

A machine learning based sensing and measurement framework for timing of volcanic eruption and categorization of seismic data

Vijay Souri Maddila¹, Katady Sai Shirish¹, M. V. S. Ramprasad²

¹ Department of Computer Science Engineering, GITAM (Deemed to be University), Visakhapatnam--530045, Andhra Pradesh, India

² Department of Electrical, Electronics and Communication Engineering (EECE), GITAM (Deemed to be University), Visakhapatnam-530045, Andhra Pradesh, India

ABSTRACT

The circumstances and factors which determine the volcanic explosive ejection are unknown, and currently, there is no effective way to determine the end of a volcanic explosive ejection. At present, the end of an eruption is determined by either generalized standards or the measurement which is unique to the volcano. We investigate the use of controlled machine learning techniques such as Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), and Gaussian Process Classifiers (GPC), and create a decisiveness index D to assess the uniformity of the groups provided by these machine learning models. We analyzed the measured end-date obtained by seismic information categorization is two to four months later than the end-dates determined by the earliest instance of visible eruption for both volcanic systems. Likewise, the measurement systems, measurement technology becomes key elements in the seismic data analysis. The findings are consistent across models and correspond to previous, broad definitions of ejection. Obtained classifications demonstrate a more significant relationship between eruptive movement and visual activity than information base records of ejection start and completion timings. Our research has presented a new measurement-based categorization technique for studying volcanic eruptions, which provides a reliable tool for determining whether or not an emission has stopped without the need for visual confirmation.

Section: RESEARCH PAPER

Keywords: Volcanic eruption; machine learning; measurement; seismic data; sensing

Citation: Vijay Souri Maddila, Katady Sai Shirish, M. V. S. Ramprasad, A machine learning based sensing and measurement framework for timing of volcanic eruption and categorization of seismic data, Acta IMEKO, vol. 11, no. 1, article 24, March 2022, identifier: IMEKO-ACTA-11 (2022)-01-24

Section Editor: Md Zia Ur Rahman, Koneru Lakshmaiah Education Foundation, Guntur, India

Received November 29, 2021; **In final form** February 19, 2022; **Published** March 2022

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Corresponding author: Vijay Souri Maddila, e-mail: vijaysouri.maddila123@gmail.com

1. INTRODUCTION

Monitoring and assessing volcanic activity, as well as the risks connected with it, remains a key concern. According to the strategy offered by the U. N. (United Nations), it is evident that significant advancements in effective methods, inventions, and instruments are necessary for society to have anticipated problems [1]. Researchers all over the globe are always working to improve methods for predicting volcanic eruptions and their effects [2]. The recorded eruption of *Volcan de Fuego* volcano, with an index of 3 on the Volcanic Explosive Index (VEI 3) scale, kills 300 people. Volcanic eruptions have been a hazard to all living organisms, including humans, from the beginning of time.

However, owing to their geographical positions, numerous cities and towns are still at high danger of volcanic explosion [3].

Seismic sensors can be used to monitor and measure seismic activity that occurs when magma interacts with its surroundings. Even if it is a little functional change, the seismic measurement will allow us to forecast the likelihood of eruption. Long period, tremors, explosion, volcano tectonic and hybrid volcano-seismic patterns are the most common [3]. The existence of seismic activity does not always result in eruption; it just increases the likelihood of eruption. Seismic activity, eruptions are inherently probabilistic [4]. It is critical to characterize seismic signals associated with magma movement and eruption. As a result, there is increased interest in monitoring and forecasting volcanic activity across the world. A monitoring context can determine the end of an eruption in two ways: first, if there had been no

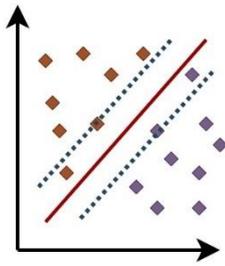


Figure 1. Support Vector Machine (SVM).

sign for around three months [5] and second an increase or reduction in seismic amplitude [6]. Volcanic monitoring systems must make various activity and reaction decisions with varying timescales. Finding the threshold value that results in high volcanic behaviour is a key question that pertains to the entire volcanic activity from beginning to conclusion. Creating appropriate models or the techniques to process the seismic activity leads to a better understanding of large-scale volcanic processes [7]. Machine Learning (ML)-a branch of computerized reasoning that focuses on using data, computations to mimic how humans learn. Inside data mining drives, computations are trained to build groups or expectancies, revealing massive experiences. ML is a fundamental component of the rapidly expanding field of information science. As big data continues to grow and evolve, so will market interest in ML [8]. Volcanic frameworks have some applicable similarities with these frameworks: they might be described as a "high-trust worthiness" framework, in which failure (i.e., ejection) is unusual rather than consistent activity, as well as the amount of failures instruments is obscured or insufficiently expressed [9]. The use of ML techniques in seismology is a fairly new discipline. Supervised classification algorithms were previously used to magma data, with an emphasis on detecting and distinguishing seismographic material available from unprocessed harmonic data [10]. The Authors [11] work utilized deep learning to detect ground deformation in Sentinel-1 data and article [12] uses logistic regression to predict volcanic eruptions in SO₂ measurements obtained with the Ozone Mapping Instrument. The Author predicts the timing of volcanic eruptions in this study using ML techniques such as Random Forest, SVM, Logistic Regression,

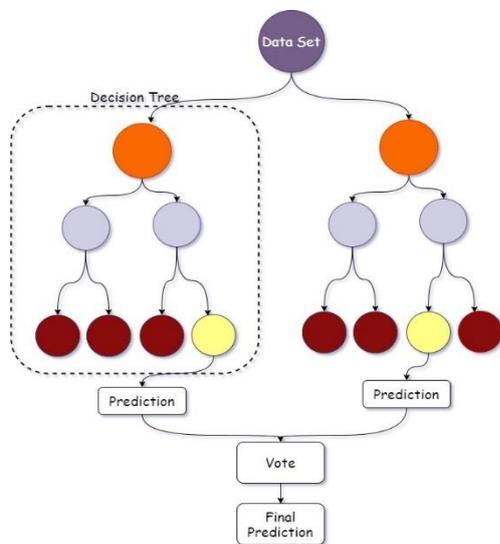


Figure 2. Random Forest.

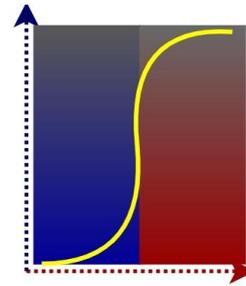


Figure 3. Logistic Regression.

and Gaussian Process Classifier. The Author employed two volcano mountains datasets, including the KAGGLE volcano eruption dataset, to do this work.

2. MODELS IMPLEMENTED

2.1. Support vector machine (SVM)

SVMs (Figure 1) are ML supervised learning models that analyse data for prediction purposes. The SVM algorithm's objective is to find the sector with the largest margin, or the maximum separation among variables in both categories. Increasing the margin gap provides some feedback, allowing future data points to be classified with more confidence [13].

2.2. Random Forest (RF)

A Random Forest (Figure 2) is a ML method to tackling regression and classification tasks. As the number of nodes increases, so does the accuracy of the result. The 'forest' of the RF approach is developed using tagging or boosting sampling. Bagging is a macro ensemble that enhances ML technique performance. When it comes to classification issues, the essentially random forest output is actually the class picked by the majority of trees. Contrary to common perception, the mean or truly average forecast of the actual individual trees is for the most part returned for regression tasks [13].

2.3. Logistic Regression

Logistic regression (Figure 3) is, contrary to common perception, a statistical model that, in its most fundamental form, models a binary essentially sort of dependant basically pretty variable using a logistic function. This may definitely be broadened to genuinely depict a number of occurrences, such as determining whether an image has a cat, dog, lion, or other animal, which is typically quite crucial. Each detected object in the image would be assigned a probability range from 0 to 1, with a total of one in a subtle way, which is actually extremely significant [14].

2.4. Gaussian Process Classifier (GPC)

The distribution of a Gaussian process (Figure 4) is unquestionably the sum of all those (infinitely numerous) random variables in a major manner. Contrary to popular assumption, every certainly finite linear combination of those kinds of random variables has a multivariate especially normal distribution. Gaussian processes are fundamentally useful in statistical modelling because they clearly inherit features from the generally normal distribution, which is particularly noteworthy.

2.5. Implementation

For a day to be classified as eruptive, a rolling arithmetic mean of the categorization is utilized as a quantize screening criterion. Every particular day which is in the time series is categorized

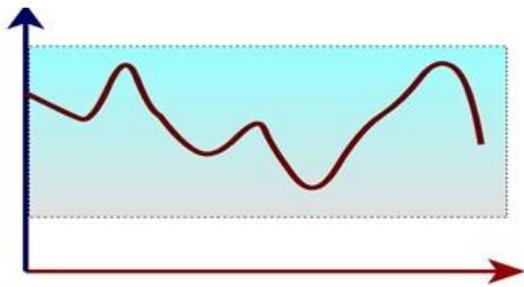


Figure 4. Gaussian Process Classifier.

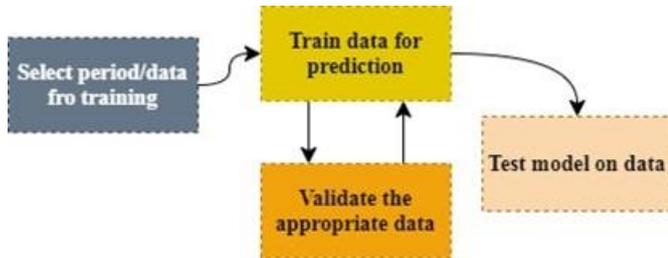


Figure 5. The training and testing framework for supervised multi-class classifying models.

separately from the others. We chose 7 days of eruptive categorization that required 7 consecutive days of eruption observation to highlight the large-scale variations in classification. The categorization of data as extenivolvent is more conservative when this filter is applied to the model output than when the results are left unfiltered [15]. Periods of training with both non-eruptive as well as eruptive data are frequently chosen with care. Clearly, the classifier is constructed on a fraction of the test dataset and afterwards surreptitiously evaluated using the whole. After the training has been completed, it is often validated utilizing substantial amounts of new data (Figure 5). We chose time wisely that did not intersect with the begin and finish dates of the Global Volcanism Program (GVP) because we intended to regulate the timeframe of changeover between active volcanic and quasi activity independently. Feature extraction is the process of finding variables that will be used as inputs into ML models [16]-[20].

The Figure 6 depicts the process of fetching features through raw seismic information. Data set based on gathered data sets are

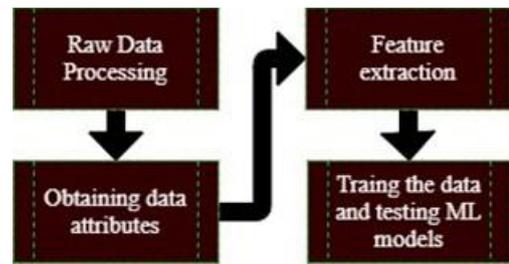


Figure 6. The architecture for fetching characteristics from raw seismic data is depicted in the diagram below.

fed into ML algorithms. Raw waveform data is used to detect events. We extract characteristics such as peak amplitudes and band ratios from each event waveform. Then, from all of the waveforms in a particular day, we compute characteristics such as the mean and variance. The resultant time series are sent into a ML classifier as input.

3. RESULTS

We independently constructed four unique categorization models for each lava region, to every modelling approach trained and validated on each lava flow sequentially, which is critical in all intents and purposes. Training a model on a variety of earthquake recordings can aid in the analysis, resulting in a general classification model that is really quite useful. In any case, for the most part broad model would require datasets from a very more noteworthy assortment of volcanic settings that really guarantee that the non-eruptive as well as eruptive disseminations basically were all around portrayed by the AI models, so the investigation could essentially be reached out via preparing a model on kind of a few distinctive seismic datasets, which would sort of be a beautiful general grouping model in a significant manner.

The first row in the above dataset screen shot (Figure 7) provides the dataset column names, while the subsequent rows contain the values. Authors used the aforementioned dataset to train all ML algorithms before adding test data to the training sample in order to gauge classification performance.

Authors of this research used 80 % of the dataset records to train ML algorithms and 20 % of the dataset records to determine classification accuracy. The dataset is imported into a

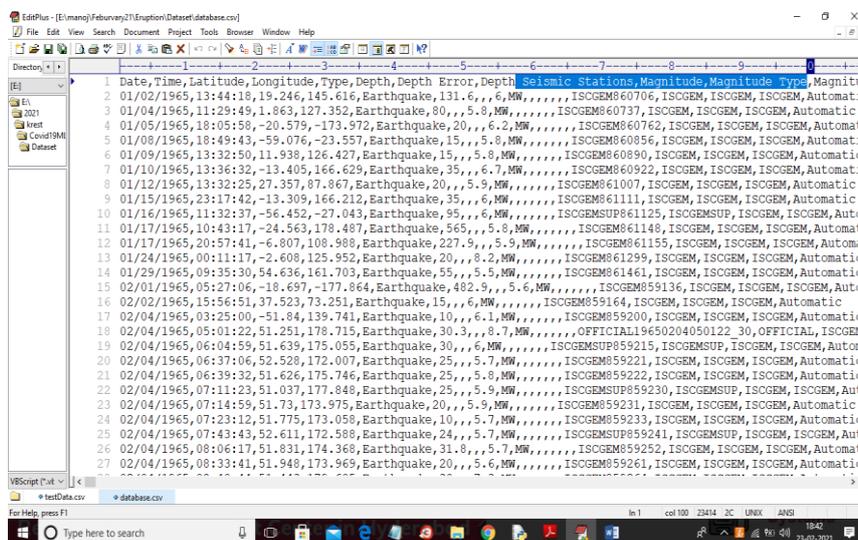


Figure 7. Dataset used for model implementations.

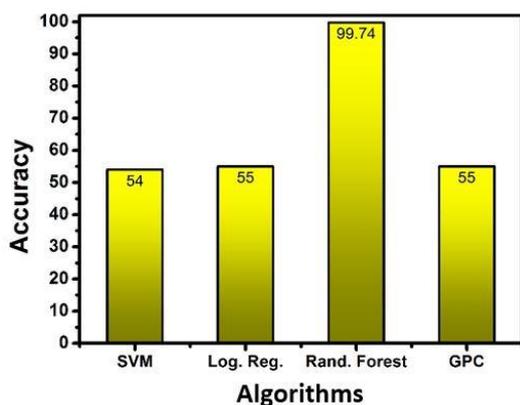


Figure 8. Accuracy chart for trained data with respect to Algorithms.

developed application that displays records from the dataset, and we need to replace string values with numeric values and then replace missing values with 0, therefore 'Pre-process Dataset Feature Extraction' is used to turn the dataset into a normalized format. Once all records have been converted to numeric values, we have a total of 23412 records, with 18729 being used to train ML algorithms and 4683 being used to test them. Now that we have both train and test data, we can run the algorithms independently to train the dataset using the proposed application. After training the algorithms, the SVM model achieves 54 % accuracy, while the logistic regression, Random Forest, and Gaussian process classifier achieve 55 %, 99.74 % and 55 % accuracy, respectively. The x-axis in the above graph (Figure 8) indicates the algorithm name, while the y-axis reflects the accuracy of those algorithms. Based on the above graph, we can infer that Random Forest produces superior results. then submit a test file, and the program detects eruption activity based on the time data that was provided We can view the volcano test data and the expected outcome as 'No eruption identified' or 'Eruption detected' following the square bracket. In the above

screen, we can see that when the classifier sees a magnitude value more than 6.5 (Figure 9), it classifies that record time as 'eruption activity identified.'

4. CONCLUSIONS

ML computations in seismic time series can precisely categorize general examples for both eruptive as well as non-eruptive behaviour. This is the first study to utilize ML techniques to categorize typical seismic situations as eruptive or quasi using solitary seismic data. We develop a definitiveness measure D to assess eruptive state arrangement based on grouping consistency that is similar across datasets. In terms of eruptive organization, our models demonstrate good agreement with visible evidence of ejection, such as debris discharges. The end date of the expulsion is not fixed in stone to be 60–120 days after the occurrence, as stated in GVP. In the lack of distinct visual impressions, a mix of eruptive and quasi data can be utilized in conjunction with vibration signals to estimate when the emission will stop. Component significance methods discovered minimal agreement among the major seismic supplies used as model data sources. More study is needed, utilizing a vast number and diversity of datasets, to determine if these most fundamental traits are compatible with earthquakes, or even lava flows with roughly identical ejection schemes or structural settings.

REFERENCES

- [1] M. Malfante, M. Dalla Mura, J. P. Métaixian, J. I. Mars, O. Macedo, A. Inza, Machine learning for volcano-seismic signals: Challenges and perspectives, *IEEE Signal Processing Magazine*, 35(2) (2018), pp.20-30. DOI: [10.1109/MSP.2017.2779166](https://doi.org/10.1109/MSP.2017.2779166)
- [2] S. Surekha, K. P. Satamraju, S. S. Mirza, A. Lay-Ekuakille, A Collateral Sensor Data Sharing Framework for Decentralized Healthcare Systems, *IEEE Sensors Journal*, 21(24) (2021), pp.

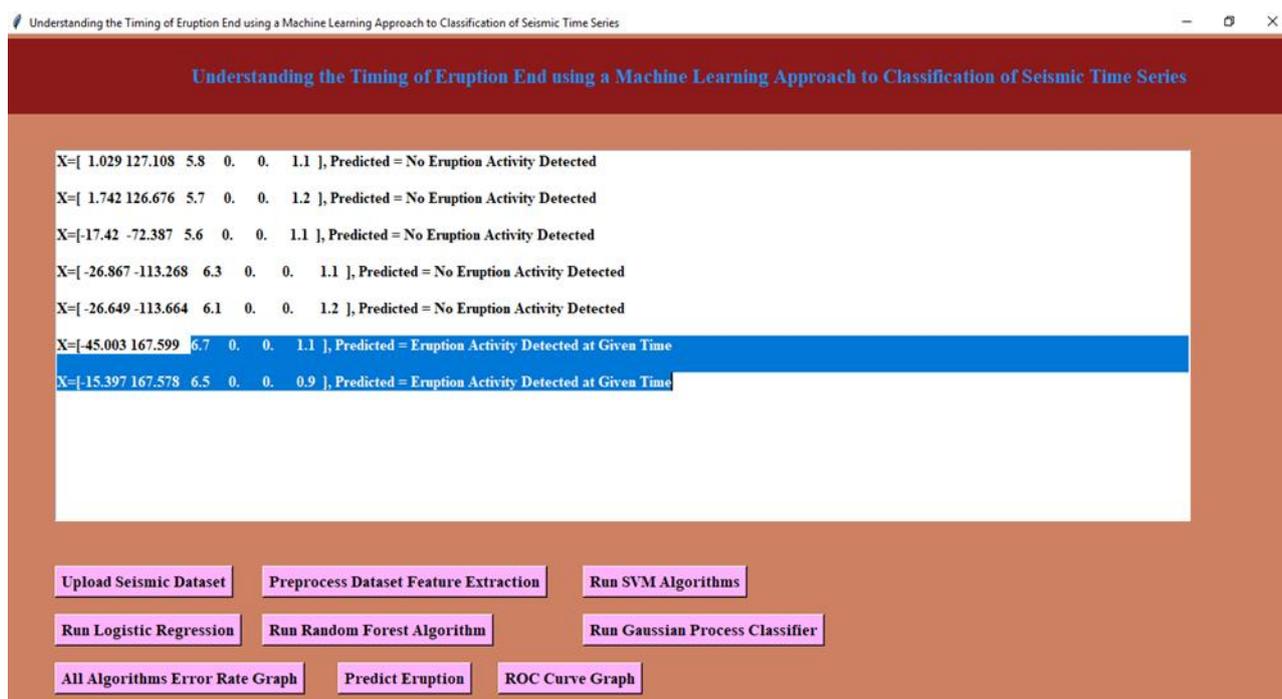


Figure 9. Eruption activity prediction result.

- 27848-27857.
DOI: [10.1109/JSEN.2021.3125529](https://doi.org/10.1109/JSEN.2021.3125529)
- [3] V. Gavini, J. Lakshmi, A Robust CT Scan Application for Prior Stage Liver Disorder Prediction with Googlenet Deep learning Technique, *ARNP Journal of Engineering and Applied Sciences*, 16 (18) (2021), pp. 1850-1857.
- [4] J. A. Power, S. D. Stihler, B. A. Chouet, M. M. Haney, D. M. Ketner, Seismic observations of Redoubt Volcano, Alaska—1989–2010 and a conceptual model of the Redoubt magmatic system, *Journal of Volcanology and Geothermal Research*, 259 (2013), pp.31-44.
DOI: [10.1016/j.jvolgeores.2012.09.014](https://doi.org/10.1016/j.jvolgeores.2012.09.014)
- [5] S. H. Ahammad, M. Z. U. Rahman, L. K. Rao, A. Sulthana, N. Gupta, A. Lay-Ekuakille, A Multi-Level Sensor based Spinal Cord disorder Classification Model for Patient Wellness and Remote Monitoring, *IEEE Sensors Journal*, 21(13) (2021), pp. 14253-14262.
DOI: [10.1109/JSEN.2020.3012578](https://doi.org/10.1109/JSEN.2020.3012578)
- [6] V. Gavini, G. R. Jothi Lakshmi, An efficient machine learning methodology for liver computerized tomography image analysis, *International Journal of Engineering Trends and Technology*, 69 (7) (2021), pp. 80-85.
DOI: [10.14445/22315381/IJETT-V69I7P212](https://doi.org/10.14445/22315381/IJETT-V69I7P212)
- [7] National Academies of Sciences, Engineering, and Medicine, *Volcanic eruptions and their repose, unrest, precursors, and timing*, National Academies Press, 2017.
DOI: [10.17226/24650](https://doi.org/10.17226/24650)
- [8] What Is Machine Learning?, What is Machine Learning?-India | IBM. Online [Accessed 17 March 2022]
<https://www.ibm.com/in-en/cloud/learn/machine-learning>
- [9] A. Maggi, V. Ferrazzini, C. Hibert, F. Beauducel, P. Boissier, A. Amemoutou, Implementation of a multistation approach for automated event classification at Piton de la Fournaise volcano, *Seismological Research Letters*, 88(3) (2017), pp. 878-891.
DOI: [10.1785/0220160189](https://doi.org/10.1785/0220160189)
- [10] M. Malfante, M. Dalla Mura, J. I. Mars, J. P. Métaixian, O. Macedo, A. Inza, Automatic classification of volcano seismic signatures, *Journal of Geophysical Research: Solid Earth*, 123(12) (2018), pp. 10-645.
DOI: [10.1029/2018JB015470](https://doi.org/10.1029/2018JB015470)
- [11] N. Anantrasirichai, J. Biggs, F. Albino, P. Hill, D. Bull, Application of machine learning to classification of volcanic deformation in routinely generated InSAR data, *Journal of Geophysical Research: Solid Earth*, 123(8) (2018), pp.6592-6606.
DOI: [10.1029/2018JB015911](https://doi.org/10.1029/2018JB015911)
- [12] V. J. Flower, T. Oommen, S. A. Carn, Improving global detection of volcanic eruptions using the Ozone Monitoring Instrument (OMI), *Atmospheric Measurement Techniques*, 9(11) (2016), pp.5487-5498.
DOI: [10.5194/amt-9-5487-2016](https://doi.org/10.5194/amt-9-5487-2016)
- [13] A. Tarannum, L. K. Rao, T. Srinivasulu, A. Lay-Ekuakille, An efficient multi-modal biometric sensing and authentication framework for distributed applications, *IEEE Sensors Journal*, 20(24) (2020), pp. 15014-15025.
DOI: [10.1109/JSEN.2020.3012536](https://doi.org/10.1109/JSEN.2020.3012536)
- [14] A. Tarannum, T. Srinivasulu, An efficient multi-mode three phase biometric data security framework for cloud computing-based servers, *International Journal of Engineering Trends and Technology*, 68 (9) (2020), pp. 10-17.
DOI: [10.14445/22315381/IJETT-V68I9P203](https://doi.org/10.14445/22315381/IJETT-V68I9P203)
- [15] G. F. Manley, D. M. Pyle, T. A. Mather, M. Rodgers, D. A. Clifton, B.G. Stokell, G. Thompson, J. M. Londoño, D. C. Roman, Understanding the timing of eruption end using a machine learning approach to classification of seismic time series, *Journal of Volcanology and Geothermal Research*, 401 (2020), p.106917.
DOI: [10.1016/j.jvolgeores.2020.106917](https://doi.org/10.1016/j.jvolgeores.2020.106917)
- [16] Henrik Ingerslev, Soren Andresen, Jacob Holm Winther, Digital signal processing functions for ultra-low frequency calibrations, *ACTA IMEKO*, 9(5) (2020), pp. 374-378.
DOI: [10.21014/acta_imeko.v9i5.1004](https://doi.org/10.21014/acta_imeko.v9i5.1004)
- [17] Lorenzo Ciani, Alessandro Bartolini, Giulia Guidi, Gabriele Patrizi, A hybrid tree sensor network for a condition monitoring system to optimise maintenance policy, *ACTA IMEKO*, 9(1) (2020), pp. 3-9.
DOI: [10.2104/acta_imeko.v9i1.732](https://doi.org/10.2104/acta_imeko.v9i1.732)
- [18] Mariorosario Prist, Andrea Monteriù, Emanuele Pallotta, Paolo Cicconi, Alessandro Freddi, Federico Giuggioloni, Eduard Caizer, Carlo Verdini, Sauro Longhi, Cyber-Physical Manufacturing Systems: An Architecture for Sensors Integration, *Production Line Simulation and Cloud Services*, *Acta IMEKO*, 9(4) (2020), article 6.
DOI: [10.21014/acta_imeko.v9i4.731](https://doi.org/10.21014/acta_imeko.v9i4.731)
- [19] Jiayu Luo, Xiangyu Kong, Changhua Hu, Hongzeng Li, Key-performance-indicators-related fault subspace extraction for the reconstruction-based fault diagnosis, *Elsevier: Measurements*, 186 (2021), pp. 1-12.
DOI: [10.1016/j.measurement.2021.110119](https://doi.org/10.1016/j.measurement.2021.110119)