Journal of Data Acquisition and Processing
ISSN: 1004-9037
https://sjcjycl.cn/
DOI: 10.5281/zenodo.7702430

# THE ROLE OF MATHEMATICS IN MACHINE LEARNING

**Dr. Mithun B Patil**
Associate Professor, N K Orchid College of Engineering, Solapur, Maharashtra, India.
ORCID: 0000-0001-7741-3645 Email:mithunbpatil2@gmail.com *

**Prof. Santosh Jadhav**
Assistant Professor, N K Orchid College of Engineering, Solapur, Maharashtra, India

**Dr. Sopan Talekar**
Assistant Professor, Department of Computer Engineering, MVPS's KBT College of Engineering, Nashik, Maharashtra, India

**Dr Vipul V Bag**
Professor, N K Orchid College of Engineering, Solapur, Maharashtra, India

**Statements and Declarations**

- The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
- The authors declare the following financial interests/personal relationships which may be considered as potential competing interests

**ABSTRACT**
Machine learning(ML) is the field of Computer Science that uses different models for prediction, classification, and analysis. Machine learning is a Mathematical model prerequisite by aggregation of linear algebra, statistics, calculus, and probability. In this paper, the art of mathematics in machine learning by referring to a different research article from repositories is identified and the need for mathematics for building machine learning models to perform calculations operations such as matrix manipulation of the data for which linear algebra is preferred. To build an uncertainty model in machine learning we prefer probability for learning model and create analysis on a given training dataset for creative prediction. ML is intrinsically data-dependent and driven. The data captured from different sources are improper and consist of much invalid information to perform the statistical analysis we need to perform data preprocessing and validation during which calculus plays a vital role in making the data ready with minimum error. The contribution of the paper is to identify the apogee of mathematical rules required for building an appropriate model.
**Keywords**: Machine Learning, Statistics, Probability, Calculus, Mathematical Model

## 1 INTRODUCTION

In Computer Science there are many tasks and problems which are strenuous using computer programs, traditional methods and instructions. Creating a dynamic game, and desktop application are enormous and complex, making the best machine for a person, making a self-driving car where the computer recognizes objects, is not so easy. These are not things that

computers cannot easily do. The world is full of data produced by different sources such as computers, phones and other devices in the form of images, music, words, spreadsheets, and videos, and it doesn't seem to be slowing down anytime soon. People analyze data and implement a system to examine an opportunity in a data outline, but the redundant and ever-growing nature of data makes it difficult to do this by hand. One way to turn this around is to make the computer learn and figure out how to improve itself through many exercises. This is machine learning (ML). Effective use of high-performance computing architectures, unmatched execution power, and productive use of statistical tools for data extraction are factors that make ML is trending technology for the scientific computing community.

ML algorithms use assumptions based on scientific models, including calculus statistics, and probability. ML promises to derive meaning from all data. ML is a mixture of tools and technologies that can be used to answer questions about your data. ML helps various e-commerce companies offer the right products to their customers to analyze the review of the customer and improve the product. To be able to do this, we need to combine it with mathematics, which has a lot of programming.

## 2 MATHEMATICAL MODEL OF A SYSTEM

The mathematical Model of a system is a core part of engineering specially in computer science & allied branches such as Machine Learning, Artificial intelligence, and Deep learning which play a major role in establishing relationships between parameters, state variables and decision variables. These are also used to set and evaluate the performance of the system. The formulation of the Mathematical model is based on the system parameter which needs to be identified and correlated. The events in the model need to be automatically generated using the parameters and the variables which result in a change in the state variable need to be recorded with system performance criteria estimated.

## 3 MACHINE LEARNING PROCESS

Machine learning has enabled pc frameworks to begin mastering without being explicitly modified. ML has given completely new abilities inside the regions of detection, estimation, prediction and category. The purpose of schooling an ML version is to create a version that first-class answers the goal question nearly every time. Following have been the principle ranges to recommend a model for a given trouble thru ML [5]. schooling in an ML version must first fulfill the records that should be taught along the version. consequently, step one of ML is to collect exceptional information. because high quality and quantity immediately decide the effectiveness of the version. now and again the statistics require tunings and handling like standardization, duplication and error correction. To do this, records training is finished, which comprises loading and dealing out information to make it suitable to be used within the model. The processed facts are then fragmented into parts, one essential element for training and the opposite for performance assessment. the next step is the selection of the version, wherein all of the algorithms which can be nicely tailored to the nature of the trouble are considered. model education then reinforces the model's predictive capacity through the use of education facts. The take a look at data is then used for assessment by using the proposed version as proven in Fig 1.

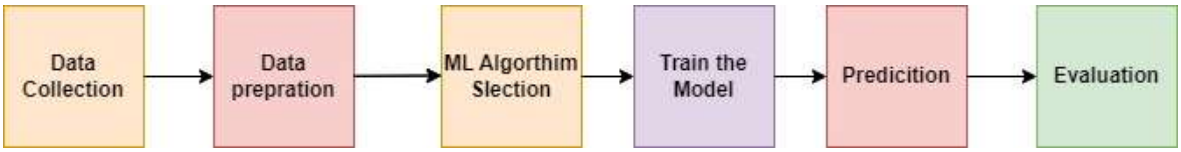


**Figure 1: Machine learning Process**

**Step 1.** Data Collection

Data from various sources are collected and tabulated where quantity and quality are directly proportional to the accuracy of the spin model.

**Step 2.** Data preparation

The collected are preprocessed and missing values of the data are accumulated and made ready for the model.

**Step 3.** ML model selection

There are many machine learning algorithms where we need to select appropriate models as per the requirement which gives appropriate solutions to the given problem.

**Step 4.** Train the Model

The selected model in step 3 is trained by a large set of data. The new data is checked with a model for a particular task like prediction or classification etc.

**Step 5.** Prediction

In this step, the prediction or classification of new objects or data is carried out.

**Step 6.** Evaluation

The model results are compared with actual results to check the errors and accuracy of the results.

**Step 7. MATHEMATICS FOUNDATION FOR MACHINE LEARNING**

ML proposes distinctive models that enforce distinctive findings, forecasting, and classification algorithms. The prediction can be what an item looks like in a photograph, what the next given load is likely to be, or what the excellent combination of drugs to treat a particular disorder is likely to be. ML is an embedded math requirement. It is used to understand why something works or why one version is better than another. Additive mathematics used in ML.ML works

through a mixture of linear algebra, statistics, calculus, and probability [7].

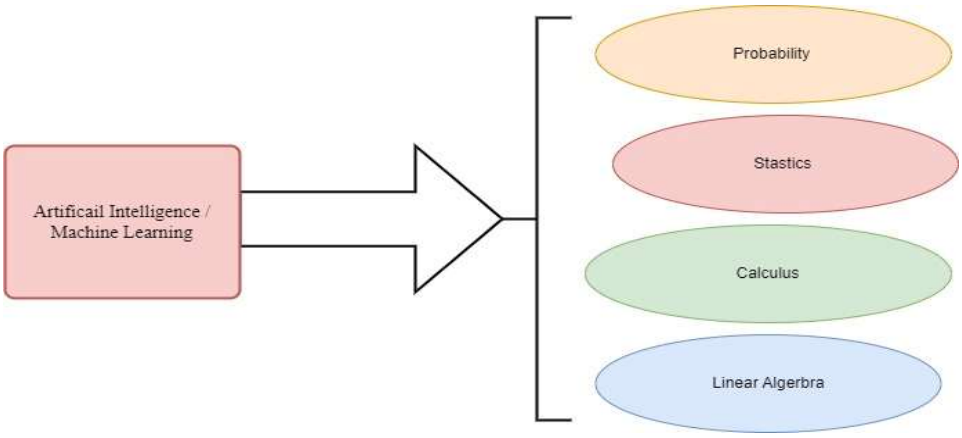


**FIGURE 2: Mathematical components in Machine Learning**

### 3.1 Statistics

Statistics is the core of everything [8]. It has been extensively used for the valuation of the value of a inhabitants parameter. Calculus expresses how to make a model learn and how a model can be optimized. Linear algebra makes running these algorithms feasible on enormous data sets [13].

**TABLE 1. Data set used to create a model of apartment cost estimator concerning the area of the apartment**

| Price Per Square Foot Price of Plot | Price Per Square Foot Price of Plot |
|---|---|
| 85 | 534760 |
| 65 | 535717 |
| 70 | 833333 |
| 15 | 728377 |
| 95 | 899945 |
| 80 | 914339 |
| 60 | 403601 |
| 55 | 437328 |

Probability assistance prediction of event occurrence. Let us consider a problem to predict the price of a plot in terms of the area of the Plot. Considering a data set, having 2 columns, one is the price per square foot of a given plot and the other is the total price of the plot. This is the only marker to predict the price of the plot as a whole. Here present some kind of correlation in the data set. The predictive model gives an idea about the correlation between dependent and independent variables which ultimately improves the capability of a model to predict the

price of the Plot as a whole provided the price per square foot. The graph shown in Fig. 3, has an x-axis measuring the price per square foot and a y-axis measuring the price of the house. It is a scattered plot. Ideally, a line can be found that intersects as many data points as possible. This line can be used for prediction.

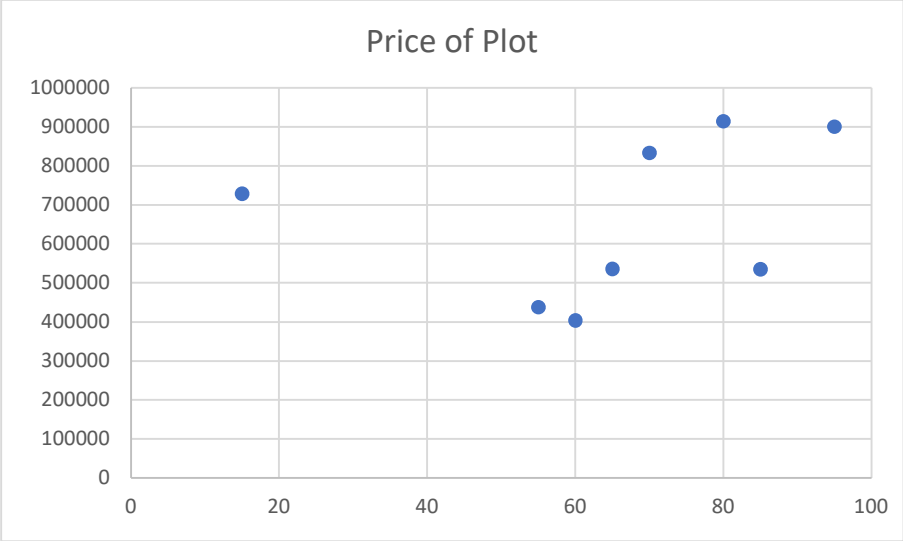


FIGURE 3. Graph showing the correlation between Price per square foot and Price of the Plot

In mathematics, the field of statistics acts as a collection of technologies that extracts useful information from data. It's a tool for creating an understanding of a set of numbers. Statistical inference is a process of predicting a larger population of data based on a smaller sample. In statistics, we try to create a line so we use a statistical inference technique called linear regression. This allows us to summarize and study the relationship between dependent and independent variables. The way linear regression is represented is by using equation 1

$$y=mx + c \qquad (1)$$

Varying each of parameters, m and b, produces different linear models that define different input-output mappings. Where y is the prediction (dependent variable) based on the input variable x (independent variable). The point of intersection of the line with the y-axis is represented by 'b' and, 'm' the slope defines the relativeness of variables. So the value of y can be predicted using the value of x provided the value of m and c is already known. The naive way to find the value of m and c is to try out a bunch of different values. To check the correctness of these values error functions are used. The error function tells how far off the actual value is from the predicted value. There are lots of different types of statistical error functions, least-squares one of them.

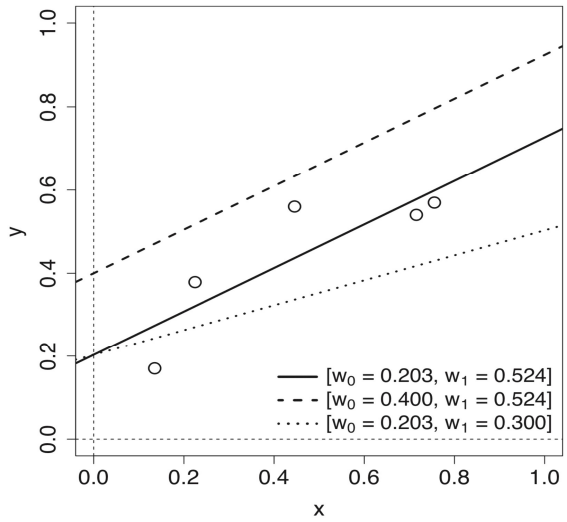


Figure 4: Line Plot of Different Line Models Produced by Varying the Slope and Intercept Taken from Deep Learning

## 3.2 PROBABILITY AND STATISTICS

Probability is the measure of the likelihood of something. Here the data does not consist of values but it is classified in different categories or classes. Logistic regression is a probability technique that can be used to find the probability of occurrence rather than predicting a value. Since the probability goes between 0 and 100, a threshold is chosen past that point it is more likely to have more probability. This gives an S-shaped curve by a sigmoid function. Optimizing this function will plug in input data and get a probability class value. Below are a few Probability tools available for building the model

| **Probability Tool Used in ML Model** | |
|---|---|
| Maximum likelihood | Decision-making in training |
| log loss and cross-entropy | Model fitness |
| Receiver Operating Characteristic curve, or ROC curve | Model evaluation |
| Bayesian optimization | configure hyper-parameters |

Probability theory is one of the major tools in building the model of predictions where probably and unlikely are very common in daily life. To predicate appropriate occurrence we use probability theory. In probability occurance of any event P(A) were ia a particular event and P(A) represent probability ocuurance od that event . The result of the occurnace of that result may be success or failure. For Example Picking a marable from sack which consists of balck and white marble or tossing a coin has event were occurance of head or tail is 0.5.

### 2.2.1 Bayes' Theorem

Bayes' theorem gives the relationship between any two event occurance like selling the bread with milk in an shopping store. Data mining is a field of Computer science that establish a relation between a such event and gives prior knowledge about the particular event to occur from a large dataset.

$$P(A|B) = \frac{P(B|A)*P(X)}{P(B)} \qquad (2)$$

Were A and B events $P(A|B)$is the conditional probability that event A occurs when event B Have already occurred.

## 4 CALCULUS

Calculus is the study of change. It helps to find the direction of change. In which direction should the unknown variable change such that the prediction is more optimal and the error is smaller? It got an optimization technique called gradient descent that will help to discover the minimum value iteratively. It uses the error for a given data point to compute the gradient of an unknown variable and the gradient can be used to update two variables. Then move on to the next point and repeat the process over and over again till the minimum value is found. If multiple variables are considered, then that is called multi-variant regression. The branch of math that consult the multivariate spaces and the linear transformation between them is called linear algebra. It provides a set of operations that can be performed on groups of numbers known as matrices. Now the data set consists of M by I matrices, where samples have 'I' features. Each variable has a weight. Calculus includes all gears for analysis of the association amid capacities and their information sources. Regularly, in AI, we are attempting to discover the sources of info that empower a capacity to best match the information. Neural systems are one of the most well-known and fruitful reasonable structures in AI. They are developed from an associated snare of neurons and enlivened by the structure of natural minds. The conduct of every neuron is affected by a lot of control boundaries, every one of which should be streamlined to best fit the information. The multivariate chain rule can be utilized to ascertain the impact of every boundary of the systems, permitting them to be refreshed during preparation. Multivariate math is required to construct numerous normal AI procedures. A wide variety of data sets is derived from devices such as biochip transponders on farm animals, automobiles with built-in sensors, smart homes, smart cities, or airplanes with sensors, heart monitoring implants, and electric clams in coastal waters. Efficient ML algorithms for such data sets can use hypotheses based on mathematical models involving both calculus and statistics. These devices or sensors used inside physical, biological, or environmental systems collect large volumes of data that follow mathematical models based on both calculus and statistics. Calculus can be used to implement learning from patterns. Various combinations of states and control are used by the ML model for analysis [15], [16].

Calculus plays an integral role in understanding the internal workings of machine learning algorithms, such as the gradient descent algorithm for minimizing an error function. Calculus provides us with the necessary tools to optimize complex objective functions as well as functions with multidimensional inputs, which are representative of different machine learning applications. A neural network model, whether shallow or deep, implements a function that maps a set of inputs to expected outputs. The function implemented by the neural network is learned through a training process, which iteratively searches for a set of weights that best enable the neural network to model the variations in the training data.

## 5 Multivariate Calculus

We recurrently attempt to predict a variable that is dependent on multiple variables. For an instant, we want to predict the weather in certain locations which are dependent on multiple factors. If the result of your function a, is dependent on the given input variable b, then it's declared as given in equation 3.

$$a = f(b) \quad (3)$$

Consequently, if the result of the variable c depends on b and c as an instance, then the function is given equation (4)

$$c = f(b, c) \quad (4)$$

Multivariate calculus has many applications in machine learning such as

- support vector algorithms, and multivariate calculus castoff to find the maximal margin.
- In the Expectation-Maximization algorithm, it is situated used to find the maxima.
- The optimization problems trust the multivariate calculus.
- In gradient descent, it is used to invent the local and global maxima.

## CONCLUSION

Machine learning is a trending technology that is booming in the field of computer science with building different models such for prediction, classification, audio or video recognition etc. Building the model of machine learning depends on mathematical concepts such as probability, calculus, and statistics. The role of the mathematical is very important for building appropriate models and getting accurate results with minimum error. Probability is used for prediction and classification-related models. Calculus is used for multivariate dependent results were statics is used in classification and sampling in Machine learning.

## REFERENCES

[1]. Khaliq, Abdul & Olumoyin, Kayode. (2022). Mathematics of Machine Learning. 10.13140/RG.2.2.25118.41284.

[2]. Linear Algebra. (2020). Mathematics for Machine Learning, 8-56. doi:10.1017/9781108679930.004.

[3]. Probability and distributions. (2020). Mathematics for Machine Learning, 152-200. Vector calculus. (2020).

[4]. Mathematics for Machine Learning, 120-151. doi:10.1017/9781108679930.007:10.1017/9781108679930.008.

**[5].** Karamitsos, Ioannis and Albarhami, Saeed and Apostolopoulos, Charalampos, "Applying Develops Practices of Continuous Automation for Machine Learning", Multidisciplinary Digital Publishing Institute, vol 11, 2020, pp 363 3.

**[6].** Srinivas Pyda, Srinivas Kareenhalli, "Mathematics and Machine Learning, International Conference on Mathematics and Computing", 2018, pp. 135-153 2

**[7].** Sanjeev Arora, "Mathematics of Machine Learning: An Introduction", Proceedings of the International Congress of Mathematics, 2019, pp. 377_390

**[8].** . 4. Md. Kosher, "A combination of Mathematics, Statistics, and Machine Learning to Detect Fraud, National Mathematics Conference", Bangladesh, 2020.

**[9].** Gennady Grabarnik, Luiza Kim-Tyan, Serge Yaskolko, "Addressing Prerrequisited for STEM Classes Using an Example of Linear Algebra for a Course in Machine Learning", The Twelfth International Conference on Mobile, Hybrid, and On-line Learning, 2020, pp. 21_26

**[10].** Somya Goel, Sanjana Rosshan, Rishabh Tyagi, Sakshi Agarwal, "Augur Justice: A Supervised Machine Learning to predict Outcomes of Divorce court cases", Fifth International Conference on Image Information Processing", 2019, pp. 280-285.

**[11].** Kevin Fong-Rey Liu, Jia-Shen Chen," Prediction and assessment of student learning outcomes in calculus A decision support of integrating data mining and Bayesian belief networks", 3rd International Conference on Computer Research and Development,2011, IEEE.

**[12].** Aaron N. Richter,b,∗, Taghi M. Khoshgoftaar, "A review of statistical and machine learning methods for modeling cancer risk using structured clinical data", Artificial Intelligence In Medicine, vol 90, 2018, pp. 1_14

**[13].** Zhenru Wang, Tijie Shi, " Prediction of the Admission Lines of College Entrance Examination based on machine learning", 2nd IEEE International Conference on Computer and Communications", 2016, pp. 332-335

**[14].** Azmoodeh, A. Dehghantanha, M. Conti, and K.-K. R. Choo, "Detecting crypto-ransomware in IoT networks based on energy consumption footprint", J. Ambient Intell. Humanized Comput., vol. 9, no. 4, 2018, pp. 1141_1152.

**[15].** N. Milosevic, A. Dehghantanha, and K.-K. R. Choo, "Machine learning aided Android malware classi_cation", Comput. Elect. Eng. , vol. 61, 2017, pp. 266_274 .

**[16].** Mohan S Acharya, Asfia Armaan, Aneeta S Antony, "A Comparison of Regression Models for Prediction of Graduate Admissions", Second International Conference on Computational Intelligence in Data Science, 2019, IEEE.

**[17].** Sankhwar, S., Gupta, D., Ramya, K.C. et al. Improved grey wolf optimization-based feature subset selection with fuzzy neural classifier for financial crisis prediction. Soft Comput 24, 101–110 (2020). https://doi.org/10.1007/s00500-019-04323-6.

**[18].** Le, DN., Parvathy, V.S., Gupta, D. et al. IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification. Int. J. Mach. Learn. & Cyber. (2021). https://doi.org/10.1007/s13042-020-01248-7.

**[19].** Anupama, C.S.S., Sivaram, M., Lydia, E.L. et al. Synergic deep learning model–based automated detection and classification of brain intracranial hemorrhage images in wearable networks. Pers Ubiquit Comput (2020). https://doi.org/10.1007/s00779-020-01492-2.

**[20].** Shankar, K., Sait, A. R. W., Gupta, D., Lakshmanaprabu, S. K., Khanna, A., & Pandey, H. M. (2020). Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model. Pattern Recognition Letters, 133, 210-216.

**[21].** Ghosh, S., Rana, A. and Kansal, V., "A Statistical Comparison for Evaluating the Effectiveness of Linear and Nonlinear Manifold Detection Techniques for Software Defect Prediction" International Journal of Advanced Intelligence Paradigms (IJAIP), 12(3/4), pp 370-391, ISSN 1755-0394, Inderscience, DOI 10.1504/IJAIP.2019.098578, 2019

[22]. R. B. Kadam, G. G Mali and B. S. Mohite, (2013). Analytical application of poly[dibenzo-18-crown-6] for chromatographic separation of thorium(IV) from uranium(VI) and other elements in glycine medium, J. Radioanal. Nucl. Chem.295 (1) 501-511.

[23]. R. B. Kadam, G. G Mali and B. S. Mohite, (2012). Development of analytical method for chromatographic separation of copper(II) using poly[dibenzo-18-crown-6]. Int. J. Analytical and Bioanalytical Chemistry ,2 (2) 139-146.

[24]. K. R. Mahanwar, S. R. Sabale, R. B. Kadam and B. S. Mohite, (2012). Reversed-phase column extractive separation of Gd(III) with poly[dibenzo-18 crown-6],Int. J. Inorganic Chemistry, 2012.

[25]. Kadam R. B. and Mohite B. S. (2014). Analytical application of poly[dibenzo-18-crown-6] for chromatographic separation of Cobalt(II) in glycine medium, Res. J. Chem. Environ.18 (3),7-18.

[26]. R.Kadam, S. Jadhav, P. Kamble, N. Madane, (2022). Chromatographic Extraction of palladium(II) ions in multi-component mixture using poly[dibenzo-18-crown-6] as sorbent, Journal of Emerging Technologies and Innovative Research,9(4),141-154.

[27]. N. S. Madane, R. B. Kadam, P. N. Kamble, (2022). Development of Reliable Analytical Method for Extraction and Separation of Zinc(II) in Kerosene, The International journal of analytical and experimental modal analysis, XIV(VII),1039-1042.

[28]. R. Kadam, S. Jadhav., (2022). Reversed-Phase Column Extractive Separation of Palladium(II) with Poly[dibenzo- 18-crown-6]: A analysis of real sample,Science, Technology and devolpement journal, XI (VII),160-164.

[29]. P.N. Kamble, N. S. Madane R.B. Kadam, (2023). Liquid-liquid extraction of Barium(II) using Cyanex 301 in Kerosene ,The International journal of analytical and experimental modal analysis ,XV (I),8-21.

[30]. Vijayalakshmi A, Vidyavathy Balraj (2017), 'Optical, thermal, laser damage threshold, dielectric studies and z-scan technique of novel semiorganic NLO material: sodium boro succinate (NaBS)', U.P.B. Scientific Bulletin- Series B, vol. 79, issue 1, pp. 221-232.

[31]. Vijayalakshmi A, Vidyavathy Balraj, Determination of basic solid state parameters and characterization of optical, dielectric and fluorescence properties of Calcium Boro Lactate(CaBL) , Journal of chemical society, Pakistan. Vol. 38 Issue 6, pp. 1092-1097.

[32]. Vijayalakshmi A, Vidyavathy Balraj, G. Vinitha, (2016),"Crystal structure, growth and nonlinear optical studies of Isonicotinamide p-nitrophenol: A new organic crystal for optical limiting applications", Journal of crystal growth, 448 pp.82-88. doi: https://doi.org/10.1016/j.jcrysgro.2016.05.002

[33]. Vijayalakshmi A, Vidyavathy Balraj , Vinitha G. , (2016) "Structure and characterization of a new organic crystal for optical limiting applications, isonicotinamide bis-p-aminobenzoic acid", Ukranian J. Phys. Opt., Volume 17, Issue 3.pp. 98-104.

[34]. Vijayalakshmi A, Vidyavathy, B, Peramaiyan, G & Vinitha, G (2017), 'Synthesis, growth, structural and optical studies of a new organic three dimensional framework: 4-(aminocarbonyl) pyridine 4 (aminocarbonyl) pyridinium hydrogen L-malate, Journal of Solid State Chemistry, vol. 246, pp. 237-244. doi: https://doi.org/10.1016/j.jssc.2016.11.025

[35]. Vijayalakshmi A, Vidyavathy Balraj , B. Gunasekaran, Abdul Razack Ibrahim, Synthesis, Structural, Optical, Thermal and LDT Characterization of Novel Semi-Organic

Non-Linear Optical Material: Calcium Borolactate, Asian Journal of Chemistry; Vol. 28, No. 12 (2016).

[36]. Padmaja, D. L., Nagaprasad, S., Pant, K., & Kumar, Y. P. (2022). Role of Artificial Intelligence and Deep Learning in Easier Skin Cancer Detection through Antioxidants Present in Food. Journal of Food Quality, 2022.

[37]. Padmaja, D. L. (2021). Performance Analysis of Different Architectures on Face Mask Detection. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(13), 377-381.

[38]. Gundu, K. S., Dhyaram, L. P., Ramana Rao, G. N. V., & Surya Deepak, G. (2023, January). Comparative Analysis of Energy Consumption in Text Processing Models. In Advancements in Smart Computing and Information Security: First International Conference, ASCIS 2022, Rajkot, India, November 24–26, 2022, Revised Selected Papers, Part I (pp. 107-116). Cham: Springer Nature Switzerland.

[39]. Ramirez-Asis, E., Guzman-Avalos, M., Mazumdar, B. D., Padmaja, D. L., Mishra, M., Hirolikar, D. S., & Kaliyaperumal, K. (2022). Metaheuristic Methods for Efficiently Predicting and Classifying Real Life Heart Disease Data Using Machine Learning. Mathematical Problems in Engineering, 2022.

[40]. Padmaja, D. L., Tammali, S., Gajavelly, N., & Reddy, K. S. (2022, May). A comparative study on natural disasters. In 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC) (pp. 1704-1709). IEEE.

[41]. Padmaja, D. L., Sruthi, B. S., Deepak, G. S., & Harsha, G. S. (2022, April). Analysis to Predict Coronary Thrombosis Using Machine Learning Techniques. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 21-27). IEEE.

[42]. Padmaja, D. L., & Sriharsha, G. K. (2022, December). Challenges in Crop Selection Using Machine Learning. In Artificial Intelligence and Data Science: First International Conference, ICAIDS 2021, Hyderabad, India, December 17–18, 2021, Revised Selected Papers (pp. 66-76). Cham: Springer Nature Switzerland.

[43]. Padmaja, D. L., Nagaprasad, S., Pant, K., & Kumar, Y. P. (2022). Role of Artificial Intelligence and Deep Learning in Easier Skin Cancer Detection through Antioxidants Present in Food. Journal of Food Quality, 2022.

[44]. Baker, M. R., Padmaja, D. L., Puviarasi, R., Mann, S., Panduro-Ramirez, J., Tiwari, M., & Samori, I. A. (2022). Implementing Critical Machine Learning (ML) Approaches for Generating Robust Discriminative Neuroimaging Representations Using Structural Equation Model (SEM). Computational and Mathematical Methods in Medicine, 2022.

[45]. Lakshmipadmaja, D., & Vishnuvardhan, B. (2018). Classification performance improvement using random subset feature selection algorithm for data mining. Big Data Research, 12, 1-12.

[46]. Padmaja, D. L., & Vishnuvardhan, B. (2018). Evaluating the influence of parameter values on the performance of random subset feature selection algorithm on scientific data. Data & Knowledge Engineering, 117, 174-182.

[47]. Padmaja, D. L., & Vishnuvardhan, B. (2016, February). Comparative study of feature subset selection methods for dimensionality reduction on scientific data. In 2016 IEEE 6th International Conference on Advanced Computing (IACC) (pp. 31-34). IEEE.

[48]. Dhyaram, L. P., & Vishnuvardhan, B. (2018). RANDOM SUBSET FEATURE SELECTION FOR CLASSIFICATION. International Journal of Advanced Research in Computer Science, 9(2).

[49]. Padmaja, D. L., & Vishnuvardhan, B. (2014). Survey of dimensionality reduction and mining techniques on scientific data. International Journal of Computer Science & Engineering Technology, 1(5), 1062-6.

[50]. Padmaja, D. L., & Vishnuvardhan, B. INFLUENCE OF DATA GEOMETRY IN RANDOM SUBSET FEATURE SELECTION.

[51]. Lakshmi Padmaja, D., & Vishnuvardhan, B. (2019). Variance-based feature selection for enhanced classification performance. In Information Systems Design and Intelligent Applications: Proceedings of Fifth International Conference INDIA 2018 Volume 1 (pp. 543-550). Springer Singapore.

[52]. Padmaja, D. L., Surya Deepak, G., Sriharsha, G. K., & Ramana Rao, G. N. V. (2021). Ensemble Methods for Scientific Data—A Comparative Study. In Information and Communication Technology for Competitive Strategies (ICTCS 2020) Intelligent Strategies for ICT (pp. 587-595). Singapore: Springer Nature Singapore.

[53]. Nagaprasad, S., Padmaja, D. L., Qureshi, Y., Bangare, S. L., Mishra, M., & Mazumdar, B. D. (2021). Investigating the impact of machine learning in pharmaceutical industry. J. Pharm. Res. Int., 33, 6-14.

[54]. Sriharsha, G. K., Padmaja, D. L., Rao, G. R., & Deepa, G. S. (2022, December). A Modified Approach of Hyper-parameter Optimization to Assess The Classifier Performance. In 2022 IEEE Pune Section International Conference (PuneCon) (pp. 1-9). IEEE.