FEATURE EXTRACTION OF INDICATOR CARD DATA USING MACHINE LEARNING

Krishna Dharavath,
Department of Electronics and Communication Engineering, Vardhaman College of Engineering, Shamshabad, 501218, Hyderabad, India

Abstract:
In this article, three feature extraction techniques for sucker-rod pump indication card data based on Fourier Descriptors (FD), Geometric Moment Vector (GMV), and Grey Level Matrix Statistics (GLMX) have been investigated, simulated, and compared. Due to the non-optimal amount of Fourier Descriptors used in the technique, the Fourier Descriptors algorithm may result in information loss in numerical tests. The Geometric Moment. While the Grey Level Matrix Statistics approach produces low-dimension feature vectors with greater time consumption and memory space, it also consumes more time and resources. Additionally, the Fourier Descriptors approach and the Geometric Moment Vector algorithm's rotational invariance property may lead to incorrect pattern identification of indicator card data when utilised for sucker-rod pump operating condition diagnostics.

1. Introduction

The most popular artificial lift method for onshore oil well production is the sucker-rod pump system [1-3]. Sucker-rod pumps are used to produce oil in around 80% of wells worldwide and 90% of those in China [4, 5]. A sucker-rod pump system requires expensive and time-consuming maintenance and optimization. The relation curve between the load and the displacement of a sucker-rod pump in an unbroken suck cycle is the indication card [6], where the -axis denotes displacement and the -axis the load. The indication card is useful for analysing the sucker-rod pump wells' down-hole operating conditions [7], as it can assess the well's operational state and offer trustworthy evidence of high efficiency, acceptable exploitation for oil well output. The card can display a form that may represent a regular functioning or a failure scenario when the system is in use. The approaches for defect diagnosis and pattern identification are used to recognise various curve forms based on various types of real-time indicator card data. Identify the type of anomalous condition that is present and determine the cause of the problem [8]. Therefore, it is crucial to accurately and quickly identify the sucker-rod pump indication card in order to diagnose any problems with the down-hole operating condition.

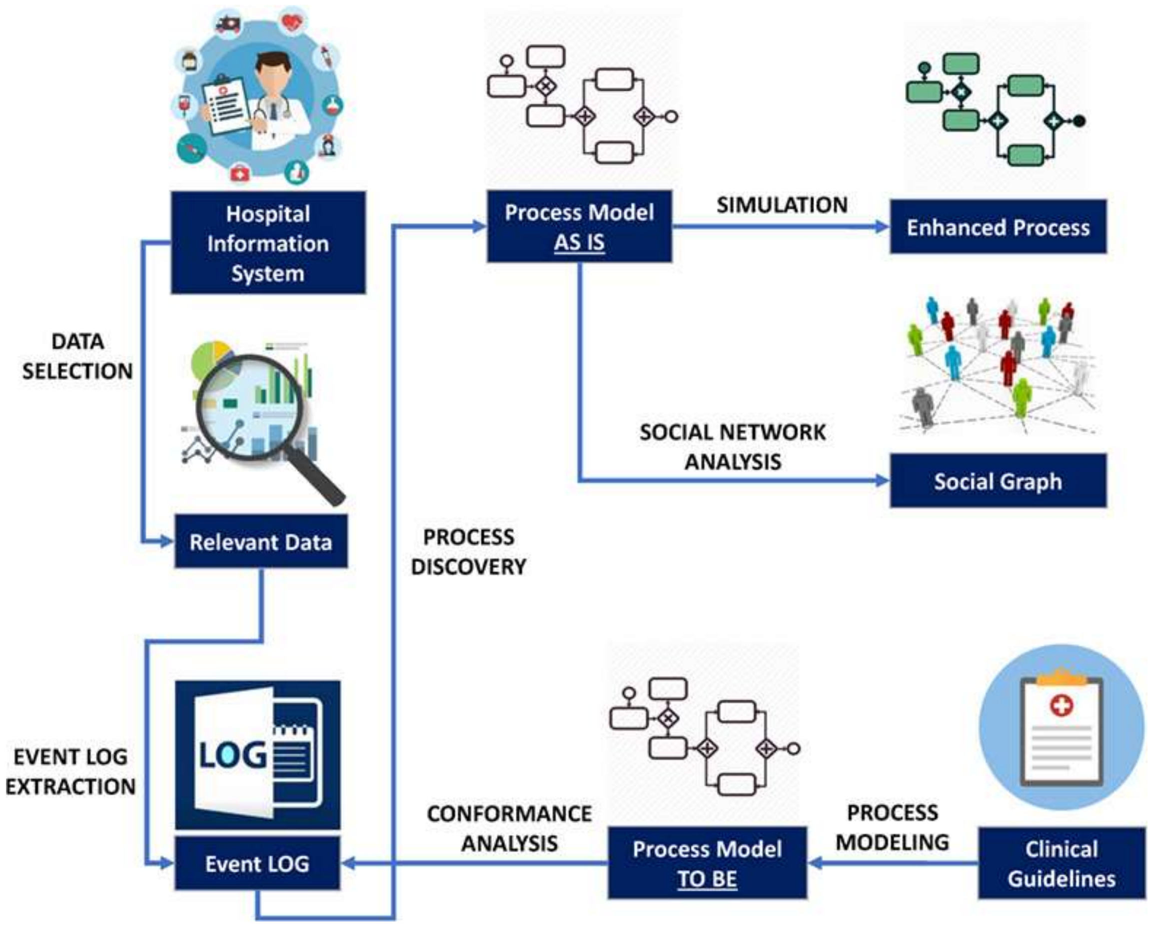

Fig.1: Feature Extraction of Indicator Card Data using Machine Learning Flow Chart

Algorithm Based on Fourier Descriptors

A common technique for picture reconstruction and classification is the Fourier transformation. The Fourier Descriptors (FD), which represent the object form in a frequency domain, are created as a full collection of complex integers. The superfluous indication card points (data) are removed using the polygonal approximation approach to lessen the computational complexity. The steps are listed below. In order to choose the feature pixels (data) of the polygon that satisfy the requirement with the highest curvature of a specified length curve, we first traverse all the digital pixels (data) of the indicator card curve in accordance with a given value, which is called.We then add the feature data to an array. According to a comparison of several calculating methods, the value () is given a value of 0.008. Take the pump-on-touch defect as an example. From the original 702 pixels, 34 feature pixels (data) are retrieved.

Algorithm Based on Gray Level Matrix Statistics

One of the classic techniques for extracting visual features is the grid approach. The grid method's processing phases are listed below. First, in the horizontal and vertical directions, respectively, we partition the picture of the indication card into a number of little grids of the same size and form. The grids that the indication card's curve travels through are then marked.

Finally, we can get the indication card image's feature parameters. In this study, we employ the grid-based Grey Level Matrix Statistics (GLMS) feature extraction approach.

The indication card curve should be transformed into a grayscale visual matrix prior to the GLMS feature extraction. The GLMS feature extraction algorithm has the following phases. The indication card curve traverses a grid to initialize the grayscale matrix mesh, in which case the grey value is set to "1". Other grids are given a grey value in accordance with the grey contour principle: if the grid is inside the curve, the grey value is equal to the initial value plus the grid's distance from the curve; if the grid is outside the curve, the grey value is equal to the initial value minus the grid's distance from the curve.

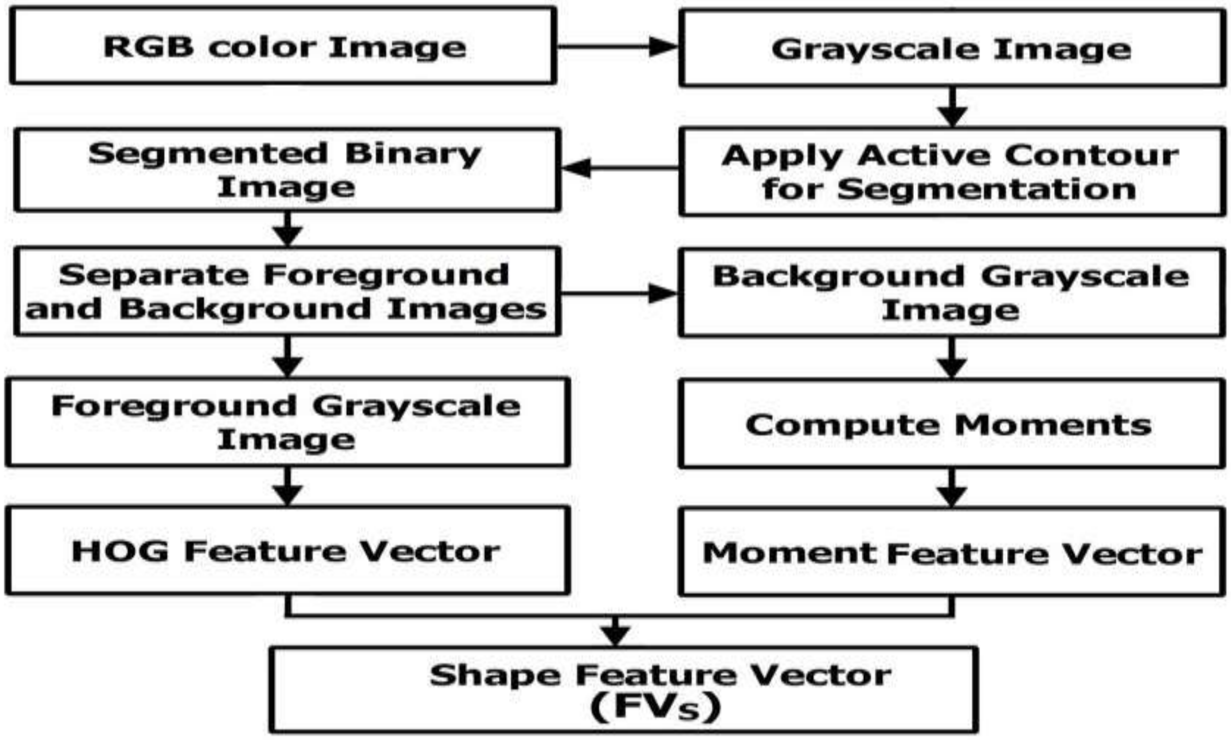

Fig.2: Feature Extraction of Indicator Card Data using Machine Learning Process.

Finally, we can get 6 statistic parameters of the gray level matrix [19]. In this article, three feature extraction techniques for sucker-rod pump indication card data based on Fourier Descriptors (FD), Geometric Moment Vector (GMV), and Grey Level Matrix Statistics (GLMX) have been investigated, simulated, and compared. The Fourier Descriptors approach takes less memory space and running time, according to numerical testing, however there may be information loss owing to non-optimal Fourier coefficients. Descriptors, the Geometric Moment Vector approach consumes more time and memory while producing low-dimension feature vectors, and the Grey Level Matrix Statistics algorithm consumes more time and memory while producing the same results. Additionally, the Fourier Descriptors approach and the Geometric Moment Vector algorithm's rotational invariance property may lead to incorrect pattern identification of indicator card data when utilised for sucker-rod pump operating condition diagnostics.

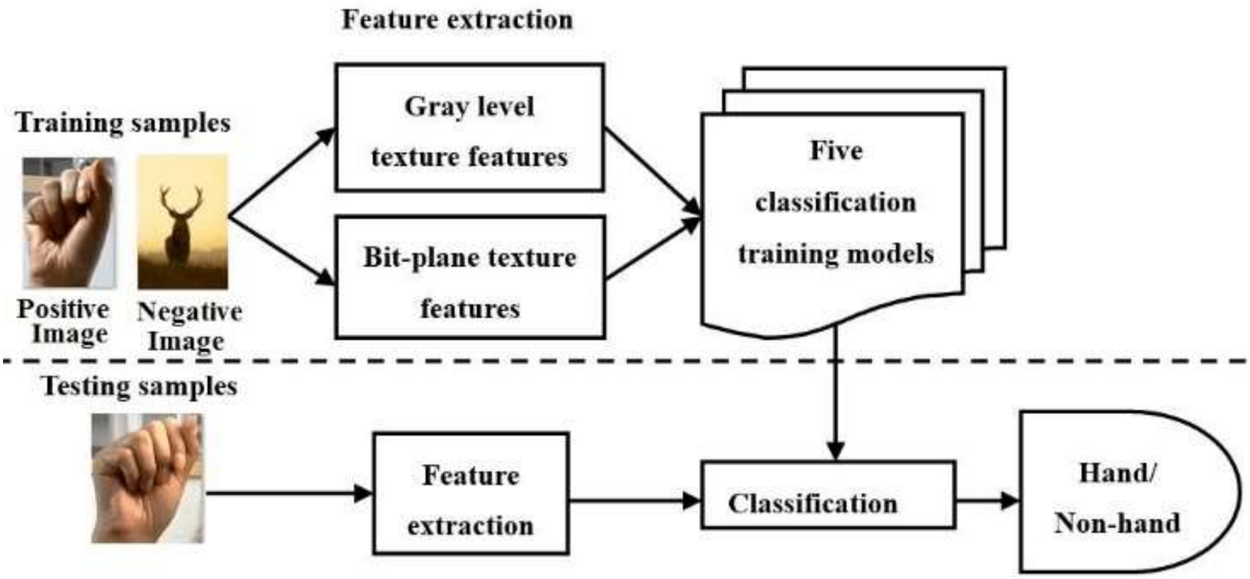

Fig.3: Feature Extraction of Indicator Card Data using Machine Learning Method.

In this article, three feature extraction techniques for sucker-rod pump indication card data based on Fourier Descriptors (FD), Geometric Moment Vector (GMV), and Grey Level Matrix Statistics (GLMX) have been investigated, simulated, and compared. According to numerical testing, the Geometric Moment Vector approach is faster and uses less memory than the Fourier Descriptors algorithm, with the possibility of information loss owing to non-optimal amounts of Fourier Descriptors. While the Grey Level Matrix Statistics approach produces low-dimension feature vectors with increased time and memory use, it also produces low-dimension feature vectors. Additionally, the Fourier Descriptors approach and the Geometric Moment Vector algorithm's rotational invariance property may lead to incorrect pattern identification of indicator card data when utilised for sucker-rod pump operating condition diagnostics.

2. Methods and Materials

The extraction function reduces the amount of tools required to correctly evaluate huge data sets. One of the major issues with advanced data analytics is the enormous number of variables involved. If you needed an algorithm to categorise the exhibits and utilise them for fresh samples, it would be beneficial if you had enough of memory and processing capacity for doing so. In order to approach these issues with the utmost accuracy, feature extraction is a broad idea that may be used to generate dynamic mixtures of variables. Texture analysis looked for a precise, simpler approach to capture the unique characteristics of textures for accurate item recognition and division. In order to analyse images and recognise patterns, the surface is crucial. Only a few processing architectures make advantage of texture extraction [7][5]. This work creates a grey surface incidence matrix to produce statistical texture features. The statistical light intensity distribution of certain points in the statistical texture analysis relative to one another is used to compute the texture qualities of the detected substances. According on the amount of pixels in each combination, the numbers are divided into first, second, and higher order categories. A method for calculating statistical texture properties of the second order is the grey level matrix (GLCM).

3. Results

The collection contains COVID-19-enhanced chest X-ray images for illness identification. Two publicly available datasets were used to get the data [87, 88]. The information is based on a publicly available dataset of chest CT images and X-rays of patients with COVID-19 or other viral or bacterial pneumonia who have positive or suspect (MERS, SARS, and ARDS) results. The direct sources of the data would be hospitals and doctors, while the indirect sources would be public records. The public has access to all images and information [4][6, 5][0], and 5][3]. PNG type images have a 256 256 pixel size. The illustration is displayed. 1824 images are utilised for analysis and simulation of the algorithms, of which 80% are training samples and 20% are validation samples. With the development of the smart oilfield, several deep learning technologies are being used to recognise the graphic characteristic of the indication diagram in order to identify the type of rod pump system malfunction and maintain the oilfield's regular output. The source data for the indicator diagrams from different oil fields, however, are impacted by diverse geographic circumstances, sensor equipment, acquisition software, etc. and display particular environmental features.

This makes using fault detection approaches based on indication diagrams challenging and calls for the employment of a more universal diagnosis paradigm. A multi-feature fusion fault diagnosis model is suggested as a solution to the problem. To increase the feature's robustness, the model combines the visual feature with the Fourier descriptor of the indication diagram as a feature. First, the two backbone networks collect features from their own networks' single-modal input data. In addition, the data from the indicator diagram and interactive fusion module teaches Fourier descriptor as features in tandem. Finally, feature categorization based on integrated features is employed to generate the network output. According to the results of the validation experiment, the accuracy of the diagnostic model when using only one feature is, respectively, 0.8233 (the graphic feature), 0.9422 (the Fourier descriptor feature), and 0.9724 when utilising the fusion of the two features. The outcomes show that the multi-input feature fusion model is superior than the single-input model in terms of performance. The method makes advantage of the characteristics' connection to realize their complimentary benefits and improve the diagnostic model's efficacy.

4. Conclusion

Three alternative feature extraction techniques, each based on a different Fourier descriptor, geometric moment vector, and grey level matrix statistic, have been examined and simulated in this article. The three methods' memory requirements and processing performance. The GMV algorithm is more time-consuming and memory-intensive due to different numbers of FDs, whereas the GLMS algorithm produces low-dimension vectors with good performance of speed and space. Numerical experiments demonstrate that the FD algorithm is with high computing speed and more memory space but possible loss of information. When employed for sucker-rod pump operating condition diagnostics, the rotational invariance property of the FD method and the GMV algorithm may result in inaccurate pattern identification of indicator card data. For greater performance, more study on feature extraction from indicator card data has to be done. The development of medical image processing has hastened the creation of intelligent detection and diagnosis software. Machine learning algorithms are widely acknowledged as a potent technique for enhancing illness diagnosis precision. To get stronger machine learning models, though, you need effective feature extraction tools. As a consequence of its ability to automatically retrieve features or make use of previously trained models, deep learning models are frequently utilised in medical imaging applications. The results of this study, which employs machine learning techniques to identify COVID-19 on chest X-rays, are highly encouraging. Using GLCM techniques, the feature extractions are carried out. KNN, SVM, LDA, NB, and CNN algorithms are utilised to categorise patients.

References

1. T. Ai, Z. Yang, H. Hou et al., "Correlation of chest CT and RT-PCR testing in coronavirus disease 2019 (COVID-19) in China: a report of 1014 cases," Radiology, vol. 296, Article ID 200642, 2020.
2. Y. Fang, H. Zhang, J. Xie et al., "Sensitivity of chest CT for COVID-19: comparison to RT-PCR," Radiology, vol. 296, no. 2, pp. E115–E117, 2020.
3. H. Zeng, C. Xu, J. Fan et al., "Antibodies in infants born to mothers with COVID-19 pneumonia," JAMA, vol. 18, pp. 1848-1849, 2020.
4. D. A. Schwartz, "An analysis of 38 pregnant women with COVID-19, their newborn infants, and maternal-fetal transmission of SARS-CoV-2: maternal coronavirus infections and pregnancy outcomes," Archives of Pathology & Laboratory Medicine, vol. 144, no. 7, pp. 799–805, 2020.
5. T. Ozturk, M. Talo, E. A. Yildirim, U. B. Baloglu, O. Yildirim, and U. Rajendra Acharya, "Automated detection of COVID-19 cases using deep neural networks with X-ray images," Computers in Biology and Medicine, vol. 121, Article ID 103792, 2020.
6. M. Li, P. Lei, B. Zeng et al., "Coronavirus disease (COVID-19): spectrum of CT findings and temporal progression of the disease," Academic Radiology, vol. 27, no. 5, pp. 603–608, 2020.
7. F. Shi, J. Wang, J. Shi, and W. Ziyan, "Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19," IEEE Rev Biomed Eng, vol. 14, pp. 1-2, 2020.
8. H. S. Maghdid, A. T. Asaad, K. Z. Ghafoor, A. S. Sadiq, and M. K. Khan, "Diagnosing COVID-19 pneumonia from X-ray and CT images using deep learning and transfer learning algorithms," 2020,
9. A. F. de Moraes Batista, J. L. Miraglia, T. H. R. Donato, and A. D. P. C. Filho, "COVID-19 diagnosis prediction in emergency care patients: a machine learning approach," medRxiv, 2020,

10. G. Hripcsak and D. J. Albers, "Next-generation phenotyping of electronic health records," Journal of the American Medical Informatics Association, vol. 20, no. 1, pp. 117–121, 2013.
11. P. B. Jensen, L. J. Jensen, and S. Brunak, "Mining electronic health records: towards better research applications and clinical care," Nature Reviews Genetics, vol. 13, no. 6, pp. 395–405, 2012.
12. J. Luo, M. Gopukumar, D. Zhao, and W. Min, "Big data application in biomedical research and health care: a literature review," Biomedical Informatics Insights, vol. 8, Article ID BII.S31559, 2016.
13. M. Ahmadi, A. Sharifi, and S. Khalili, "Presentation of a developed sub-epidemic model for estimation of the COVID-19 pandemic and assessment of travel-related risks in Iran," Environmental Science and Pollution Research, vol. 28, no. 12, Article ID 14521, 2020.
14. S. Hassantabar, M. Ahmadi, and A. Sharifi, "Diagnosis and detection of infected tissue of COVID-19 patients based on lung X-ray image using convolutional neural network approaches," Chaos, Solitons & Fractals, vol. 140, Article ID 110170, 2020.
15. M. H. Nasirpour, A. Sharifi, M. Ahmadi, and S. Jafarzadeh Ghoushchi, "Revealing the relationship between solar activity and COVID-19 and forecasting of possible future viruses using multi-step autoregression (MSAR)," Environmental Science and Pollution Research, vol. 28, no. 28, Article ID 38074, 2021.