POLITECNICO DI TORINO
Repository ISTITUZIONALE

The Design of the Principal Component Analysis (PCA)-based Fuzzy Logic Classifier on Physical Fatigue in Process Industries

Original
The Design of the Principal Component Analysis (PCA)-based Fuzzy Logic Classifier on Physical Fatigue in Process Industries / Shi, H., ALBARRAN MORILLO, C., Baldissone, G., Demichela, M.. - In: CHEMICAL ENGINEERING TRANSACTIONS. - ISSN 2283-9216. - 111:(2024), pp. 211-216. [10.3303/CET24111036]

Availability:
This version is available at: 11583/2993405 since: 2024-10-15T08:54:58Z

Publisher:
AIDIC

Published
DOI:10.3303/CET24111036

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

01 November 2026

CHEMICAL ENGINEERING TRANSACTIONS
VOL. 111, 2024
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Valerio Cozzani, Bruno Fabiano, Genserik Reniers
Copyright © 2024, AIDIC Servizi S.r.l.
ISBN 979-12-81206-11-3; **ISSN** 2283-9216
DOI: 10.3303/CET24111036

The Design of the Principal Component Analysis (PCA)-based Fuzzy Logic Classifier on Physical Fatigue in Process Industries

Huxiao Shi*, Carlos Albarran Morillo, Gabriele Baldissone, Micaela Demichela

Applied Science and Technology Department, Politecnico di Torino, Turin, Italy
huxiao.shi@polito.it

Currently, Physical Fatigue Management is considered a vital part of process industries because of the diverse work scenarios, the high degree of non-standardized operations, and the high reliance on manual labor. The way an individual's performance deteriorates with the accumulation of fatigue can vary based on both the worker and the workplace conditions. However, the widely used method for assessing physical fatigue, the 'Borg Rating of Perceived Exertion Scale', has several limitations. These include high subjectivity, introducing more variability among individuals in a population, and weakness in dynamic measurements, potentially missing the optimal recovery time for operators. In this study, a Principal Component Analysis (PCA)-based Fuzzy Logic Classifier has been designed to aid the development of customized warning systems for physical fatigue, and also support the intelligent decision-making process in process industries.

1. Introduction

The human element plays a vital role in the consideration of safety and risk management according to the analysis of numerous accident reports within process industries (Yang and Demichela, 2023). Physical fatigue is considered one of the main contributors to human error in industrial incidents and accidents, especially in petrochemical, oil and gas plant's operations (Mariana et al., 2018). The stresses from repetitive manual labor and demanding tasks stimulate physical fatigue, which hinders the development of effectiveness, safety and even employee well-being in workplaces. Currently, the adoption of Physical Fatigue Management is suggested to follow four phases as the framework: detection, identification, diagnosis and recovery, and consider the big potential of applying wearable sensors in dynamic monitoring (Sedighi Maman et al., 2020). In industrial domains, particularly in the process industry which features multiple production patterns, non-standard operations and complex workplaces involved, it is more valuable to achieve dynamic monitoring.
Combining wearable devices with Physical Fatigue Management becomes more feasible and useful with their high availability right now, and strength in real-time measuring of physical fatigue in the operational environment while minimally influencing the primary job (Morillo and Demichela, 2023). There exist several wearable devices available in the current market that allow to record the physiological parameters of the users. However, there are still several difficulties faced during the practical application of the wearable devices:

- Process multiple variables recorded by wearable devices, in other words, the high dimensional dataset.
- Build the connection between the monitored data and physical fatigue.
- Address the variability from high subjectivity of participants when doing the physical fatigue test by 'Borg Rating of Perceived Exertion Scale' (shortened in 'Borg Scale').
- Guide the decision-making process with more intelligent and customized strategies.

In this paper, these limitations are addressed through the design of a Principal Component Analysis (PCA)-based fuzzy logic classifier. The designed classifier is regarded as beneficial to support the development of customized warning systems and the intelligent decision-making process in process industries.

Paper Received: 26 February 2024; Revised: 10 June 2024; Accepted: 17 June 2024
Please cite this article as: Shi H., Albarran Morillo C., Baldissone G., Demichela M., 2024, The Design of the Principal Component Analysis (PCA)-based Fuzzy Logic Classifier on Physical Fatigue in Process Industries, Chemical Engineering Transactions, 111, 211-216 DOI:10.3303/CET24111036

2. Methodology

In this section, the built database, research framework and applied methods are introduced.

2.1 Database built

The database analyzed in this research involves condition-monitored data for 33 participants (21 males and 13 females) recruited voluntarily from a fitness facility, with ages ranging from 21 to 41 years. The mean age of the participants was 25.6 ± 4.4 years. This study adhered to the Declaration of Helsinki guidelines. Prior to data collection, each participant received an informed consent form that included detailed information on the study's nature, potential benefits, risks, and alternatives.

The study uses a fitness setup to simulate industrial tasks safely and effectively. It uses weightlifting, resistance band exercises, isometric exercises, and bodyweight exercises. Weightlifting targets muscle groups involved in lifting, pushing, and pulling heavy objects. Resistance band exercises simulate tasks requiring pulling and pushing actions. Isometric exercises replicate holding positions suitable for mobility limitations or injury recovery. Participants can choose exercises based on fitness levels and preferences.

During this process, a smartwatch is deployed to support dynamically monitoring operators' performance parameters, where the data about electrodermal activity, skin temperature, pulse rate, accelerometer data, step counts and activity counts are collected. The 'Borg Scale' is applied to define whether participants are in states of (non-)fatigue. This approach allows people to subjectively assess their degree of physical exertion on a scale of 6 (no exertion) to 20 (highest intensity). Finally, 502 records are collected as the database.

2.2 Research framework and methods

Principal Component Analysis (PCA) is proven to have strengths in optimizing the selected feature indicators in the domain of intelligent monitoring and fatigue state detection (Chen et al., 2023). With applied multiple variables which may increase system complexity, PCA helps to convert those potentially correlated variables into a set of linearly uncorrelated variables through orthogonal transformation, which are called principal components. The information that principal components have compared with original multiple variables is explained by the term of a sum of squares of deviations or variances. Thus, PCA is broadly applied to achieve dimensionality reduction, and decrease the complexity of the target system (Chen et al., 2021).

Fuzzy logic is an approach to support reasoning under uncertainty and partial information and facilitate modeling logical reasoning with imprecise or vague statements (Nguyen et al., 2023). The degree of belonging of an element x to a fuzzy set A is quantified by a membership function $F_A(x)$, providing each element with a value between 0 and 1 according to Eq (1), where X denotes the universe of discourse, and the value given to each element by the membership function is called membership value.

$$A = \{x, F_A(x) \mid x \in X\} \tag{1}$$

Fuzzy logic has the potential to analyze cases where one variable is vaguely divided into categorical categories (Baldissone et al., 2019). In this paper, fuzzy logic facilitates defining the vague zone between fatigue and non-fatigue categories.Figure 1 shows the framework and methods of this research. Firstly, face to the multiple dimensional performance parameters, Principal Component Analysis (PCA) is applied to extract key features related to physical fatigue. In order to solve the variability caused by individual differences and subjectivity of participants when conducting the Borg Test, the Fuzzy Set Theory is considered to distinguish the state between physical fatigue and non-physical fatigue. Then the PCA-based fuzzy logic classifier is accordingly designed. The benefits of the designed classifier are discussed in the last part.

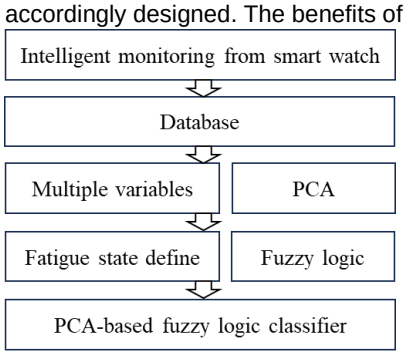


Figure 1. Research framework

3. Result

The designed classifier with all results is shown in this section.

3.1 Feature extraction

Firstly, KMO and Bartlett's Test was conducted to check whether there is a correlation between variables in the data set, the result from this method supports the feasibility and suitability of feature-extracted techniques like PCA. According to Table 1, the KMO value is 0.761 with a significant value of <0.001***, which indicates that the chi-square value is statistically significant, meaning that the null hypothesis can be rejected (that the variables are not completely independent). Thus, this database is considered to be suitable for PCA.

Table 1: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy		0.761
Bartlett's Test of Sphericity	Approx. Chi-Square	1111.187
	Degree of freedom	15
	Significance	<0.001***

Table 2 and Figure 2 show the PCA results from the analysis. The choice of component numbers is based on their eigenvalues whose value is greater than 1. In this study, components 1 and 2 (PC1 and PC2) are selected with 66.079% of the variance could be explained. Considering the high complexity of the original database, 66% is regarded to be enough to remain the key feature of the database. Table 3 provides the component matrix about each variable's coefficient, which explains how each feature contributes to two components. In this way, the original 6-dimensional features could be reduced to a 2-dimensional feature variable.

Table 2: Results of PCA

	Initial Eigenvalues			Extraction Sums of Squared Loadings		
Component	Total	Explained variance %	Cumulative variance %	Total	Explained variance %	Cumulative variance %
1	2.900	48.338	48.338	2.900	48.338	48.338
2	1.064	17.741	66.079	1.064	17.741	66.079
3	0.954	15.901	81.979			
4	0.569	9.476	91.455			
5	0.292	4.862	96.317			
6	0.221	3.683	100.000			

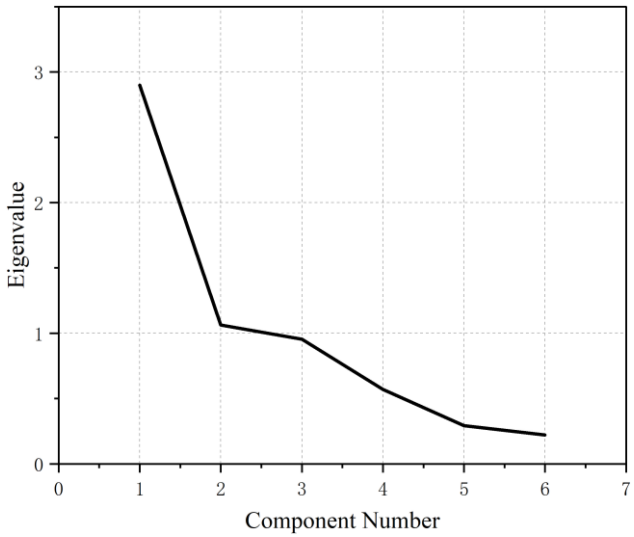


Figure 2: Scree Plot which supports to choose the number of components

Table 3: Component matrix with the loading of each variable

Variable	Component	
	1	2
Electrodermal activity	0.456	-0.626
Pulse rate	0.657	-0.338
Skin temperature	-0.212	0.593
Accelerometer data	0.833	0.363
Step counts	0.852	0.143
Activity counts	0.892	0.233

3.2 The selection of components that contribute to physical fatigue

The mapping from the extracted two components to physical fatigue is checked by a visualized density plot (Figure 3), where component 1 (PC1) shows a relatively clear classifying performance. Component 2 was discarded here since its high overlapping zones between two states, which is not suitable for developing the classifier.

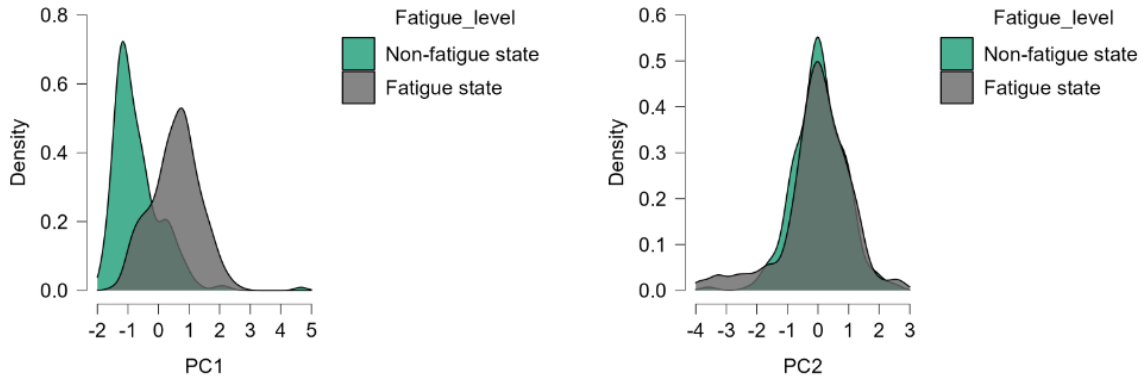


Figure 3. Density plot of components 1 and 2 according to the fatigue level

As for the common area shared by two fatigue states in the density plot of PC1, three potential reasons are considered:

- Individual physiological differences:

Participants vary in their cardiorespiratory endurance, muscle strength, overall fitness, etc., which affect their level of perceived fatigue.

- Subjective interpretation and perceived differences:

Even under similar physiological conditions, different participants may understand the concept of 'fatigue' differently. They may also differ in how they interpret the scale and map their feelings onto the scale.

- Exercise Differences:

Participants completed exercises designed to imitate industrial duties, with modifications in exercise type, trial time, and rest intervals. These variations in workout routines might result in variable exhaustion levels among individuals. Some people may feel more fatigued because of the nature and intensity of the workouts, as well as disparities in recovery time between trials and rest periods.

To better deal with the variabilities, the approach of fuzzy logic in the fuzzy set theory was applied to support developing the classifier.

3.3 PCA-based fuzzy logic classifier

Component 1 extracted from the last part was set as the input of the designed classifier. Taking into account the advantages of the normal distribution in terms of good explanation ability, operation simplification, and the ability to fit negative numbers in the data set, the normal distribution is applied to fit the (non-)fatigue states. Maximum likelihood Estimation (MLE) supports to estimate the parameters of distributions of two states, which are shown in Eq (2) and Eq (3). Then, the membership function could be designed according to the two states' Cumulative Density Function (CDF), which is shown in Figure 4a.

The Probability Density Function (PDF) of the Non-fatigue state:

$$PDF_{nonfatigue}(x) = \frac{1}{0.7201\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x+0.7079}{0.7201})^2} \quad (2)$$

The PDF of the Fatigue state:

$$PDF_{fatigue}(x) = \frac{1}{0.7805\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-0.5491}{0.7805})^2} \quad (3)$$

According to the membership function, it is possible to further define the area between two states based on the degree of belonging. The division could be more specialized based on the different needs of application scenarios, which inspires the adoption of different strategies or actions. In this case, prompt alerts would be available for operators once they entered the areas characterized by elevated levels of fatigue. Here are the general rules (Figure 4b):

- Area 1 (A1： pure non-fatigue state): in this area, all belonging values of the fatigue state are 0, e.g., [-3.6, 0]. Operators in this state are considered to be in a state of high mental concentration, and there is no need to intensively focus on the fatigue level of operators.
- Area 2 (A2: higher non-fatigue state): in this area, belonging values of the non-fatigue state are higher than those of the fatigue state. The exhaustion of the operators began to accumulate gradually, but the impact on the work is not significant. Thus, intermittent rest is recommended.
- Area 3 (A3: higher fatigue state): in this area, belonging values of the fatigue state are higher than those of the non-fatigue state. The impact on operators from physical fatigue should not be underestimated, longer resting time is supposed to be taken into account.
- Area 4 (A4: pure fatigue area): in this area, all belonging values of the non-fatigue state are 0. Operators are fully tired, stricter regulations should be considered here, e.g., stopping work immediately and taking compulsory rests.

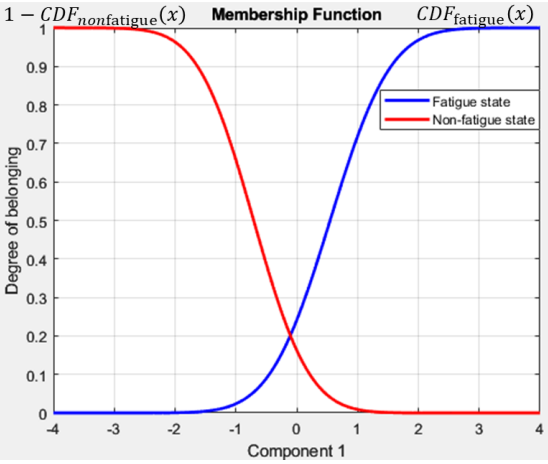


Figure 4(a). The designed membership function of the classifier

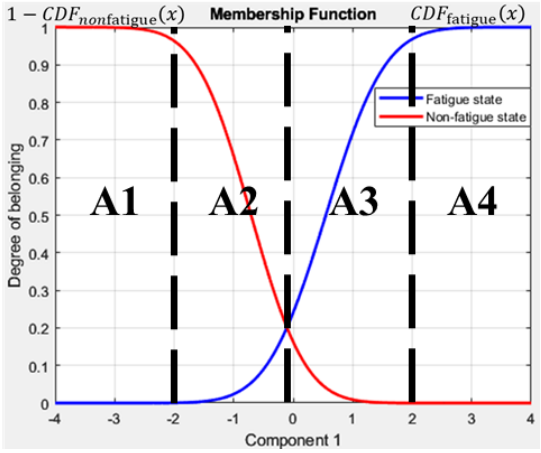


Figure 4(b). The division of physical fatigue states

4. Discussion and conclusions

In process industries, the diversity of work scenarios increases the difficulty of measuring the physical fatigue state of workers. This study focuses on four difficulties: 1. Apply Principal Component Analysis (PCA) to address the complexity from high dimensional dataset. 2. Determine the most relative component contributing to physical fatigue. 3. Fuzzy logic methods are considered to address the variability from high subjectivity of participants. 4. A PCA-based fuzzy logic classifier is designed to support intelligent decision-making processes. The finding of this study benefits practical activities in process industries. With the classifier, it is possible to evaluate workers' fatigue levels, with the consideration of the complexities associated with the subjective nature of fatigue experienced by workers. And the implementation of this classifier is meaningful for enhancing worker welfare and productivity in process industries.

This approach departs from the conventional static or infrequent assessments, which makes sure that fatigue levels are continuously assessed by providing real-time and continuous tracking with the help of deployed smart watches. Secondly, this approach helps to address the multiple data sources and data types faced in the process industries. Variables that related to physical fatigue can be furtherly determined as critical indexes, which makes it easier to identify fatigue early on and enables prompt responses that avoid severe exhaustion also with the consideration of cost consumption at the same time. Moreover, the classifier plays a crucial role in determining the optimal recovery times for workers. By distinguishing more clearly between states of fatigue

and non-fatigue, the system can recommend personalized rest periods and recovery strategies. This customization ensures that each worker receives the necessary downtime to recover fully, thereby preventing over-fatigue. As a result, workers are more likely to perform at their best during active work periods, which is highlighted in process industries with complex and intensive manual labor operations. Finally, the flexibility of the classifier allows for the development of agile, tailored warning systems. These systems can be adapted to meet the specific needs of different work environments within the process industry, enhancing safety and well-being. The adoption of such customized strategies not only improves the immediate environment for workers but also boosts overall efficiency and productivity in the industry, with even the potential of setting new standards for worker care and operational excellence.

The limitations of this study come from two aspects. In this study, multiple movements are generally considered as labor behavior without further classifying different operation types in different working scenarios. From the perspective of designing warning systems, it is not enough to divide in-between fatigue states in a general way, the diversity of different operation tasks and other considerations about potential consequences caused by physical fatigue in human errors should be also taken into account.

The future work is expected to define different rules/strategies applied in different scenarios of process industries based on different membership values. Another meaningful work is to combine the considerations from different decision-making levels: e.g., operators, controllers, managers, etc., which furtherly guide the development of useful warning systems. Moreover, it is also valuable to encompass the development of a personalized scale for determining the optimal rest period, aligning with our research on fatigue states.

Acknowledgments

This research is supported by the program of China Scholarship Council (No. CSC202207930021).

References

Baldissone, G., Demichela, M., Leva, C., 2019. Multivariable based decision-making for the maintenance strategy of process equipment. Chemical Engineering Transactions 77, 1009–1014. https://doi.org/10.3303/CET1977169

Chen, S., Xu, K., Yao, X., Ge, J., Li, L., Zhu, S., Li, Z., 2021. Information fusion and multi-classifier system for miner fatigue recognition in plateau environments based on electrocardiography and electromyography signals. Computer Methods and Programs in Biomedicine 211, 106451. https://doi.org/10.1016/j.cmpb.2021.106451

Chen, Y., Lu, C., Tian, S., Gu, Q., Jiang, S., Li, X., Zou, Y., 2023. Monitoring and detecting coal miners' fatigue status using MPA-LSSVM in the vision of smart mine. Process Safety and Environmental Protection 179, 774–783. https://doi.org/10.1016/j.psep.2023.09.054

Mariana, M., Sahroni, T.R., Gustiyana, T., 2018. Fatigue and Human Errors Analysis in Petrochemical and Oil and Gas Plant's Operation.

Morillo, C.A., Demichela, M., 2023. A Data-Driven Framework to Model Physical Fatigue in Industrial Environments Using Wearable Technologies, in: Neuroergonomics and Cognitive Engineering. Presented at the AHFE (2023) International Conference, AHFE Open Access. https://doi.org/10.54941/ahfe1003016

Nguyen, D.T., Trodahl, M., Pedersen, T.A., Bakdi, A., 2023. Verification of collision avoidance algorithms in open sea and full visibility using fuzzy logic. Ocean Engineering 280, 114455. https://doi.org/10.1016/j.oceaneng.2023.114455

Sedighi Maman, Z., Chen, Y.-J., Baghdadi, A., Lombardo, S., Cavuoto, L.A., Megahed, F.M., 2020. A data analytic framework for physical fatigue management using wearable sensors. Expert Systems with Applications 155, 113405. https://doi.org/10.1016/j.eswa.2020.113405

Yang, S., Demichela, M., 2023. A data-driven narratives skeleton pattern recognition from accident reports dataset for human-and-organizational-factors analysis. Journal of Loss Prevention in the Process Industries 83, 105047. https://doi.org/10.1016/j.jlp.2023.105047