.sa

Publisher: Elsevier Ltd

ISSN: 17468094

Language of Original Document: English

Abbreviated Source Title: Biomed. Signal Process. Control

2-s2.0-85189746466

Document Type: Article

Publication Stage: Final

Source: Scopus

ELSEVIER

Copyright © 2024 Elsevier B.V. All rights reserved. Scopus® is a registered trademark of Elsevier B.V.

 **RELX Group™**