# Analysis of Image Classification using SVM

Sai Surya Teja Gontumukkala[1]
*Computer Science*
*Amrita Vishwa Vidyapeetham*
Bengaluru, India
saisurya029@gmail.com

Yogeshwara Sai Varun Godavarthi[1]
*Computer Science*
*Amrita Vishwa Vidyapeetham*
Bengaluru, India
gysaivarun31@gmail.com

Bhanu Rama Ravi Teja Gonugunta[1]
*Computer Science*
*Amrita Vishwa Vidyapeetham*
Bengaluru, India
ravitejagonugunta150@gmail.com

Subramani R[2]
*Mathematics*
*Amrita Vishwa Vidyapeetham*
Bengaluru, India
r_subramani@blr.amrita.edu

K Murali[2]
*Mathematics*
*Amrita Vishwa Vidyapeetham*
Bengaluru, India
k_murali@blr.amrita.edu

***Abstract*— Image classification is one of the classical image processing problems. There are various approaches such as Support Vector Machine, Artificial Neural Networks, Convolutional Neural Networks, K-Nearest Neighbors and Decision Tree for solving this problem. In this paper Support Vector Machine (SVM) is used to classify Images and we are trying to understand SVM and then understand how to draw a decision boundary and try to make it optimal and use it for classification. Two Datasets "Dogs Vs Cats" and "Color Classification" are used in this paper.**

**Keywords—SVM, Threshold, Kernel, Multi Class Classification**

## I. INTRODUCTION

In Artificial Intelligence, SVM is supervised learning model which is associated with learning algorithms, used to analyze data, recognize patterns, classification and regression analysis. SVM has its roots in Statistics, gained prominence due to their robust nature [1-2], accuracy and its effect on even small datasets [5]. Generally, SVMs are binary linear classifiers but also, they can be developed to work on non-linear classification tasks using kernel trick [7,12,11]. Two Datasets "Dogs Vs Cats" and "Color Classification" are used in this paper. The "Dogs Vs Cats" dataset is comprised of photos of dogs and cats. The dataset is divided as training set and testing set and the SVM model is trained using training set and the testing set is used for predicting whether the image is of a cat, or a dog and the Accuracy of the model will be calculated. This Classification is binary and "polynomial" kernel is used to make this classification. The "Color Classification" dataset is comprised of photos of 8 different colors. Even This dataset is divided as training set and testing set and the SVM model is trained using training set and the testing set is used for predicting whether the color in the images and the Accuracy of the model will be calculated. This is a Multi Class Classification [3,13] and kernel is not used to make this classification.

## II. RELATED WORKS

In [1], Theory of SVM, its formulation and implementation were discussed and experiments were done with some variation. In [2], the SVM approaches one against one and one against all are applied on land cover mapping and they are evaluated. In [5], image classification is done using SVM and CNN, results were compared and evaluated. In [7], SVM combined with the deep quasi-linear kernel to classify images of large-scale datasets. In [8], SVM is used to classify real time burning images. In [9], it is discussed about performing Multi Class Classification in an efficient manner by the usage of Binary tree Architecture. In [10], SVM is solved using Kernel technique to classify satellite images. In [12], Non-Linear SVM is used to classify encrypted data. Previously, classifying images using SVM and its techniques were discussed on a single dataset. In this paper SVM is performed on two different datasets with two different sizes i.e., one is of larger size and the other one is very small, the results of both datasets are analyzed to know on which SVM works better.

## III. MATH CONCEPTS

Let us consider a 1D Data of observations which classifies as getting an admission or not based on GRE Score of a student.

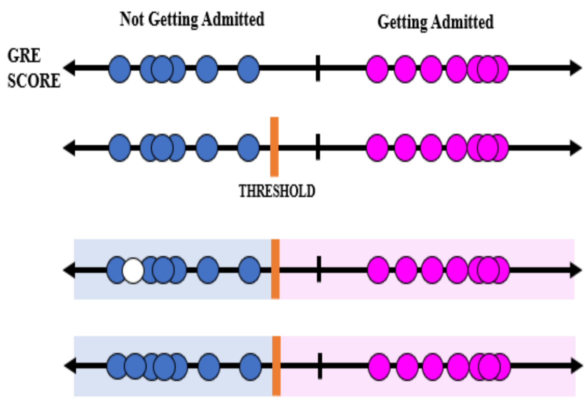


Fig. 1. Threshold

As shown in the Fig. 1, first we will take a threshold which classifies the Observations into getting admitted and not getting addmitted. When a new observation is given, if it lies left to the threshold it is classifed as not getting addmitted and if it lies right it is classified as getting admitted. As shown in the Fig. 1 the new observation is classified as Not getting admitted and in the Fig. 2 the next new observation is classified as getting admitted.

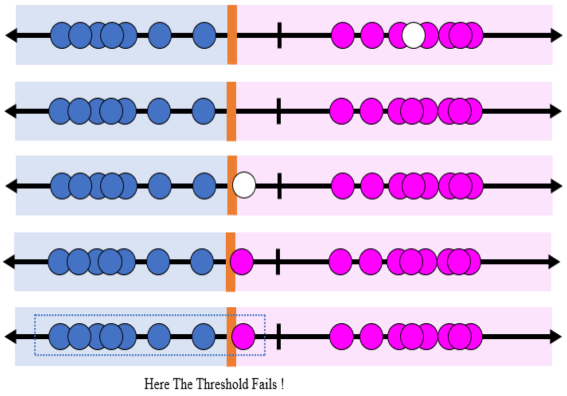


Fig. 2. Failure of Threshold

As shown in the Fig. 2, the last new observation is classified as getting admitted, but it is very near to the not getting admitted observations and far from the getting admitted observations, to overcome this we will take the mid point of the observations as threshold as shown in Fig. 3 so when a new observation comes same as the previous one it will be classified as not getting admitted.

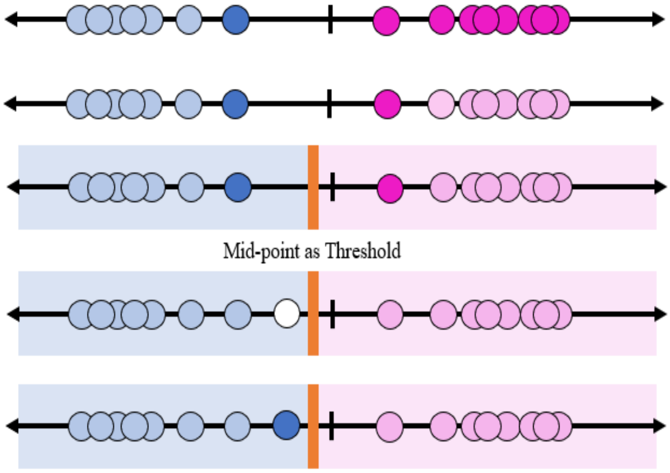


Fig. 3. Mid-Point as Threshold

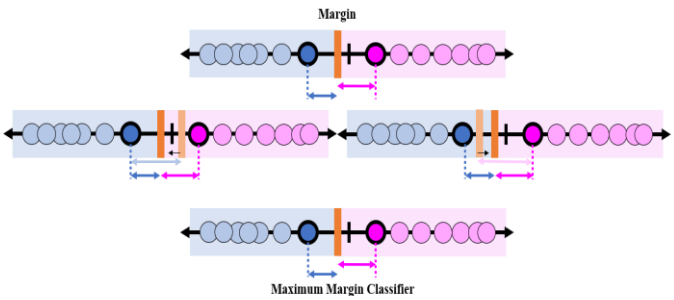


Fig. 4. Maximum Margin Classifier

The minimum distance between the data points and the threshold is known as margin, in the above case as shown in Fig. 4 if we move the threshold left or right the margin gets minimized so the above classifier is known as maximum margin classifier. Now suppose we have an outlier in our data and when maximum margin classifier is used, the margin is very close to getting admitted observations and far from the not getting admitted observations and if a new observation comes as shown in the Fig. 5 it will be classified as not getting admitted even though it is far from the not getting admitted observations, so Maximum Margin Classifier is sensitive to outliers.

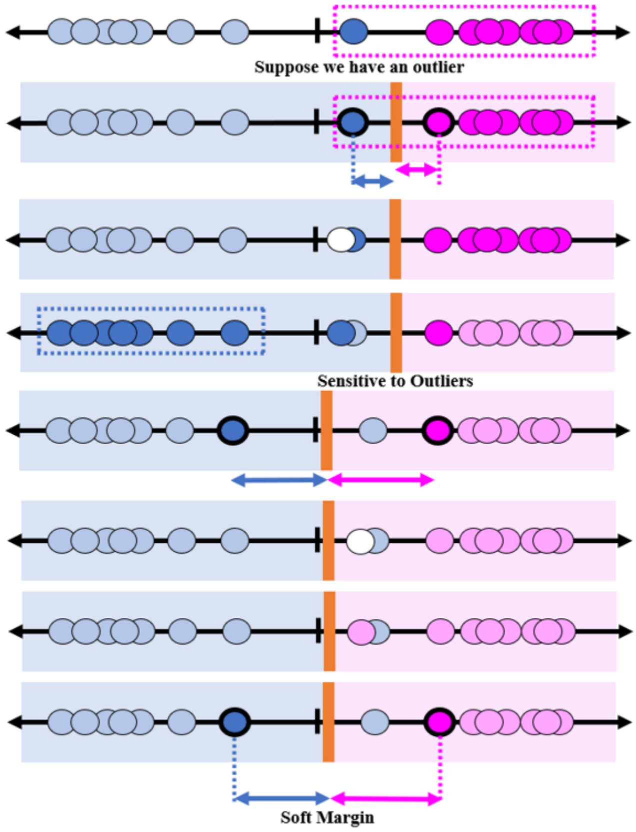


Fig. 5. Soft Margin

Now, we take the threshold as the midpoint of the points shown in Fig. 5 by allowing misclassifications and when a new observation comes it is classified as getting admitted. So, in the previous classification bias was less but it had more variance whereas here the bias is more, but it has low variance i.e., it works well for testing data. Then this margin is called as soft margin.

There will be many soft margins out of which one best soft margin will be chosen by cross validation, let us assume that the following soft margin was best. So, when we take threshold such as it allows misclassifications but works well for testing data then it is called as soft margin classifier a.k.a. Support Vector Classifier.

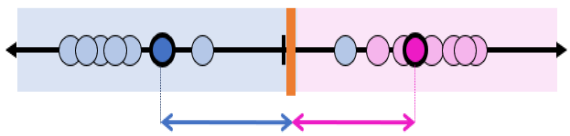

Fig. 6. Support Vector Classifier!!!

Suppose we have a 2D Data i.e., 2 parameters GRE Score, CGPA for getting admitted.

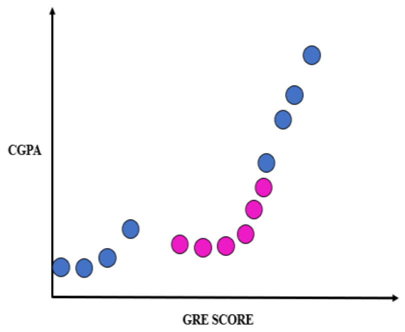


Fig. 7. 2D Data

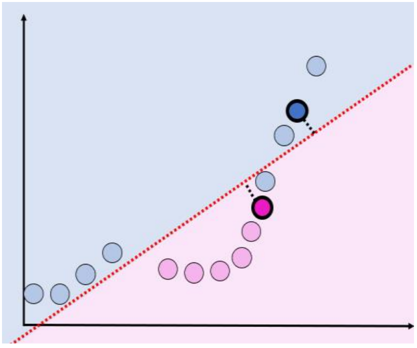

Fig. 8. Line as a Threshold

Then the threshold would be a line instead of a point and the observations which lie on one side of the line are classified as getting admitted and which lie on the other side are classified as not getting admitted.

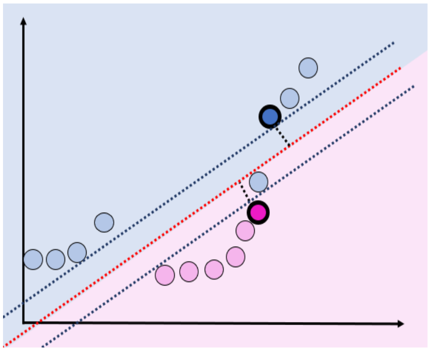

Fig. 9. Distance from the nearest Points

Suppose we have a 3D Data i.e., 3 parameters GRE Score, CGPA and TOEFL Score for getting admitted [6].

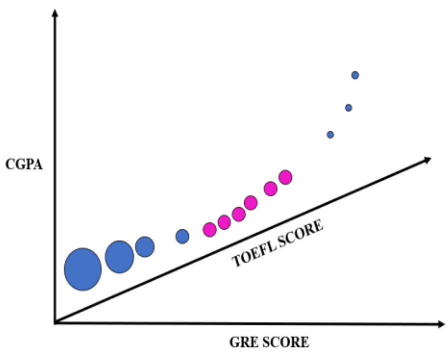


Fig. 10. 3D Data

Then the threshold would be a plane instead of a line and the observations which lie on one side of the line are classified as getting admitted and which lie on the other side are classified as not getting admitted [4].

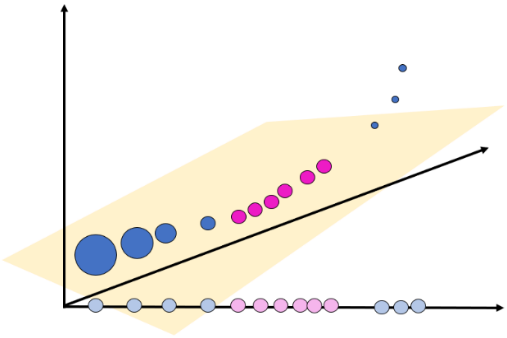

Fig. 11. Plane as a Threshold

If the data is in even higher dimension than 3-D then the threshold used for classification will be hyperplane [8].

## A. *Polynomial kernel*

Polynomial kernel is a type of kernel function which is commonly used with SVMs. Basically; it computes the relationships between the pair of observations.
Let's see an example by considering an example of a dataset. Here the blue points are considered as patients not cured and purple points represents patients cured. The patients who took high dosage and low dosage didn't get cured.

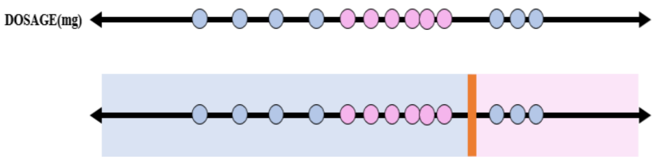


Fig. 12. Drug Dosage Data

As this training dataset has many overlaps, it is difficult to find a support vector classifier which satisfies and helps in separating the patients who are cured from the patients who are not cured.

So, we square original dosage and give them as each point on y-axis original dosage measurements. An SVM along with polynomial kernel trick implemented in order to compute the higher dimensional relationships between the observations and then find a support vector classifier which satisfies based on the high dimension relationships.

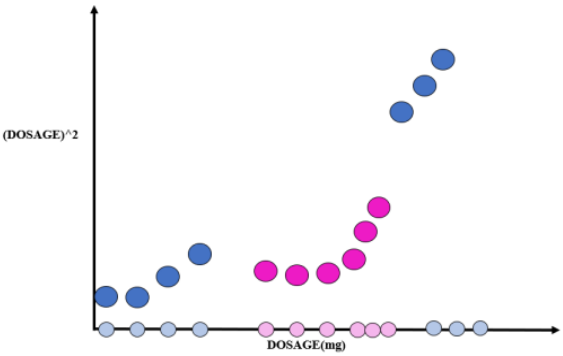


Fig. 13. Plotting data into higher dimension

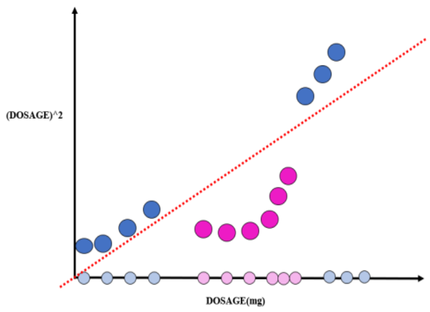

Fig. 14. Drawing the threshold

The kernel notation is defined as

$$(a \times b + r)^d$$

Degree
Coefficient of the polynomial
Observation 2
Observation 1

Where a, b are the two observations, d is the degree of the polynomial and r is the coefficient of the polynomial. For example, let's consider r= ½ and d=2 since we are squaring the term.

$$\left(a \times b + \frac{1}{2}\right)^2 = \left(a \times b + \frac{1}{2}\right)\left(a \times b + \frac{1}{2}\right)$$

$$= ab + a^2b^2 + \frac{1}{4}$$

$$\left(a, a^2, \frac{1}{2}\right) \cdot \left(b, b^2, \frac{1}{2}\right)$$

The dot product gives the high dimensional coordinates for the data, 1st terms(a) represent x-axis coordinates, the 2nd terms(b) represent y-axis coordinates and the 3rd terms represents the z-axis coordinates. Since both 3rd terms are same, we can ignore them.
It implies that if we need to calculate the high dimensional relationships, we need to calculate dot product between each data point.
For example, if we want to find out the high dimensional relationships between the 2 observations shown below. We find the dot product of the points and we get 5,256.25 so this is an example where we have to establish the relation of the data points in 2 dimension and find the support vector classifier even though we didn't transform the data completely to 2-dimension.

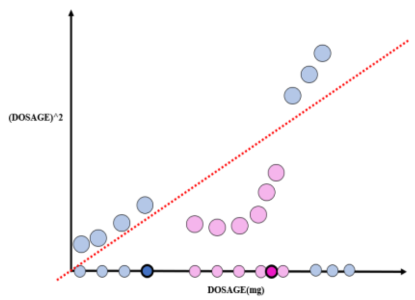

Fig. 15. Finding the relationship between two data points

$$(6 * 12 + \frac{1}{2})^2 = (72 + \frac{1}{2})^2 = 72.5^2 = 5{,}256.25$$

So, polynomial kernel is used to compute the relationships between pairs of observations and r and d are determined using cross validation. Then after deciding the values of r and d we just insert the observations in the kernel and calculate the high dimensional relationships [4,6,10].

### *B. Multi Class Classification*

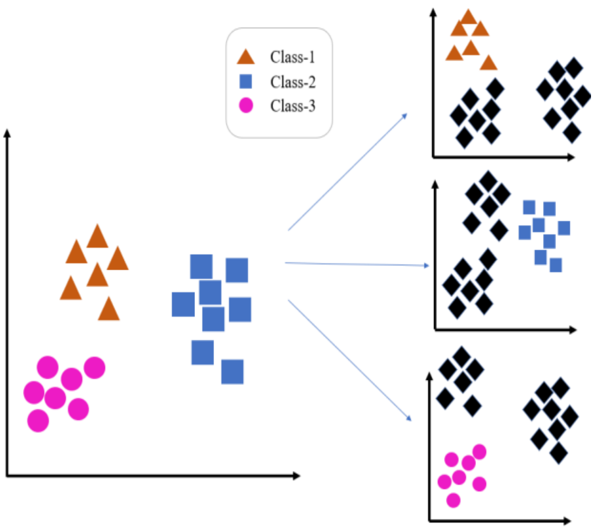

Fig. 16. Multi Class Classification

In the Fig. 16 there are 3 classes, we take each class at a time and considered the rest of the 2 classes as one i.e., as binary and it finds the probability for the newly coming observation in the 3 classifications and the class with highest probability will be given as the final classification [2, 9].

## IV. IMPLEMENTATION

In this implementation two datasets were used, one is "Dogs Vs Cats" Dataset, it contains pictures of dogs and cats so this classification is binary and since this data will be a scattered one i.e., we can't just decide by color or size or anything on that matter of fact to classify it as a dog or cat, so here the kernel trick would be applied i.e., polynomial kernel. Whereas "Color Classification" Dataset pictures of contains 8 different colors so this is a multi-class classification and since this data can be classified without any kernel because color for which same kind of data will be at same place i.e., not scattered. So, for both SVM models will be trained accordingly with its data split into 75 % for training and the rest for testing. Since "Dogs Vs Cat Dataset" is a large Dataset Pickle is used for faster computation.

## V. RESULTS

SVM gave 80 percent accuracy on "Dogs Vs Cats" Dataset and the Fig. 17 below is the table showing top 20 elements of original testing set and the predicted ones.

| | original | predicted |
|---|---|---|
| 0 | 1 | 0 |
| 1 | 1 | 1 |
| 2 | 0 | 1 |
| 3 | 0 | 1 |
| 4 | 0 | 1 |
| 5 | 0 | 0 |
| 6 | 1 | 0 |
| 7 | 1 | 1 |
| 8 | 1 | 1 |
| 9 | 1 | 1 |
| 10 | 0 | 0 |
| 11 | 1 | 1 |
| 12 | 0 | 0 |
| 13 | 0 | 1 |
| 14 | 0 | 1 |
| 15 | 0 | 0 |
| 16 | 1 | 0 |
| 17 | 1 | 0 |
| 18 | 0 | 1 |
| 19 | 0 | 0 |

Fig. 17. Output of Dogs Vs Cats

SVM gave 96 percent accuracy on “Colors Classification” Dataset and the Fig. 18 below is the table showing top 20 elements of original testing set and the predicted ones.

| | original | predicted |
|---|---|---|
| 0 | 7 | 7 |
| 1 | 6 | 6 |
| 2 | 0 | 0 |
| 3 | 3 | 3 |
| 4 | 7 | 7 |
| 5 | 3 | 3 |
| 6 | 0 | 6 |
| 7 | 1 | 1 |
| 8 | 1 | 1 |
| 9 | 2 | 2 |
| 10 | 5 | 5 |
| 11 | 3 | 3 |
| 12 | 7 | 7 |
| 13 | 1 | 1 |
| 14 | 6 | 6 |
| 15 | 6 | 6 |
| 16 | 2 | 2 |
| 17 | 7 | 7 |
| 18 | 0 | 0 |
| 19 | 1 | 1 |

Fig. 18. Output of Colors

## VI. CONCLUSION

SVM is one of the best suitable algorithms for Image Classification as it works with binary, multi classification as well as linear and non-linear classifications using Kernel technique and gives good accuracy even with smaller data sets and this paper shows about the real time image classification for identification of a cat or a dog and identification of a color using SVM. It might not always give the best accuracy for large datasets as they have more possibility of having more overlaps and it even takes large amount of time for large datasets.

## REFERENCES


[1] Theodoros Evgeniou and Massimiliano Pontil, "Support Vector Machines: Theory and Applications" 2001 Machine Learning and Its Applications, Advanced Lectures, DOI:10.1007/3-540-44673-7_12

[2] Anthony Gidudu, Gregory Hulleyand and Tshilidzi Marwala, "Image classification using SVMs: One-Against-One Vs One-against-All" 2007 Computing Research Repository (CoRR), vol.8(5), pp.544-550.

[3] Bhaswitha Maddula, Manikantha Vallabhaneni, Sarada Jayan, Subramani R, 2020. “Chaotic Evolutionary Algorithm”, *2020 IEEE International Conference for Innovation in Technology*, 1 – 4. DOI: 10.1109/INOCON50539.2020.9298233.

[4] Josh Starmer, Support Vector Machines, Clearly Explained!!!, STATQUEST!!!, Oct. 2019. Accessed on: Apr. 4 2021. [online]. Available: https://statquest.org/support-vector-machines-clearly-explained/

[5] S. Y. Chaganti, I. Nanda, K. R. Pandi, T. G. N. R. S. N. Prudhvith and N. Kumar, "Image Classification using SVM and CNN," 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), 2020, pp. 1-5, doi: 10.1109/ICCSEA49143.2020.9132851.

[6] S. Tamuly, Jyotsna C, and Amudha, J., “Deep Learning Model for Image Classification”, Advances in Intelligent Systems and Computing, vol.1108. Springer International Publishing, Cham, 2019.

[7] P. Liang, W. Li, D. Liu and J. Hu, "Large-scale image classification using fast SVM with deep quasi-linear kernel," 2017 International Joint Conference on Neural Networks (IJCNN), 2017, pp. 1064-1071, doi: 10.1109/IJCNN.2017.7965970.

[8] T.S. Hai, L.M. Triet, L.H. Thai, and N.T. Thuy, “Real Time Burning Image Classification Using Support Vector Machine” 2017 EAI Endorsed Transactions on Context-aware Systems and Applications, DOI:10.4108/eai.6-7-2017.152760

[9] G. Madzarov and D. Gjorgjevikj, "Multi-class classification using support vector machines in decision tree architecture," IEEE EUROCON 2009, 2009, pp. 288-295, doi: 10.1109/EURCON.2009.5167645.

[10] S. Manthira Moorthi, I. Misra, R. Kaur, N. P. Darji and R. Ramakrishnan, "Kernel based learning approach for satellite image classification using support vector machine," 2011 IEEE Recent Advances in Intelligent Computational Systems, 2011, pp. 107-110, doi: 10.1109/RAICS.2011.6069282.

[11] Manikantha Vallabhaneni, Bhaswitha Maddula, Sarada Jayan, Subramani R (2020) “Chaotic Hooke-Jeeves Algorithm using Cubic map with MATLAB code”, *2020 IEEE International Conference for Innovation in Technology*, 1 – 4, DOI: 10.1109/INOCON50539.2020.9298244.

[12] Anthony Barnett, Jay Santokhi, Michael Simpson, Charlie Stainton-Bygrave,Nigel P. Smart, Srnivas Vivek and Adrian Waller, “Image classification using non-linear Support vector Machines on Encrypted Data” 2017

[13] Shekhar Pandey, Supriya M, and Abhilash Shrivastava, Data Classification Using Machine Learning Approach, 2018 Advances in Intelligent Systems and Computing, vol.683, pp.112-122