iJRASET
International Journal For Research in
Applied Science and Engineering Technology

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

**Volume:** 10 **Issue:** XI **Month of publication:** November 2022

**DOI:** https://doi.org/10.22214/ijraset.2022.47252

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue XI Nov 2022- Available at www.ijraset.com*

# Solar Radiation Prediction by Machine Learning Algorithm & their Techniques

Yogesh Singh[1], Amarendra Singh[2],
*[1]M.tech Scholar, [2] Assistant Professor, Department of Civil Engineering, Institute of Engineering and Technology, Lucknow-226021 UP, India*

***Abstract: For the power grid to run smoothly or for the best control of the energy flows into the solar system, it is necessary to forecast the output power of solar systems. It is crucial to concentrate the prediction on solar irradiance before predicting the solar system's output. There are numerous ways to anticipate the global solar radiation, but the two main categories are machine learning algorithms and cloud images mixed with physical models. This paper's goal is to provide an overview of machine learning-based algorithms for solar irradiation forecasting in this setting. Despite the fact that many studies explain techniques like support vector regression or neural networks. Due to the variability of the data collection, time step, forecasting horizon, setup, and performance measures, ranking the performance of such methods is challenging. Overall, the prediction error is very similar. Some write. Global solar radiation suggested using hybrid models or an ensemble forecast technique to boost prediction accuracy.***

***Keywords: Forecasting solar radiation, Decision Tree Method, Deep Learing Method, and Generalized Learning Method.***

## I. INTRODUCTION

Global solar radiation arriving on the Earth's surface is of fundamental significance for various applications, such as meteorology, hydrology, and particularly for the design and use of renewable solar energy systems. However, direct measurements of Global solar radiation are not readily available for most worldwide locations, especially in developing countries, most likely due to the high costs of installation and the difficulty in maintenance of the measuring instruments. Since observed Global solar radiation data are not always accessible, different techniques have been developed to estimate Global solar radiation, such as empirical models by establishing the linear/nonlinear relationships between the meteorological variables and global solar radiation, machine learning models to simulate the complex and nonlinear mapping from meteorological variables to global solar radiation, satellite-based methods for continuously monitoring the spatiotemporal changes in solar radiation on global and regional scales and radiative transfer models to simulate the solar radiation scattering and absorption in the atmosphere. There are also international databases providing large-scale global solar radiation data such as Meteonorm, SolarGIS, and NASA-SSE (Surface meteorology and Solar Energy). Among the above techniques, empirical and machine learning models are more commonly used in practice as a result of their low computational costs and high prediction accuracy, respectively. Based on the horizontal Global solar radiation, global solar radiation on PV panel surfaces with particular tilt angles can be further estimated by the isotropic models and anisotropic models. machine learning models generally provide more accurate Global solar radiation pre-diction compared with empirical models, but the prediction accuracy of various types of machine learning models, particularly their computational efficiency on large-scale dataset for predicting Global solar radiation have been rarely compared in different regions of the world. For instance, Wang et al. only compared three ANN models, e.g., MLP, RBF and GRNN models, for daily Global solar radiation estimation with sunshine duration and other meteorological variables at 12 stations across China in terms of prediction accuracy. It was found that the MLP and RBF models outperformed the GRNN model. Zou et al. have tested the performance of the ANFIS model for predicting daily Global solar radiation at three stations in Hunan Province of China, compared with two empirical models (e.g. Bristow-Campbell Model and Yang's Hybrid Model). It was concluded that the ANFIS model gave more accurate Global solar radiation estimates than the two empirical models. Wang et al. further compared the ANFIS and M5Tree models for daily Global solar radiation estimation at 21 stations across China. The results indicated that the ANFIS model was superior to the M5Tree and empirical models. Fan et al. also compared two machine learning models (e.g., Deep Learning) for daily Global solar radiation prediction in the humid subtropical China. They found that the Deep Learning models outperformed the studied empirical models, and recommended the Deep Learning model as a promising machine learning model for Global solar radiation estimation due to better model stability, efficiency and comparable prediction accuracy.

## II. METHEDOLOGY

### *A. Data Collection*

The satellite and other data utilised in this research are the same as those used for images from Meteosat 5 and 7 (HRI-VIS channel).Twelve images a day were utilised in the satellite calculations.

### *B. Data Cleaning*

The initial data obtained includes some missing units that have to be filled; otherwise, the further analysis will not work out. To deal with the missing units, we remove the columns that are missing a significant amount of data and delete some specific samples which lack integrity.

### *C. Feature Selection*

The data collected have a ton of meaningless features that should be removed. Most of the features will not contribute to the result of the prediction; some of them even worsen the situation. For those features that represent some very irrelevant aspects, the filtration is vital. Other than figuring out each feature and estimated their relativeness through their meanings, a mathematic analysis would be more convincing and efficient

## III. MACHINE LEARNING METHOD

Machine learning is a subfield of computer science and it is classified as an artificial intelligence methods. It can be used in several domains and the advantage of this method is that a model can solve problems which are impossible to be represented by explicit algorithms In the reader can find a detailed review of some machine learning and deterministic methods for solar forecasting. The machine learning models find relations between inputs and outputs even if the representation is impossible; this characteristic allow the use of machine learning models in many cases, for example in pattern recognition, classification problems, spam filtering, and also in data mining and forecasting problems. The classification and the data mining are particularly interesting in this domain because one has to work with big datasets and the task of pre processing and data preparation can be undertaken by the machine learning models. After this step, the machine learning models can be used in forecasting problems. As already mentioned machine learning is a branch of artificial intelligence. It concerns the construction and study of systems that can learn from data sets, giving computers the ability to learn without being explicitly programmed. Below techniques are used for solar radiation prediction.

### *A. Deep Learning*

Machine learning, which is simply a neural network with three or more layers, is a subset of deep learning. These neural networks make an effort to mimic how the human brain functions, while they fall well short of doing so, enabling it to "learn" from vast volumes of data. Deep learning is the sub class of machine learning which mimics the brain. It could also otherwise call as large neural network. But the concept is not knotted with the neural networks alone. Since the usage of deep learning is well explained thorough neural networks it is mostly understood that it is the large neural network.

### *B. Decision Tree Method*

A hierarchy of branches is used to create a decision tree model. A classification decision rule is represented by each route from the root node through internal nodes to a leaf node. These 'if-then' rules can also be used to express these decision tree pathways. Decision tree is a tool for machine learning which supports decision making using a tree like model finds simple tree-like models which are easy to understand. A decision tree consists of 3 kinds of node that is decision node, chance node, and end nodes. Decision Tree is a predictive modelling approach used in classification and prediction tasks that were first introduced by Breiman et al. in 1983. Decision trees use the divide and conquer technique to split the problem search space into subsets. A decision tree is a tree where the root and each internal node is labelled with a question. The arcs emanating from each node represent each possible answer to the associated question. Each leaf node represents a prediction of a solution to the problem under consideration.

### *C. Generalized Learning Method*

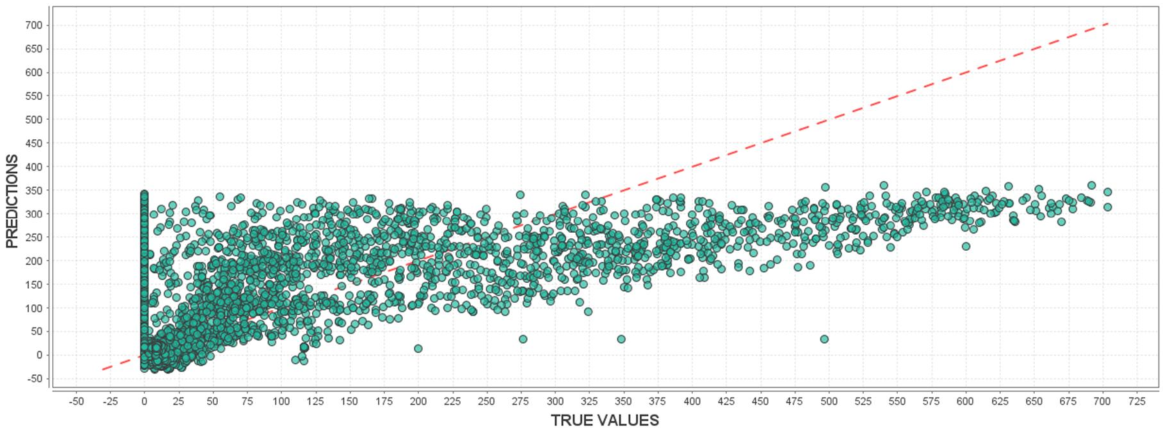
Generalization describes our model's capacity to appropriately respond to novel, previously unobserved data derived from the same distribution as the model. Generalized linear model (GLM) in statistics, is a flexible generalization of ordinary linear regression which allows for response variables that have error distribution models other than a normal distribution. GLM generalizes linear regression by permitting the linear model to be related to the response variable through a link function and by considering the magnitude of the variance of each measurement to be a function of its predicted value. Some studies improve the regression quality using a coupling with other predictors like Kalman filter.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
*ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538*
*Volume 10 Issue XI Nov 2022- Available at www.ijraset.com*

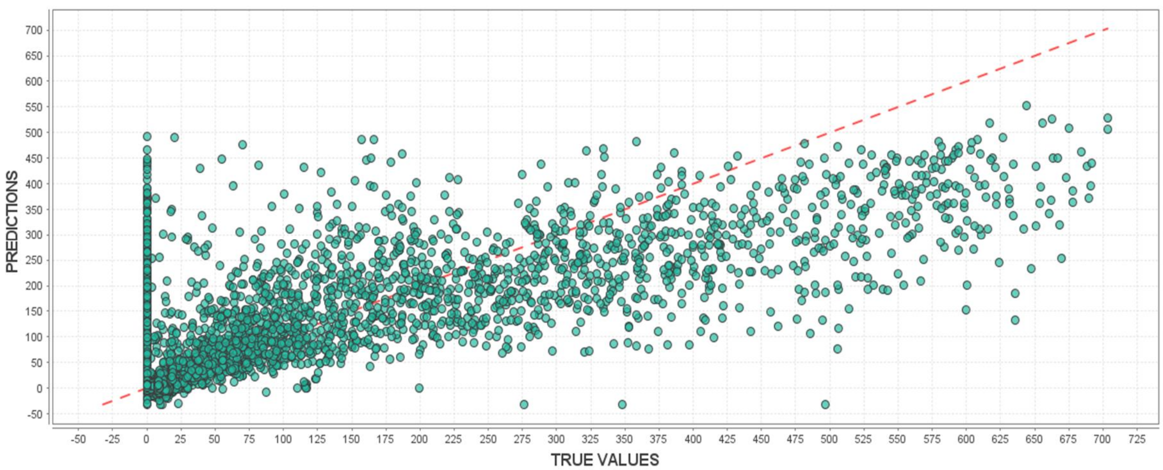
## IV. RESULT AND DISCUSSION

Machine Learning algorithm is the better and efficient method than other traditional methods that is we said because of the result that we have got, After applying the different machine learning techniques like Deep Learning, Decision Tree Learning, Generalized Learning Method then we got some interesting results which are as follows,

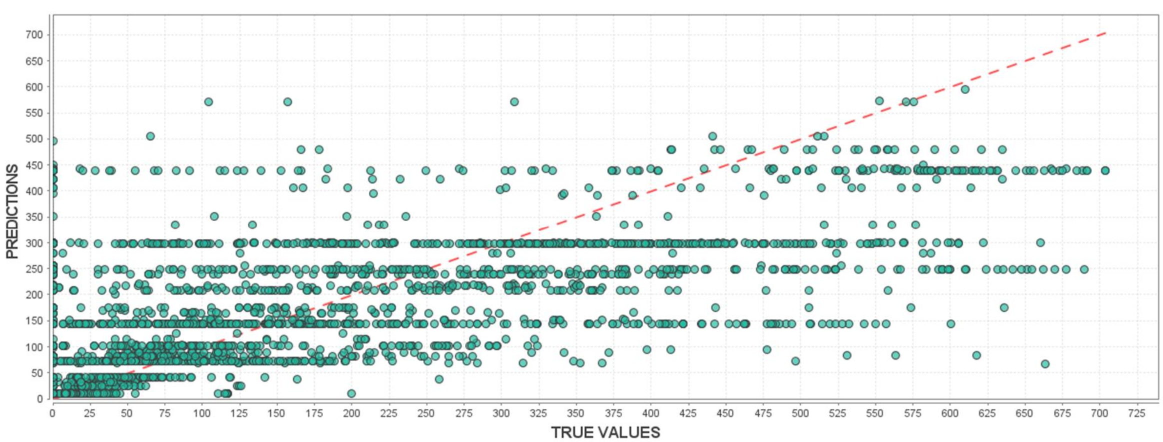
### *A. Prediction Charts*




GENERALIZED LEARNING METHOD




DEEP LEARNING




DECISION TREE METHOD

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 10 Issue XI Nov 2022- Available at www.ijraset.com

### B. Some Predictive Values

#### 1) Deep Learning

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Solar Radiation (SR) | 3.77 | 3.58 | 4.45 | 156.49 | 151.06 | 4.05 | 4.02 | 4.19 |
| prediction(SR) | 8.739028 | 4.653043 | 8.583868 | 219.8195 | 90.90542 | 0.432582 | 2.734777 | 6.980148 |

#### 2) Decision Tree

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Solar Radiation (SR) | 3.77 | 3.58 | 4.45 | 156.49 | 151.06 | 4.05 | 4.02 | 4.19 |
| prediction(SR) | 10.02057 | 10.02057 | 10.02057 | 219.5606 | 114.5713 | 10.02057 | 10.02057 | 10.02057 |

#### 3) Generalized Learning Method

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Solar Radiation (SR) | 3.77 | 3.58 | 4.45 | 156.49 | 151.06 | 4.05 | 4.02 | 4.19 |
| prediction(SR) | 25.20451 | 25.09222 | 25.01659 | 147.7936 | 140.4441 | 23.95128 | 25.09281 | 25.06323 |

### C. Validation

Validation is the method in which we find the different validation points like r square, squared error, relative error value, root mean square error value, absolute error value that helps us to find the accuracy of prediction models. Validation provides the better image of our models so that we can easily opt the best prediction model for our working. Validation simply find the error values to decide their accuracy in the results.

| MODEL | R SQUARE VALUE | SQUARED ERROR VALUE | RELATIVE ERROR VALUE | ROOT MEAN SQUARE ERROR VALUE | ABSOLUTE ERROR VALUE |
|---|---|---|---|---|---|
| Generalized Linear Model | 0.541750087 | 10002.54558 | 0.661260767 | 100.0046921 | 59.84911737 |
| Deep Learning | 0.642784893 | 7823.378998 | 0.557837468 | 88.44027849 | 47.30625407 |
| Decision Tree | 0.585571852 | 9076.72601 | 0.419110147 | 95.26723317 | 49.03868087 |

The mean absolute error (MAE) is appropriate for applications with linear cost functions, i.e., where the costs resulting from a poor forecast are proportional to the forecast error:

$$MAE = \frac{1}{N} \times \sum_{i=1}^{N} \left| \hat{y}(i) - y(i) \right|$$

The mean square error (MSE) uses the squared of the difference between observed and predicted values. This index penalizes the highest gaps:

$$MSE = \frac{1}{N} \times \sum_{i=1}^{N} \left(\hat{y}(i) - y(i)\right)^2$$

MSE is generally the parameter which is minimized by the training algorithm.

The root mean square error (RMSE) is more sensitive to big forecast errors, and hence is suitable for applications where small errors are more tolerable and larger errors cause disproportionately high costs, as for example in the case of utility applications. It is probably the reliability factor that is most appreciated and used:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \times \sum_{i=1}^{N} \left(\hat{y}(i) - y(i)\right)^2}$$

The mean absolute percentage error (MAPE) is close to the MAE but each gap between observed and predicted data is divided by the observed data in order to consider the relative gap.

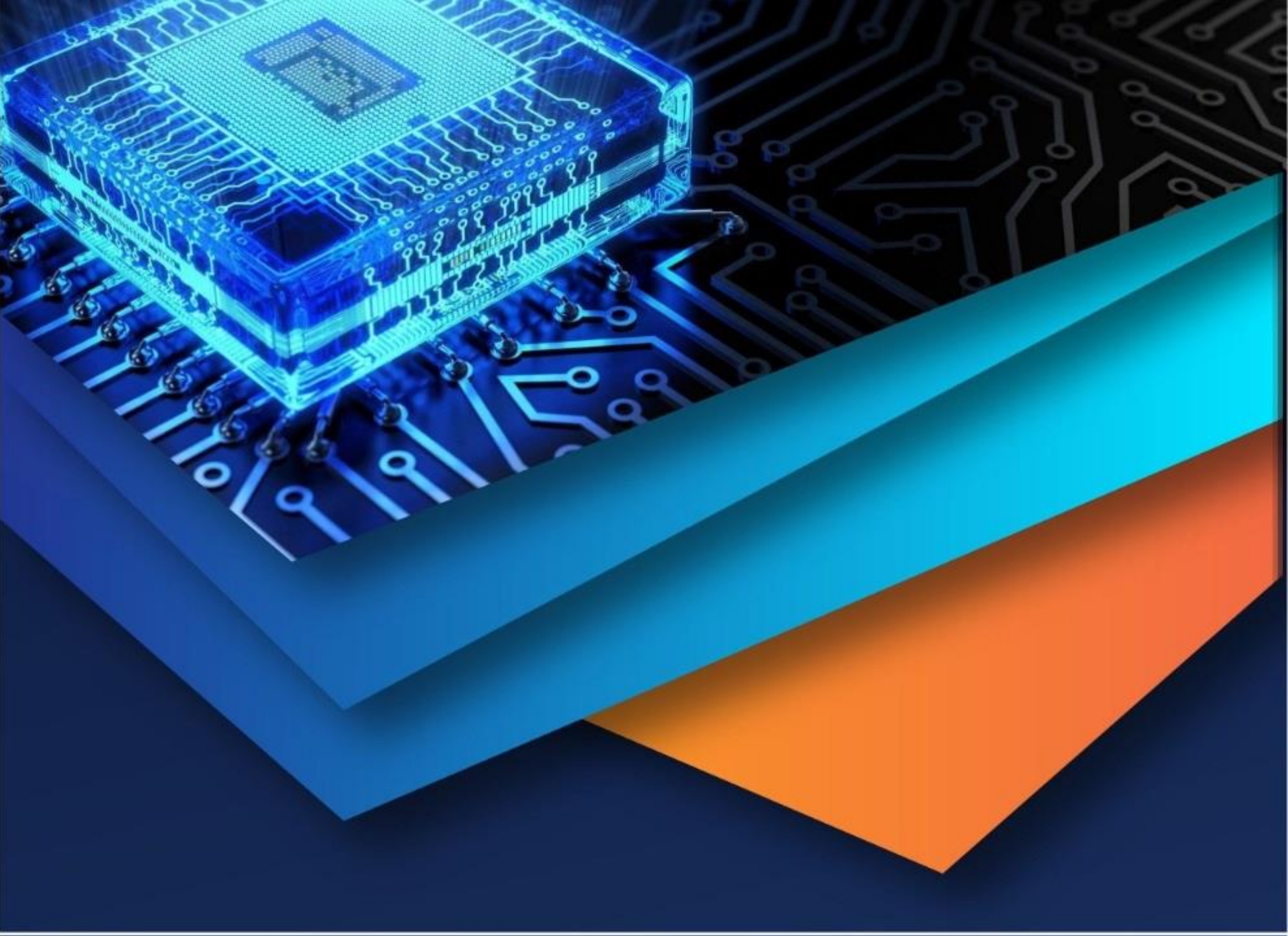
$$MAPE = \frac{1}{N} \times \sum_{i=1}^{N} \left| \frac{\hat{y}(i) - y(i)}{y(i)} \right|$$

## V. CONCLUSION

For the first time, this study shows that the prediction model (Deep Learning Method) are able to accurately forecast Global Solar Radiation based on hydrological, geographical parameters etc, confirming the power of machine-based prediction in solving one of the oldest problems in solar power industry. But the other prediction model gives less accuracy with the future prediction data because of the different machine learning techniques have the different mechanism of prediction but machine learning approach is better than any other traditional methods and the ease of future data findings and the time required for prediction work is very frequent than other methods. As automated data collection is becoming routine, developing, training and testing such predictive models is applicable to identify solar radiation, thereby minimizing losses. So we can say that machine learning is the best method as per their ease of working and the accuracy that we have got from the machine learning is great.

## REFERENCES

[1] K.R. Petrovski, M. Trajcev, G. Buneski, A review of the factors affecting the costs of bovine mastitis, J. S. Afr. Vet. Assoc. 77 (2006) 52–60.
[2] H. Seegers, C. Fourichon, F. Beaudeau, Production effects related to mastitis and mastitis economics in dairy cattle herds, Vet. Res. 34 (2003) 475–491.
[3] W. Vanderhaeghen, T. Cerpentier, C. Adriaensen, J. Vicca, K. Hermans, P. Butaye, Methicillin-resistant Staphylococcus aureus (MRSA) ST398 associated with clinical and subclinical mastitis in Belgian cows, Vet. Microbiol. 144 (2010) 166–171.
[4] D. Watson, M. McColl, H. Davies, Field trial of a staphylococcal mastitis vaccine in dairy herds: clinical, subclinical and microbiological assessments, Aust. Vet. J. 74 (1996) 447–450.
[5] A. Henderson, C. Hudson, A.J. Bradley, V. Sherwin, M.J. Green, Prediction of intramammary infection status across the dry period from lifetime cow records, J. Dairy Sci. 99 (2016) 5586–5595.
[6] S.C. Archer, A.J. Bradley, S. Cooper, P.L. Davies, M.J. Green, Prediction of Streptococcus uberis clinical mastitis risk using Matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF MS) in dairy herds, Prev. Vet. Med. 144 (2017) 1–6.
[7] D.B. Jensen, H. Hogeveen, A. De Vries, Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis, J. Dairy Sci. 99 (2016) 7344–7361.
[8] S. Sharifi, A. Pakdel, M. Ebrahimi, J.M. Reecy, S.F. Farsani, E. Ebrahimie, Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle, PLoS One 13 (2018), e0191227.




10.22214/IJRASET

INDEX COPERNICUS

45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429


# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY