

Recent state of art feature/attribution selection models for medical data classification

Ashok

Annai College of arts and science - Kovilacheri

ABSTRACT

Feature selection (FS) methods serves a vital role in any machine learning (ML) issue. Choice of the finest process earns optimal feature subset thus raising the accurateness and diminishing the period needed for training. In high dimensional datasets cases it becomes beneficial in eliminating the unrelated structures. This study provides a new method of surveying the familiar FS method particularly utilized in medical data categorization, through assuming the subsequent kinds of medicinal data— numerical, signals, and images. It is more helpful to authors in gathering first-hand data then they studied the numerous features like— choice of classifier, available medical datasets, FS techniques, problems in classifying the FS method, examination of major FS methods and comprehensive mechanisms thereof. It has achieved sample experiments on the standard medical datasets from UCI and examined the effects on period and presentation by using common classifiers.

Keywords: Feature selection (FS), Electronic health record, Data mining, Clinical decision support systems, Machine learning

I. INTRODUCTION

Medical data classification alludes to gaining arrangement models from clinical datasets and intends to work on the nature of medical care [1, 2]. Clinical information arrangement can be utilized for conclusion and anticipation purposes. Clinical information show novel elements including commotion coming about because of human as well as orderly mistakes, missing qualities and even scantiness. The nature of information has an enormous ramifications for the nature of the mining results [3-7]. It is important to perform preprocessing steps to eliminate or if nothing else ease a portion of the issues related with clinical information. Notwithstanding, each dataset is unique, and there is no preprocessing strategy that is best across all datasets [8-11]. Choosing the best mix of preprocessing techniques for a particular dataset is unimaginable without preliminary and examinations [12-17]. Then again, during the last many years, there has been a huge expansion in the degree of interest in the utilization of counterfeit clever instruments in medication and the arrangement of medical care [18-20]. Most of exploration endeavors in this space have involved the improvement of symptomatic devices that are utilized to give second sentiments to doctors in an analysis [21-26].

Presently, upkeep of clinical databases has turned into a critical errand in clinical field. The patient data comprising of different features and diagnostics connected with infection ought to be placed with absolute attention to detail to offer quality types of assistance [27-33]. As the data put away in clinical databases might contain missing qualities and excess data, mining of the clinical data becomes bulky. As it can influence the aftereffects of mining, it is fundamental to have great data planning and data

decrease prior to applying data mining calculations. Expectation of infection turns out to be speedy and simpler assuming data is exact and steady and liberated from commotion [34]. Feature Selection (FS) is an effective data preprocessing strategy in data mining for diminishing dimensionality of data. In clinical finding, it is vital to distinguish most critical gamble factors connected with infection. Important feature ID helps in the expulsion of superfluous, excess ascribes from the sickness dataset which, thusly, gives speedy and improved results [35].

Characterization and forecast is a data mining procedure which first purposes preparing data to foster a model and afterward the came about model is applied on testing data to come by aftereffects of expectation [36, 37]. Different order calculations have been applied on infection datasets for the conclusion of persistent sickness and the outcomes have been viewed as extremely encouraging. There is a most extreme need to foster a clever arrangement strategy which can speed up and work on the course of conclusion of ongoing sickness [38-40].

This study provides a new method of surveying the familiar FS method particularly utilized in medical data categorization, through assuming the subsequent kinds of medicinal data— numerical, signals, and images. It is more helpful to authors in gathering first-hand data then they studied the numerous features like— choice of classifier, available medical datasets, FS techniques, problems in classifying the FS method, examination of major FS methods and comprehensive mechanisms thereof. It has achieved sample experiments on the standard medical datasets from UCI and examined the effects on period and presentation by using common classifiers.

II. LITERATURE REVIEW

[41] developed DGA for FS to foster a CAD framework to analyze lung issues from chest CT cuts. The whole datasets are been parted into two kinds [42-45]. The set containing 10% of the whole dataset has been utilized as testing set. The goal work has been characterized as the amount of the squared deviation of every data in the preparation set of each class from every data in the approval set of the comparing class. GA has been utilized for FS by limiting the proposed objective capacity, coming about in the proposed DGA. The GA has been iterated north of a few ages to acquire people that are best fit regarding the goal work. Order has been performed utilizing - NN classifier to group the RoIs into one of 4 class labels. A typical precision of 88.16% with FS and a typical exactness of 86.46% without FS have been accomplished [46-50].

[51] carried out FS utilizing a covering model which utilizes the MFO calculation with the precision of - NN classification model. [52] introduced a calculation called CSO by displaying the normal way of behaving of felines. The CSO calculation thought about two organic qualities of felines, specifically, looking for mode and following mode [53-58]. Felines invest most extreme of the energy when they are conscious on resting. In any case, during their rests, their discernment is extremely high, and they are very much aware of what's going on around them. Felines constantly notice their current circumstance shrewdly and deliberately. In spite of the fact that resting, they move their position warily and gradually [59, 60]. Looking for mode has been utilized to address this way of behaving into the CSO, and the following mode has been utilized to address the way of behaving of felines progressing towards a prey into the CSO [61].

[62] presented a multitude insight calculation named KH calculation to address streamlining assignments and is focused on the impersonation of the grouping conduct of krill swarms as for exact natural and ecological cycles [63-66]. The wellness capacity of every krill individual has been characterized as minimal distance of every individual krill from food and from the most elevated thickness of the group. Three fundamental activities considered to characterize the time-subordinate place of a singular krill are, one, development prompted by other krill people, two, scrounging movement, and three, irregular dissemination [67]. The KH calculation is tried utilizing twenty benchmark works and contrasted and eight calculations. Trial and error results demonstrate that the KH calculation can outflank these natural calculations.

III. CONCLUSION

This study provides a new method of surveying the familiar FS method particularly utilized in medical data categorization, through assuming the subsequent kinds

of medicinal data— numerical, signals, and images. It is more helpful to authors in gathering first-hand data then they studied the numerous features like— choice of classifier, available medical datasets, FS techniques, problems in classifying the FS method, examination of major FS methods and comprehensive mechanisms thereof. It has achieved sample experiments on the standard medical datasets from UCI and examined the effects on period and presentation by using common classifiers.

REFERENCES

- [1] Vaiyapuri, T., & Gupta, M. (2021). Traffic accident severity prediction and cognitive analysis using deep learning. *Soft Computing*, 1-13.
- [2] Nayak, D. R., Padhy, N., Mallick, P. K., Bagal, D. K., & Kumar, S. (2022). Brain Tumour Classification Using Noble Deep Learning Approach with Parametric Optimization through Metaheuristics Approaches. *Computers*, 11(1), 10.
- [3] Vaiyapuri, T., Mohanty, S. N., Sivaram, M., Pustokhina, I. V., Pustokhin, D. A., & Shankar, K. (2021). Automatic vehicle license plate recognition using optimal deep learning model. *Computers, Materials and Continua*, 67(2), 1881-1897.
- [4] Jain, D. K., Li, Y., Er, M. J., Xin, Q., Gupta, D., & Shankar, K. (2021). Enabling Unmanned Aerial Vehicle Borne Secure Communication with Classification Framework for Industry 5.0. *IEEE Transactions on Industrial Informatics*.
- [5] Lan, X., Zhang, W., Zhang, S., Jain, D. K., & Zhou, H. (2019). Robust multi-modality anchor graph-based label prediction for RGB-infrared tracking. *IEEE Transactions on Industrial Informatics*.
- [6] Pavithran, P., Mathew, S., Namasudra, S., & Srivastava, G. (2022). A novel cryptosystem based on DNA cryptography, hyperchaotic systems and a randomly generated Moore machine for cyber physical systems. *Computer Communications*, 188, 1-12.
- [7] Chinnasamy, P., Deepalakshmi, P., Dutta, A. K., You, J., & Joshi, G. P. (2021). Ciphertext-Policy Attribute-Based Encryption for Cloud Storage: Toward Data Privacy and Authentication in AI-Enabled IoT System. *Mathematics*, 10(1), 68.
- [8] Padmaa, M., Jayasankar, T., Venkatraman, S., Dutta, A. K., Gupta, D., Shamshirband, S., & Rodrigues, J. J. (2022). Oppositional chaos game optimization based clustering with trust based data transmission protocol for intelligent IoT edge systems. *Journal of Parallel and Distributed Computing*, 164, 142-151.
- [9] Alqaralleh, B. A., Vaiyapuri, T., Parvathy, V. S., Gupta, D., Khanna, A., & Shankar, K. (2021).

- Blockchain-assisted secure image transmission and diagnosis model on Internet of Medical Things Environment. *Personal and ubiquitous computing*, 1-11.
- [10] Nair, R., Ragab, M., Mujallid, O. A., Mohammad, K. A., Mansour, R. F., & Viju, G. K. (2022). Impact of Wireless Sensor Data Mining with Hybrid Deep Learning for Human Activity Recognition. *Wireless Communications and Mobile Computing*, 2022.
- [11] Althobaiti, M. M., Almulihi, A., Ashour, A. A., Mansour, R. F., & Gupta, D. (2022). Design of Optimal Deep Learning-Based Pancreatic Tumor and Nontumor Classification Model Using Computed Tomography Scans. *Journal of Healthcare Engineering*, 2022.
- [12] Ragab, M., Eljaaly, K., Alhakamy, N. A., Alhadrami, H. A., Bahaddad, A. A., Abo-Dahab, S. M., & Khalil, E. M. (2021). Deep ensemble model for COVID-19 diagnosis and classification using chest CT images. *Biology*, 11(1), 43.
- [13] Paikrao, P., Mukherjee, A., Jain, D. K., Chatterjee, P., & Alnumay, W. (2021). Smart emotion recognition framework: A secured IOVT perspective. *IEEE Consumer Electronics Magazine*.
- [14] A. A. Eshmawi, H. Alhumyani, S. A. Khalek, R. A. Saeed, M. Ragab et al., "Design of automated opinion mining model using optimized fuzzy neural network," *Computers, Materials & Continua*, vol. 71, no.2, pp. 2543–2557, 2022.
- [15] Gupta, A., & Namasudra, S. (2022). A novel technique for accelerating live migration in cloud computing. *Automated Software Engineering*, 29(1), 1-21.
- [16] Shankar, K., Perumal, E., Díaz, V. G., Tiwari, P., Gupta, D., Saudagar, A. K. J., & Muhammad, K. (2021). An optimal cascaded recurrent neural network for intelligent COVID-19 detection using Chest X-ray images. *Applied Soft Computing*, 113, 107878.
- [17] Kumar, S., & Nezhurina, M. I. (2019). An ensemble classification approach for prediction of user's next location based on Twitter data. *Journal of Ambient Intelligence and Humanized Computing*, 10(11), 4503-4513.
- [18] Vaiyapuri, T., Dutta, A. K., Punithavathi, I. H., Duraipandy, P., Alotaibi, S. S., Alsolai, H., ... & Mahgoub, H. (2022, April). Intelligent Deep-Learning-Enabled Decision-Making Medical System for Pancreatic Tumor Classification on CT Images. In *Healthcare* (Vol. 10, No. 4, p. 677). MDPI.
- [19] Escorcia-Gutierrez, J., Gamarra, M., Beleño, K., Soto, C., & Mansour, R. F. (2022). Intelligent deep learning-enabled autonomous small ship detection and classification model. *Computers & Electrical Engineering*, 100, 107871.
- [20] Namasudra, S., Sharma, P., Crespo, R. G., & Shanmuganathan, V. (2022). Blockchain-Based Medical Certificate Generation and Verification for IoT-based Healthcare Systems. *IEEE Consumer Electronics Magazine*.
- [21] Joshi, G. P., Alenezi, F., Thirumoorthy, G., Dutta, A. K., & You, J. (2021). Ensemble of deep learning-based multimodal remote sensing image classification model on unmanned aerial vehicle networks. *Mathematics*, 9(22), 2984.
- [22] Vaiyapuri, T., Dutta, A. K., Sikkandar, M. Y., Gupta, D., Alouffi, B., Alharbi, A., ... & Kadry, S. (2022). Design of Metaheuristic Optimization-Based Vascular Segmentation Techniques for Photoacoustic Images. *Contrast Media & Molecular Imaging*, 2022.
- [23] Dutta, A. K. (2021). Detecting phishing websites using machine learning technique. *PLoS one*, 16(10), e0258361.
- [24] Pustokhina, I. V., Pustokhin, D. A., Lydia, E. L., Elhoseny, M., & Shankar, K. (2021). Energy efficient neuro-fuzzy cluster based topology construction with metaheuristic route planning algorithm for unmanned aerial vehicles. *Computer Networks*, 196, 108214.
- [25] Manjari, K., Verma, M., Singal, G., & Namasudra, S. (2022). QUEST: Quantized and Efficient Scene Text Detector using Deep Learning. *Transactions on Asian and Low-Resource Language Information Processing*.
- [26] Ismail, S. S., Mansour, R. F., El-Aziz, A., Rasha, M., & Taloba, A. I. (2022). Efficient E-Mail Spam Detection Strategy Using Genetic Decision Tree Processing with NLP Features. *Computational Intelligence and Neuroscience*, 2022.
- [27] Vaiyapuri, T., Parvathy, V. S., Manikandan, V., Krishnaraj, N., Gupta, D., & Shankar, K. (2021). A novel hybrid optimization for cluster-based routing protocol in information-centric wireless sensor networks for IoT based mobile edge computing. *Wireless Personal Communications*, 1-24.
- [28] Shankar, K., Mohanty, S. N., Yadav, K., Gopalakrishnan, T., & Elmisery, A. M. (2021). Automated COVID-19 diagnosis and classification using convolutional neural network with fusion based feature extraction model. *Cognitive Neurodynamics*, 1-14.
- [29] Tyagi, S. K. S., Mukherjee, A., Boyang, Q., & Jain, D. K. (2021). Computing resource optimization of big data in optical cloud radio access networked industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, 17(11), 7734-7742.
- [30] Vaiyapuri, T. (2021). Deep learning enabled autoencoder architecture for collaborative filtering recommendation in iot environment. *CMC-COMPUTERS MATERIALS & CONTINUA*, 68(1), 487-503.

- [31] Namasudra, S., Chakraborty, R., Majumder, A., & Moparthy, N. R. (2020). Securing multimedia by using DNA-based encryption in the cloud computing environment. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 16(3s), 1-19.
- [32] Nair, R., Ragab, M., Mujallid, O. A., Mohammad, K. A., Mansour, R. F., & Viju, G. K. (2022). Impact of Wireless Sensor Data Mining with Hybrid Deep Learning for Human Activity Recognition. *Wireless Communications and Mobile Computing*, 2022.
- [33] Shankar, K., Perumal, E., Tiwari, P., Shorfuzzaman, M., & Gupta, D. (2021). Deep learning and evolutionary intelligence with fusion-based feature extraction for detection of COVID-19 from chest X-ray images. *Multimedia Systems*, 1-13.
- [34] Sekhar, P., Lydia, E. L., Elhoseny, M., Al-Akaidi, M., Selim, M. M., & Shankar, K. (2021). An effective metaheuristic based node localization technique for wireless sensor networks enabled indoor communication. *Physical Communication*, 48, 101411.
- [35] Namasudra, S. (2020). Fast and secure data accessing by using DNA computing for the cloud environment. *IEEE Transactions on Services Computing*.
- [36] Mukherjee, A., Jain, D. K., & Yang, L. (2020). On-demand efficient clustering for next generation IoT applications: A hybrid NN approach. *IEEE Sensors Journal*, 21(22), 25457-25464.
- [37] Mukherjee, A., Jain, D. K., Goswami, P., Xin, Q., Yang, L., & Rodrigues, J. J. (2020). Back propagation neural network based cluster head identification in MIMO sensor networks for intelligent transportation systems. *IEEE Access*, 8, 28524-28532.
- [38] Althobaiti, M. M., Kumar, K. P. M., Gupta, D., Kumar, S., & Mansour, R. F. (2021). An intelligent cognitive computing based intrusion detection for industrial cyber-physical systems. *Measurement*, 186, 110145.
- [39] M. Ragab, A. Algarni, A. A. Bahaddad and R. F. Mansour, "Machine learning applied to problem-solving in medical applications," *Computers, Materials & Continua*, vol. 69, no.2, pp. 2277–2294, 2021.
- [40] Das, A. K., Mishra, D., Das, K., Mallick, P. K., Kumar, S., Zymbler, M., & El-Sayed, H. (2022). Prophesying the Short-Term Dynamics of the Crude Oil Future Price by Adopting the Survival of the Fittest Principle of Improved Grey Optimization and Extreme Learning Machine. *Mathematics*, 10(7), 1121.
- [41] Mansour, R. F., Escorcía-Gutiérrez, J., Gamarra, M., Gupta, D., Castillo, O., & Kumar, S. (2021). Unsupervised deep learning based variational autoencoder model for COVID-19 diagnosis and classification. *Pattern Recognition Letters*, 151, 267-274.
- [42] Dutta, A. K., Elhoseny, M., Dahiya, V., & Shankar, K. (2020). An efficient hierarchical clustering protocol for multihop Internet of vehicles communication. *Transactions on Emerging Telecommunications Technologies*, 31(5), e3690.
- [43] Swati, S., Kumar, M., & Namasudra, S. (2022). Early prediction of cognitive impairments using physiological signal for enhanced socioeconomic status. *Information Processing & Management*, 59(2), 102845.
- [44] Jain, D. K., Zareapoor, M., Jain, R., Kathuria, A., & Bachhety, S. (2020). GAN-Poser: an improvised bidirectional GAN model for human motion prediction. *Neural Computing and Applications*, 32(18), 14579-14591.
- [45] Al-Wesabi, F. N., Obayya, M., Hamza, M. A., Alzahrani, J. S., Gupta, D., & Kumar, S. (2022). Energy Aware Resource Optimization using Unified Metaheuristic Optimization Algorithm Allocation for Cloud Computing Environment. *Sustainable Computing: Informatics and Systems*, 35, 100686.
- [46] Lian, J., Jia, W., Zareapoor, M., Zheng, Y., Luo, R., Jain, D. K., & Kumar, N. (2019). Deep-learning-based small surface defect detection via an exaggerated local variation-based generative adversarial network. *IEEE Transactions on Industrial Informatics*, 16(2), 1343-1351.
- [47] Ragab, M., Alshehri, S., Aldawsari, H. M., Noor, A., Ashary, E. B., Abou-Taleb, S. A. K., & Abdelazim, G. (2022). COVID-19 Identification System Using Transfer Learning Technique With Mobile-NetV2 and Chest X-Ray Images. *Frontiers in Public Health*, 102.
- [48] Agrawal, D., Minocha, S., Namasudra, S., & Gandomi, A. H. (2022). A robust drug recall supply chain management system using hyperledger blockchain ecosystem. *Computers in biology and medicine*, 140, 105100.
- [49] Alqaralleh, B. A., Mohanty, S. N., Gupta, D., Khanna, A., Shankar, K., & Vaiyapuri, T. (2020). Reliable multi-object tracking model using deep learning and energy efficient wireless multimedia sensor networks. *IEEE Access*, 8, 213426-213436.
- [50] Ragab, M., Albukhari, A., Alyami, J., & Mansour, R. F. (2022). Ensemble Deep-Learning-Enabled Clinical Decision Support System for Breast Cancer Diagnosis and Classification on Ultrasound Images. *Biology*, 11(3), 439.
- [51] Namasudra, S., & Roy, P. (2017). Time saving protocol for data accessing in cloud computing. *IET Communications*, 11(10), 1558-1565.

- [52] Mansour, R. F., Escorcia-Gutierrez, J., Gamarra, M., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence with big data analytics-based brain intracranial hemorrhage e-diagnosis using CT images. *Neural Computing and Applications*, 1-13.
- [53] Pustokhina, I. V., Pustokhin, D. A., Aswathy, R. H., Jayasankar, T., Jeyalakshmi, C., Díaz, V. G., & Shankar, K. (2021). Dynamic customer churn prediction strategy for business intelligence using text analytics with evolutionary optimization algorithms. *Information Processing & Management*, 58(6), 102706.
- [54] Shankar, K., Perumal, E., Elhoseny, M., Taher, F., Gupta, B. B., & El-Latif, A. A. A. (2021). Synergic Deep Learning for Smart Health Diagnosis of COVID-19 for Connected Living and Smart Cities. *ACM Transactions on Internet Technology (TOIT)*, 22(3), 1-14.
- [55] Kumar, S., Toshniwal, D., & Parida, M. (2017). A comparative analysis of heterogeneity in road accident data using data mining techniques. *Evolving systems*, 8(2), 147-155.
- [56] Shankar, K., Taniar, D., Yang, E., & Yi, O. (2021). Secure and Optimal Secret Sharing Scheme for Color Images. *Mathematics*, 9(19), 2360.
- [57] Metawa, N., Nguyen, P. T., Le Hoang Thuy To Nguyen, Q., Elhoseny, M., & Shankar, K. (2021). Internet of things enabled financial crisis prediction in enterprises using optimal feature subset selection-based classification model. *Big Data*, 9(5), 331-342.
- [58] M. Ragab, E. B. Ashary, M. F. S. Sabir, A. A. Bahaddad and R. F. Mansour, "Mathematical modelling of quantum kernel method for biomedical data analysis," *Computers, Materials & Continua*, vol. 71, no.3, pp. 5441–5457, 2022.
- [59] Dutta, A. K., Aljarallah, N. A., Abirami, T., Sundarajan, M., Kadry, S., Nam, Y., & Jeong, C. W. (2022). Optimal Deep-Learning-Enabled Intelligent Decision Support System for SARS-CoV-2 Classification. *Journal of Healthcare Engineering*, 2022.
- [60] Ragab, M., Ashary, E. B., Aljedaibi, W. H., Alzahrani, I. R., Kumar, A., Gupta, D., & Mansour, R. F. (2022). A novel metaheuristics with adaptive neuro-fuzzy inference system for decision making on autonomous unmanned aerial vehicle systems. *ISA Transactions*.
- [61] Zhou, S., Luo, P., Jain, D. K., Lan, X., & Zhang, Y. (2019). Double-domain imaging and adaption for person re-identification. *IEEE Access*, 7, 103336-103345.
- [62] Sharma, P., Moparthi, N. R., Namasudra, S., Shanmuganathan, V., & Hsu, C. H. (2021). Blockchain-based IoT architecture to secure healthcare system using identity-based encryption. *Expert Systems*, e12915.
- [63] Jain, D. K., Boyapati, P., Venkatesh, J., & Prakash, M. (2022). An Intelligent Cognitive-Inspired Computing with Big Data Analytics Framework for Sentiment Analysis and Classification. *Information Processing & Management*, 59(1), 102758.
- [64] Vaiyapuri, T., Lydia, E. L., Sikkandar, M. Y., Díaz, V. G., Pustokhina, I. V., & Pustokhin, D. A. (2021). Internet of Things and Deep Learning Enabled Elderly Fall Detection Model for Smart Homecare. *IEEE Access*, 9, 113879-113888.
- [65] Jain, D. K., Tyagi, S. K. K. S., Neelakandan, S., Prakash, M., & Natrayan, L. (2021). Metaheuristic optimization-based resource allocation technique for cybertwin-driven 6G on IoE environment. *IEEE Transactions on Industrial Informatics*.
- [66] Dutta, A. K. (2020). Managing natural hazards in smart cities in Kingdom of Saudi Arabia using a technique based on interior search algorithm. *Electronic Government, an International Journal*, 16(1-2), 155-169.
- [67] Mansour, R. F., El Amraoui, A., Nouaouri, I., Díaz, V. G., Gupta, D., & Kumar, S. (2021). Artificial intelligence and Internet of Things enabled disease diagnosis model for smart healthcare systems. *IEEE Access*, 9, 45137-45146.