International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-2, December, 2019

Classification of Spinal Muscle Atrophy Disease using SVM in Machine Learning

B. Ganga Bhavani, G. L. N. V. S. Kumar, M. L. Rekha, B. P. N. Madhu Kumar, Raja Rao P. B. V.

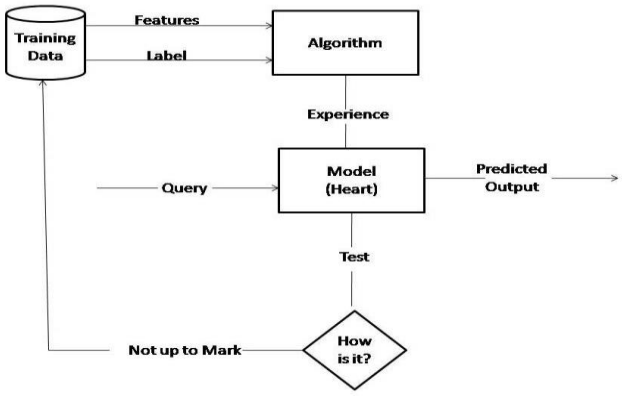
Abstract: ***SMA is a genetic neuromuscular disease. It is a rare disease. It is caused by mutations in the survival motorneuron (SMN) gene that encodes SMN Protein. Maindifficult area of SMA is muscle weakness, causing withdifficulty with moving, swallowing or breathing. Thereare four types of SMA's. The primary objective of thispaper is to classify the SMA's by using support vectormachine classifier. Then we can easily predict the life span of the children based on the group of SMA. This disease is classified on the basis of age of onset and clinical course.***

Keywords: SMN1, SMN2, SVM, SVC, CPK, SMA Linear, RBF, Polynomial.

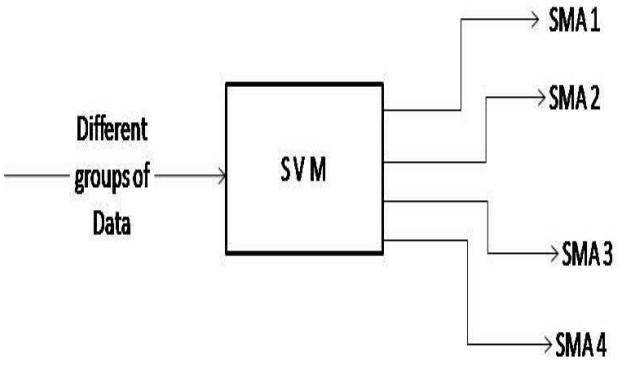
I. INTRODUCTION

The task of constructing groups of SMAs can be categorized into subtasks. There are four different types of SMAs. To classify the groups of SMAs we use Support vector machine algorithm. Support vector machine algorithm is purely supervised machine learning algorithm. In this first we take the training data, then trains the algorithm with proper features and labels then we get important output call it as model. This model is the heart of the machine. This model is trained by the experienced algorithm. We just pass a query to this model. It produces a predicted output. In some cases, the model is not up to mark we can again retrain the model. This process is continuous of testing and training.

This paper here focusses on the different groups of SMA's Such as SMA0, SMA1, SMA2, SMA3, SMA4 using support vector machine algorithm. SMA0: The most severe form. Children usually succumb to the disease before the age of 6 months.



SMA1: Symptoms appear in the first few months of life. Children rarely survive passed their second birthday
SMA2: Serious muscle weakness symptoms usually appears between 7 to 18 months of age.
SMA3: children are able to stand and walk, but worsen with time. Symptoms usually appear after 18 months of age.
SMA4: Not a lifetime threatening condition. Symptoms appear in adulthood.



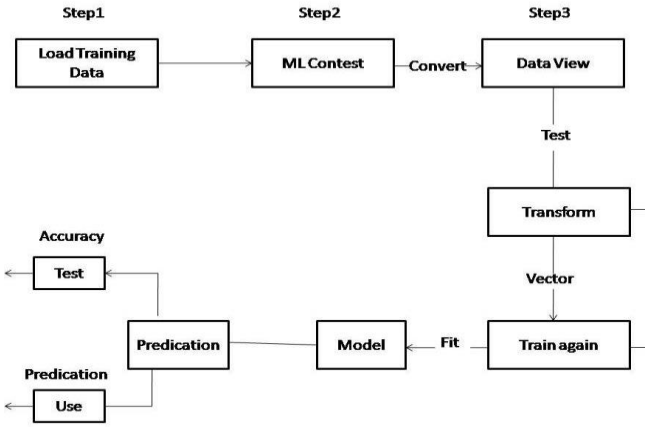
II. SVM ALGORITHM

2.1 The Basis of SVM Algorithm:

Fundamentally, SVM is used for both classification and regression problems.
Support vector machine is a supervised classification method that separates data using hyperplanes.The distance between the support vectors and hyperplanes is maximum.
So margin is always maximum.There are two types of SVM's.Linear SVM and Nonlinear SVM.
We can separate the data into two groups in linear SVM by using straight line..In non linear SVM we can separate the data into two groups by using 3D.
Steps in Supervised learning algorithm:

Revised Manuscript Received on December 30, 2019.
* Correspondence Author
Billa Ganga Bhavani, Department Of CSE, BVC Engineering College, Odalarevu, JNTUK Kakinada, AP, India. E-mail: bhavanicse10@gmail.com
G. L. N. V. S. Kumar, Department of MCA, BVCITS, Amalapuram, JNTUK Kakinada, AP, India. E-mail: kumar4248@gmail.com
M. Lakshmi Rekha, Department of MCA, BVCITS, Amalapuram, JNTUK Kakinada, AP, India. E-mail: rekhamoram@gmail.com
B. P. N. Madhu Kumar, Department Of CSE,BVC Engineering College, Odalarevu, JNTUK Kakinada, AP, India. E-mail:tpo.bvce@bvcgroup.in
Raja Rao P. B. V., Department Of CSE, BVC Engineering College, Odalarevu, JNTUK Kakinada, AP, India. E-mail rajaraopbv@gmail.com

Step1: Load the training data
Step2:Prepare ML context
Step3:Convert the ML conext into Dataview
Step4:transform the text into vector form
Step5:prepare the training algorithm
Step6:fot that trainig and testing the data into the MODEL.
Step7:make the prediction for accuracy



2.2 Support vector machine algorithm:

SVM is used to classify data by a hyper plane that linearly separates the dat from different classes.
Step 1:labelled the sample data
Step 2:hyper plane can be represented by <w,x>+b=0 where w represents weights
Step 3:dimensinal vectors is defined as
<x,y>=x1y1+x2y2+……xnyn
For 2D Plane <w,x>=w1x1+w2x2
Hyperplane is w1x1+w2x2+b=0
Step 4:ForLinear SVM
Yi(<xi,w>+b)-1>=0∀i
The points closest to hyperpane are called support vectors
Min1/2||w||2 such that
yi(xi.w+b)-1>=0∀i
Step 5:For nonlinear SVM
According to soft-margin classification
Step 6:Apply kernal function
According to Linear
According to RBF
K(x,x1)=exp-||x1-x||2/2µ2
According to Polynomial kernal
K(x,x1)=(x.x1+c)d

III. CLASSIFICATION OF SMA DISEASE

SMA is a spinal muscle atropy.the main goal of applying support vectr machine algorithm is to classify the grous of SMA's to srvive the life time of human beings.
Dataset:
SMA calculations of the performance of SVM algorithmare carried out on 50 real world data sets drawn from the Rainbow hospital,NIMBS hospital,neuro muscular disorder university of california at data repositary[ ].The datasets are chosen vary across a number of dimnsions including the type of the blood samples and the number of instances.Thses are datasets with both discrete and continuous data.
After loading the datasets into our model the graphical representation of our data is as shown in table1

Table 1: Loading the dataset

	type	AGE	cpk	GMFM	HFMS	FVC	disease
0	SMA1	0.1	250	22	8	10	1
1	SMA1	0.2	500	12	14	12	1
2	SMA1	0.5	243	14	18	14	1
3	SMA1	0.4	287	15	17	10	1
4	SMA1	0.3	296	17	2	12	1

Statastics:

For the disease SMA following attributes are considered in the calculations which plays an important role for the classification of SMA.
CPK:Creatine phosphokinase(based on muscle biopsy)
GMFM:gross motor function measure
HFMS:Hamersmith functional motor scale
FVC:forced vital capacity
After applying statastical method on the datasests the results are shoen in table2

Table 2: Attribute values after applying statastical methods

	AGE	cpk	GMFM	HFMS	FVC	disease
count	59.000000	59.000000	59.000000	59.000000	59.000000	59.000000
mean	11.218644	1926.135593	23.966102	33.152542	29.067797	2.677966
std	13.062495	1784.400000	15.963283	19.941132	12.455250	1.209740
min	0.100000	224.000000	10.000000	2.000000	10.000000	1.000000
25%	0.750000	624.500000	13.500000	16.500000	17.500000	2.000000
50%	4.000000	1223.000000	18.000000	28.000000	31.000000	3.000000
75%	23.000000	2555.500000	26.000000	46.500000	40.000000	4.000000
max	50.000000	6047.000000	84.000000	84.000000	49.000000	4.000000

Graphical Representation:

After loading the datasets into our model the graphical representation of our data is as shown in fig1

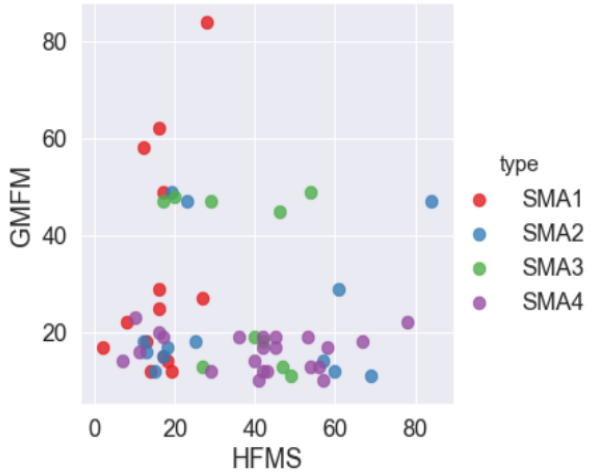


Fig 1: load the data

Fit the Model:

Here we apply the SVM (Support vector machine) to fit the model and apply the SVC (Supprt vector classifier) to classify the data.

After applying support vector machine the hyperplane divides the data into two parts as shown in fig2.

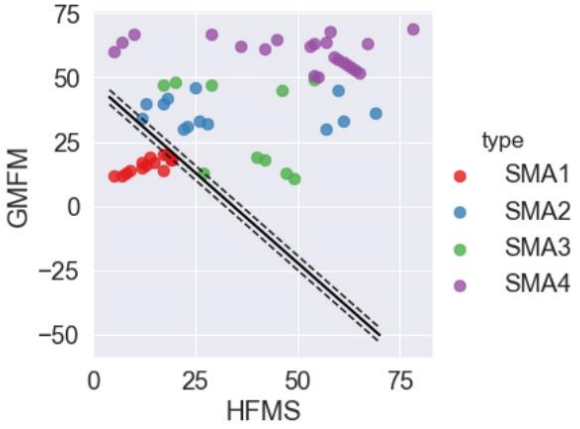


Fig2: Hyperplane divides the data into two sets.

The points which are nearest to dashed line indicates support vectors.The distance between the two hyprplanes indicates the margin.

The darked line indicates decision boundary.

The points which are below the line comes under SMA1.The points which are above the line coes under SMA2,SMA3 and SMA4.

Here SVC(Support vector classifier)classify the data into two parts.For multiclasification we use kernel functions

Applying Kernal functions:

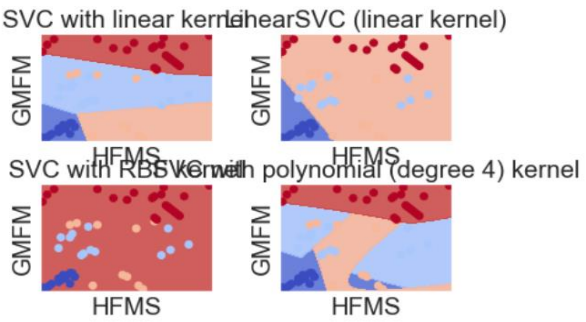
1.SVC with linear kernel

2.linear svc

3.svc with RBF

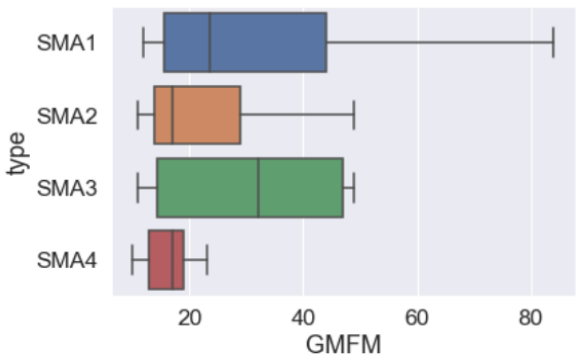
4.SVC with polynomial kernel

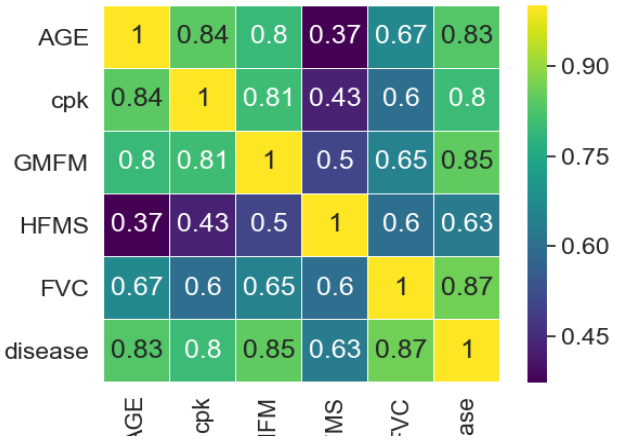
Among all those 4 types of kernel functions SVC with polynomial is the best kernel function to classify the types of SMA's disease

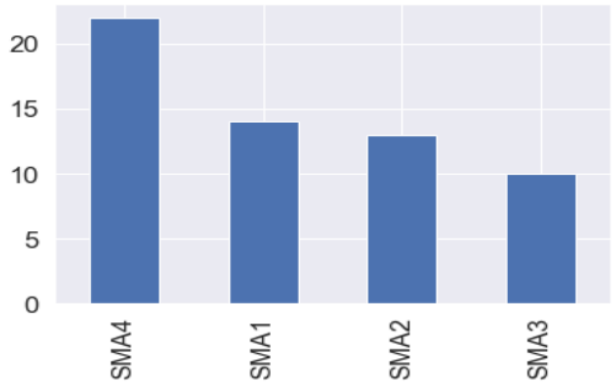


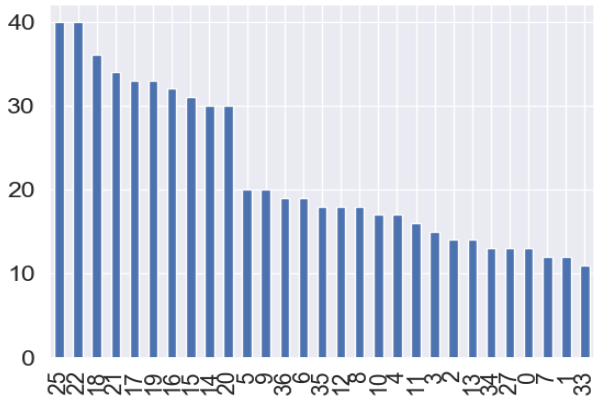
Experimental results:

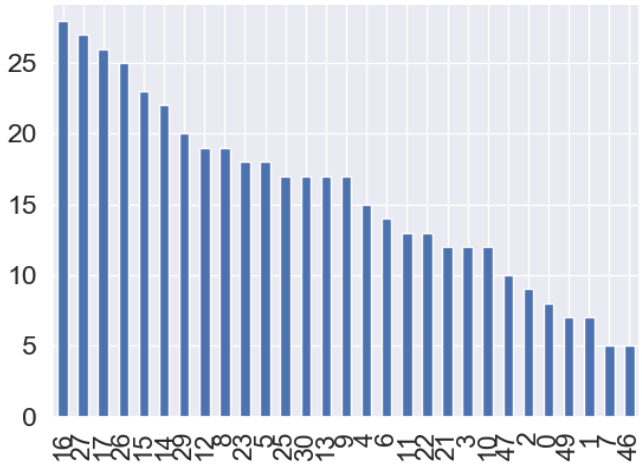
Graphical representation

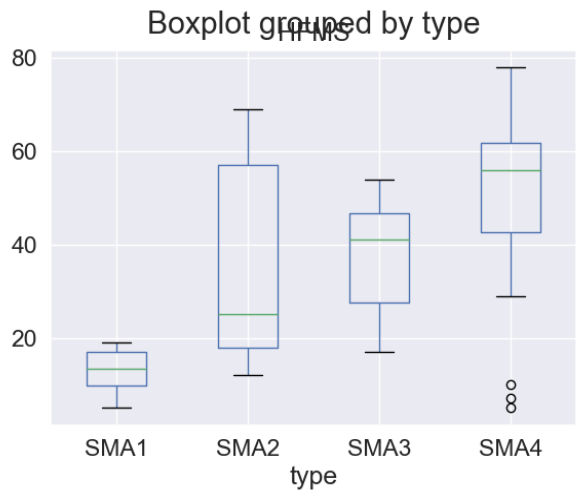


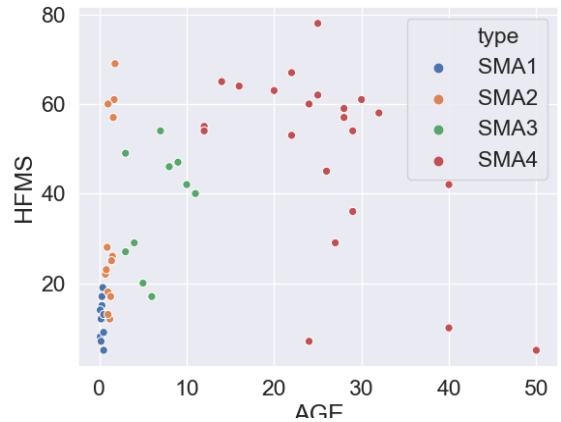


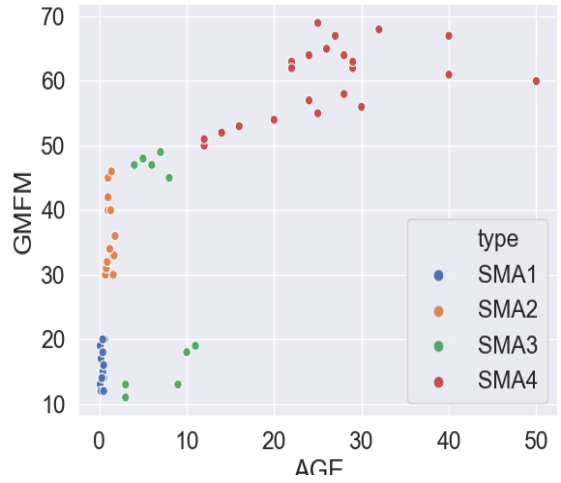












Prediction calculation by using Gaussian NB:
Step1:Load dataset
Step2:organize our data
Step3:split our data
Step3:intialise our classifier
Step4:train our classifier
Step5:make predictions
Step6:evaluate accuracy

```
x=df[['HFMS','GMFM']].as_matrix()
y=df['disease'].as_matrix()
```

```
train, test, train_labels, test_labels = train_test_split(x,
                                                          y,
                                                          test_size=0.33,
                                                          random_state=42)
```

```
gnb = GaussianNB()
```

```
model = gnb.fit(train, train_labels)
```

```
model
```

```
GaussianNB(priors=None, var_smoothing=1e-09)
```

```
preds = gnb.predict(test)
print(preds)
```

```
[1 1 3 1 4 4 4 2 4 1 4 1 3 4 1 2 1 1 4 2]
```

```
print(accuracy_score(test_labels, preds))
```

```
0.95
```

IV. CONCLUSION

By giving the blood samples of children with respect to CPK,AGE,HFMS and GMFM my model easily classify the groups of SMA's and also identify whether the person is effected by which group of spinal muscle antropy.According to this classification we can easily survive the lifespan of children caused by the disease.

REFERENCES

1. The library won the IJCNN 2001 Challenge by solving two of three problems: the Generalization Ability Challenge (GAC) and the Text Decoding Challenge (TDC). For more information,see: http://www.csie.ntu.edu.tw/~cjlin/papers/ijcnn.ps.gz.
2. Bennett, K. P. & Campbell, C. (2000). Support vector machines: Hype or hallelujah? SIGKDD Explorations, 2(2). http://www.acm.org/sigs/sigkdd/ explorations/issue2-2/bennett.pdf.
3. Sch¨olkopf, B., Smola, A., Williamson, R. C., & Bartlett, P. (2000). New support vector algorithms. Neural Computation, 12, 1207–1245.
4. Hagenbuchner M, Cliff D P, Trost S G, Van Tuc N and Peoples G E 2015 Prediction of activity type in preschool children using machine learning techniques J. Sci. Med. Sport 18 426-31
5. N. Krishnaiah, B.Ganga Bavani "Automatically Prospecting Feature for Queries from Their Search Impact", International Journal of Engineering and Advanced Technology (IJEAT),ISSN: 2249-8958, Vol 9, Issue 1, October-2019.
6. Neumann D L and Thomas P R 2009 The relationship between skill level and patterns in cardiac and respiratory activity during golf putting Int. J. Psychophysiol. 72 276-82
7. N. Krishnaiah, "Design of Hierarchy Scheme for Mobile App", International Journal of Recent Technology of Engineering (IJRTE), ISSN: 2277-3878, Vol 8, Issue 1, May-2019.
8. Abdullah M R, Eswaramoorthi V, Musa R M, Maliki M, Husin A B, Kosni N A and Haque M 2016 The Effectiveness of Aerobic Exercises at difference Intensities of Managing Blood Pressure in Essential Hypertensive Information Technology Officers J. Young Pharm. 8000

AUTHORS PROFILE

Mrs. B. Ganga Bhavani, received her B.Tech and M.Tech degree in Computer Science and Engineering, JNTU in 2005, 2009 respectively. Her research interests are in Artificial Intelligence, Deep Learning and Machine Learning

Mr. G. L. N. V. S KUMAR, received his MCA degree, MK University, Madurai in 2010. His research interests are in Artificial Intelligence, Deep Learning and Machine Learning.

Ms. M. Lakshmi Rekha received her M. Tech degree in Computer Science and Engineering, JNTUK in 2015. Her research interests are in Artificial Intelligence, Deep Learning and Machine Learning

Mr. B. P. N. Madhu Kumar received his M.Tech. Degree in CSE, JNTUK in 2009, His research interests are in Artificial Intelligence, Deep Learning and Machine Learning

Mr. RAJA RAO P B V, received his M. Tech degree in CSE, JNTUK in 2013. His research interests are in Artificial Intelligence, Deep Learning and Machine Learning.

.