Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 **ISSN: 2321-8363**
Impact Factor: 4.123

THE MULTIPLE TIME SERIES CLINICAL DATA PROCESSING USING OPTIMIZATION ALGORITHM

Priyanga M[1], Dr. K. Sasi Kala Rani[2] *M.E.,Phd.*, **Yamuna Devi S[3], Pavithra M[4]**
Student, Head of the Department, Student, Student
Department of Computer Science and Engineering, Hindusthan Institute of Technology, Coimbatore, Tamilnadu, India
Priyankame00@gmail.com

Abstract

The main objective of this research is to discover patient acuity or severity of illness has immediate practical use for clinicians.The use of multivariate time series modelling along with multiple model is evaluated. As large-scale multivariate time series data become increasingly common in application domains, such as health care and traffic analysis, researchers are challenged to build efficient tools to analyze it and provide useful insights. In many situations, analyzing a time-series in isolation is reasonable. Initially MMSVM is applied for multiple measurements of time series dataset to classify the accurate results. But it has problem along with imbalanced dataset and it leads misclassification results. In the existing system, to overcome the unbalanced data and classification performance used improved Particle swarm optimization algorithm (IPSO). The unbalanced dataset is handled by using the improved PSO algorithm and it reduced the irrelevant feature in the given time series data. In the proposed system, introduced Modified Artificial Bee Colony Algorithm (MABCA) to solve the multiple time series problems by increasing the selection of optimal feature information. The MABA algorithm is used to improve the most appropriate features globally and global optimization is increased as well as dimensionality reduction. The MABCA reaches lighter designs along with a better convergence rate. In the proposed system, the optimal features are increased by tuning the parameters. The proposed MABCA with Transductive SVM (TSVM) is used to improve the classification performance and Artificial Neural Network (ANN) is used for making long term multi-step prediction. It reduces the training process time and minimize the error rate more significantly. From the experimental result, the conclusion decides that the proposed system is superior to existing system. Application/Improvements: The findings of this work prove that the graph search based method provides better result than other approaches.

Keyword :Data mining, multiple measurements, Transductive support vector machine (TSVM) , particle swarm optimization (PSO), modified artificial bee colony algorithm (MABCA) and artificial neural network (ANN).

1. Introduction

Data mining is the process of mining the patterns form data. Generally, data mining is the search for hidden patterns that could be present in huge databases. Data mining scans via a huge volume of data to find out the patterns and correlations between patterns. Data mining is becoming gradually more important tool to make over this data into information. Data mining requires the use of data analysis tool to determine previously unknown, valid patterns and relationships in huge volume data. Such kind of tool can enclose statistical model, mathematical algorithms and machine learning methods. Thus, data mining consists of more than gathering and running data, it also contains analysis and prediction.

Data mining tools includes variety of tasks. The major functionality of the task is analyzing the data and generates the results. It becomes an increasingly common in both private and public sectors. Associations used the data mining concept and tools for surveying the customer information, avoid fraud and misuse, and help in medical research. Data mining process includes three stages as follows. The initial exploration, model constructing or pattern identification with verification or validation and deployment.

The data mining process is applied in business transactions, medical information data, scientific data, satellite data and software engineering data. The huge volume of data is stored in files, databases and repositories,it is progressively significant in recent years. To increase potentially means then analysis and interpretation of information data will be extracting the interesting knowledge which is used to proper decision making [1] [2].

The data mining process consists of data cleaning, data integration, data selection, data transformation, data mining, pattern assessment and knowledge demonstration. Data cleaning and data integration has been implemented together in a preprocessing step to make a data warehouse. Data selection and transformation could be merged to discover the knowledge representation. Data mining techniques are such as support vector machine (SVM), regression algorithms and optimization algorithms used to predict the decision from the specified dataset efficiently.

2. Analysis of research methodology

In [3], Chao-Hui Lee et.al (2011) suggested a novel data mining method to improve the efficiency and effectiveness of patient monitoring. This method is used for recognizing attacks of chronic diseases through considering of both patients bio-signals and environmental factors. The pattern based decision tree and association rule mining mechanisms are used to integrate sequential pattern mining algorithm to mine asthma attacks features and builds the classifiers effectively. In [4], Themis P. Exarchos et.al (2009) presented an optimized sequential pattern matching technique for sequence classification. The technique introduced in this scenario is used to focus automatically produces a sequence categorization model, depends on the two phase procedures. However this scenario does not produced better results and not received too much awareness.

In [5], Damian Bargiel et.al (2011) suggested the multi temporal classification of agricultural land use based on high resolution spotlight TerraSAR-X images. This scenario opens several opportunities to obtain knowledge about impacts caused through agricultural land utilize and its variations. However, this scenario wants to improve the classification results of single classes instead of class groups. In [6], IyadBatal et.al (2012) suggested a pattern mining approach for classifying multivariate temporal data. The incorporation of classification and pattern extraction approach is currently drawn in the data mining research and also is successfully utilized in static data, graph data and sequence data. In this research scenario, the minimal prognostic temporal patterns approach and efficient algorithm is presented and improved to extract these patterns.

In [7], Mohamed F Ghalwash et.al (2012) discussed about the early classification of multivariate temporal observations using extraction of interpretable shapelets. For the early categorization task, we introduced a technique named as multivariate shapelets detection (MSD). It mines the patterns from all dimensions on the time series datasets. However it has issue along with the running time complexity while incorporating parallelism in the algorithm. In [8], Zhong Yin et.al (2014) suggested identification of temporal variations in mental workload by using several data mining algorithms. This scenario is introduced locally linear embedding (LLE) which is used for discovering the low dimensional manifold in the high dimensional manifold EEG markers from various cortical regions. To identify the mental workload (MWL) to discrete levels by using MWL indicators and small sized training samples, a new EEG approach by merging LLE, support vector clustering as well as support vector data description methods are introduced also estimated with the help of measured data. However, the reliability of the scenario is reduced significantly in this approach.

In [9], ChandrimaSarkaret,al presented improved feature selection mechanism to develop the classification accuracy. In this research scenario, we use rank aggregation based feature selection method to choose appropriate donor genotype features. The efficient data mining approach is used to select the important features to identify the optimal donor for patients. It handles the high dimensional dataset more effectively. In [10], Yi-Ju Tseng et.al discussed multiple time series clinical data processing for classification with merging algorithm and statistical measures. To progress the precision of medical outcome classification using multiple measurements, a novel multiple time series data processing approach with merging algorithm is enhanced. To compare the data merging approach, the classification performance by using processed multiple measurements is compared to classification using single measurements.

In [11], Hayder M et.al suggested particle swarm optimization (PSO) to improve the time series classification data accuracy. The proposed PSO optimization algorithm is focused on the reduction of number of iterations to reach optimal solution. Gaussiam maximum likelihood with PSO is used to reduce the errors significantly and increase the speed of the algorithm. In [12], Nhat-Duc Hoang et.al presented a hybrid approach called as support vector regression with artificial neural network algorithm to deal with time series classification dataset. The artificial neural network algorithm is a searching algorithm which is used to identify the suitable parameters for increasing the classification performances.

Thus this method has been identified as the better technique as the approach provides deep analysis of higher accuracy prediction for the specified datasets. The efficiency of this approach is better than the other techniques and also it provides pathway for further improvement.

3. Materials and Methods

3.1 Pre-processing

In this module, the pre processing technique is performed to obtain the more accurate classification results. Data cleaning is the process of discovering and correcting inaccurate records from the specified dataset. Used mainly in databases, the term refers to identifying incomplete, incorrect, inaccurate, irrelevant, etc. data integration is the process of combine the various information data from heterogeneous data sources but with semantic meaning. It is used to increase the classification accuracy results for the specific query. Data transformation is used to change the set of data values from the source data system to destination data system. By using the pre-processing method, the accuracy of classification performance is increased in terms of reduction of noise and missing values.

3.2 Feature selection

In this module we have to perform the feature selection process on the time series dataset. It is used to provide relevant feature for the training and testing process. To remove the redundant and irrelevant features, the feature selection based random forest is introduced. An ensemble classifier algorithm is enhanced which contains bagging and random feature selection methods. The frequency of a feature's appearance in the classification trees represents the importance of the feature. The library random forest is used to execute the random forest feature selection process. All the features are ranked according to the weight assigned to them by random forest.

3.3 Algorithm for Merging Multiple Features Based on Defined Time Periods

In this module, we have to merge the important features by using the merging algorithm more efficiently. Based on the algorithm 1 we evaluated the time series data. The algorithm is as follows.

Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 **ISSN: 2321-8363**
Impact Factor: 4.123

Algorithm 1

Start

Read D_m= *m* days period

T_{event}= the time of specific event

R_B = all records before T_{event}, sorted by record date in descending order

F_B= all features inR_B

Initialize merged records array $M_m$based on *m* days period and F_B

FOR each record $R_{B\ k}$in R_B, k=1,2…N

T_k = the time of $R_{B,k}$, recorded

i= $T_{event} - T_k/D_m$

$M_{m\ i}$ = the *i*th merged record based on *m* days period

FOR each feature F_q in *RB k* (*q* = 1, . . . , *O*)

Set the value *Wq*of *Fq*in *Mm i* as the most recent value of *Fq*from all the *RB k* in *RB* and *i*th period

ENDFOR

ENDFOR

If statistical measures mode

FOR each period *i* in *Mm*

FOR each time-related feature *Ft* in *FB*

//time-related laboratory data in Supplementary Data 1

*Ft*_ M ax*i*= maximum of all the *Ft* in *RB* within period *i*

*Ft*_ M in*i*= minimum of all the *Ft* in *RB* within period *i*

Ft _A vg *i* = average of all the *Ft* in *RB* within period *i*

Ft _SD *i* = standard deviation of all the *Ft* in *RB* within period *i*

Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 ISSN: 2321-8363
Impact Factor: 4.123

Ft _C or*i*= Pearson's correlation coefficient of all the *Ft* in *RB* within period *i*

Ft _S lp*i* = slope of trend line of all the *Ft* in *RB* within period *i*

Add *Ft* M ax*i*,*Ft* M in*i*, *Ft* A vg *i* , *Ft* SD *i* , *Ft* C or*i*, *Ft* S lp*i* as addition features into the *i*th merged record *Mm i*

ENDFOR

ENDFOR

OUTPUT *Mm*

END

The central idea of this merging algorithm is to choose only one value to stand for a feature in one period. Because the time of the target event, such as therapy for HCC, is set as the key time with regard to data processing, the value that is closer to event time could be more significant than others. Therefore, the most recent value is selected to represent a feature in a period, and therefore some valuable information in the original data might be omitted by the merging algorithm.

3.4 Calculation of statistical measure

In this module, statistical measure is calculated for describing the data distribution in each period. There is a probability that information in the original data, such as the tendency and feature distribution may disappear after data merging. To protect the information, maximum and minimum statistics are used in this scenario. Average is a method for deriving the central tendency of a feature space, and standard deviation is an extensively utilized measurement of variability. Pearson's correlation coefficient is showing, how the feature pair is strongly related within the range of -1 to +1.

3.5 Prediction model establishment

In this module, the data mining approaches are such as support vector machine (SVM) and random forest used for single and multiple measurements respectively. The SVM builds the classification model for a binary class and it uses nonlinear mapping to change the data into higher dimensional data. Along with a suitable nonlinear mapping, two classes are divided through a hyperplane. The library SVM is focused to execute the SVM prediction process. The kernel function with radial basis function is used for SVM model establishment. For multiple measurements, the prediction outcomes are decided through voting method where more features belonged to similar group and majority vote of class is considered as final prediction result.

Algorithm 2

BEGIN

*S*m = the test dataset selected from the merged records based on m days period

P V m = the patient list of test dataset *S*m

*R*m = the training dataset selected from the merged records based on *m* days period

PMm= the predictive model established based on selected features in Rm , and

imported parameters

FOR each patient Pi in $P\ V\ m$

Initialize voting result of Pi ,$VR\ i$ to zero

If the type of predictive model is classification

FOR each merged record $P\ Sm\ i$ of Pi in period i in Sm

$Rm\ i$ = prediction result of $P\ Sm\ i$ by usingPMm

//recurrence = 1, non − recurrence = −1

$VR\ i = Rm\ i + VR\ i$

ENDFOR

If $VR\ i$ >= 0

Predict Pi as a positive case //recurrence

Else

Predict Pi as a negative case //non-recurrence

Else Ifthe type of predictive model is regression

$VR\ i$ = the average of prediction result of all merged record of Pi in period i in

Sm by usingPMm

Predict Pi by $VR\ i$ //regression result

ENDFOR

OUTPUT performance ofPMmbased on the prediction results

END

We evaluated the dataset by using MMSVM and MMRF algorithm efficiently.

Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 ISSN: 2321-8363
Impact Factor: 4.123

3.6 IPSO classification

The algorithm proposed in this work is based on the particle swarm optimization technique. PSO is an optimization algorithm that optimizes a given solutions through applying mathematical rules and after computing the fitness of a current solutions changes their coordinates into the search space. PSO is originally introduced by Kennedy, Eberhard, and Shi as an optimization technique inspired by the social behavior of bird flocks and fish herds. PSO utilizes a certain number of solutions, called particles that form a swarm. Every such particle has position and velocity coordinates in the search space. The velocity represents the change of the particle position from iteration to iteration. The change of the particle's position is dictated by the best so far known particle's position as well from the best position in the overall swarm.

This is used to improve the speed of the process by using important and relevant information features in the dataset. It reduces the number of iterations by selecting the best solutions for time series dataset. The IPSO algorithm is as follows

Algorithm 3

UpdatePSO

{

Do

ForEach Particle in Swarm

For j = 0 to ParticleLength

Partcle.Velocity[j] = W * Partcle.Velocity[j] + C1*R1*Particle.BestPosition[j] - Particle.Position[j] +C2*R2*BestParticle.Position[j] - Particle.Position[j]

EndFor

For j = 0 to ParticleLength

Partcle.Position[j] += Partcle.Velocity[j]

EndFor

CheckCandidate (Particle)

If (Particle.BestInfoGain>BestParticle.BestInfoGain)

BestParticle = Particle

EndIf

EndForEach

OldBestGain = NewBestGain

NewBestGain = GetSwarmBestInformationGain

While ((OldBestGain - NewBestGain) > EPSILON)

BestShapelet = BestParticle

}

IPSO

CheckCandidate(Particle)

{

Distances ← Initialize

ForEachTimeSeries in TrainDataSet_ClassA_And_ClassB

Distance = MinDistance(Particle.Position, TimeSeries)

Distances ← Add(Distance)

EndForEach

Histogram = OrderDistances(Distances)

InforGain = CalculateInformationGain(Histogram)

If (InforGain>Particle.BestInfoGain)

Particle.BestInfoGain = InfoGain

Particle.BestPosition = Particle.Position

EndIf

In this proposed work, improved PSO approach is used for effective summarization process. Particle swarm optimization (PSO) is a computational algorithm that optimizes a problem by iteratively trying to progress a candidate solution along with regard to a given measure of quality.The c1 and c2 are cognitive parameters, r1 and r2 are random parameters. It is used to choose the best solutions from the multiple time series data. PSO optimizes a problem by having a population of candidate solutions, here dubbed particles, and moving these particles around in the search-space according to simple mathematical formulae over the particle's position and velocity. Each particle's movement is influenced by its local best known position but, is also guided toward the best known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

In the proposed system, we introduced improved PSO algorithm to increase the classification accuracy. For the given input datasets, the similarities of multiple features are extracted optimally by using PSO parameters. The main aim of the PSO algorithm is to select the potential and relevant features by generating best fitness function value. Also it is effectively used for multiple time features along with several features. It takes minimum execution time by searching globally and also it updates new best similarity values quickly. Hence it increases the classification accuracy higher for the given specified datasets and best features are retrieved by using improved PSO algorithm more accurately.

3.7 Modified Artificial Bee Colony Algorithm (MABCA)

1. Initialize parameters as n, m, I, a and ec-length

n = Number of employed bees

m = Number of onlooker bees (m>n)

Iteration I : Maximum iteration number

a: initial value of penalty parameter for jth agent

ec-Length: Length of ejection chain neighborhood

2. Build primary employed bee colony results

For each bee evaluate fitness function f = $\sum_{j=1}^{m}\sum_{i=1}^{n} c_{ij}x_{ij} + \alpha\sum_{j=1}^{m} max\left\{\sum_{i=1}^{n} b_{ij}x_{ij} - a_j\right\}$

3. I =0

4. Repeat

5. N = 0

6. Repeat the process which is given below

7. If fit(ShiftNeighbour) <fit(EmployedBee) then

8. Employed Bee = Shift Neighbour

9. If fit(DoubleShiftNeighbour) <fit(EmployedBee) then

10. Employed Bee = DoubleShift Neighbour

11. Discover the probabilities using objective function Pi=$\frac{\Sigma(^1/_{fit})^{-1}}{fit}$

12. Allocate onlooker bees to employed bees

13. For all Onlooker Bees

14. Ejection **-**Chain Neighborhood

15. Find best Onlooker, replace with respective Employed Bee

16. iffit(Best Onlooker)< fit(Employed)

17. Find best Feasible Onlooker, replace with Best solution,

18. if fit (BestFeas Onlooker)

19. N=N+1;

20. until (N=employed bee)

21. I = I+1

22. until (I=maxIteration)

The algorithm describes the concept of universal gravitation into the consideration of the affection among employed bees and the onlooker bees. By allocating diverse values of the control parameter, the universal gravitation is concerned for the artificial bee colony algorithm while there are different quantities of employed bees and the single onlooker bee. Consequently, the investigation capability is converted about on typical in this algorithm.

3.8 Artificial Neural Network Algorithm (ANN)

Input: N training samples,

Output: predicted class

For each sample do

Input[i] =sample

Foreach i in neural metwork do

Output [i] = module. ForwardPropagate(input[i])

Input[i+1]=output[i]

End

Predictedclass = criterion (output)

If training then

Check error attributes

For each [k-i] in neural network do

Output = module. Backward Propagate

Input[i+1]=output[i]

End

End

End

For the number of input training sample we have to perform the analysis. For each input sample the neuron network perform the output sample based on the prior knowledge. It is used to search the more accurate results along with hidden layer and this layer is used to map the similarity for corresponding input sample. Hence it avoids the error values also it progress the speed of process more efficiently.

4. Results and Discussion

In this section, the performance metrics are evaluated using existing and proposed methodologies. The performance metrics are such as accuracy, precision and recall. The existing random forest, support vector machine and improved PSO algorithm is used to classify the multiple measurement of specified dataset. However the existing system has shown the lower performance in the classification results. The proposed Modified Artificial Bee Colony Algorithm (MABCA) has shown the higher performance in the classification results. The proposed MABCA with Transductive SVM (TSVM) and artificial neural network (ANN) provides superior classification accuracy results. From the experimental result, we can conclude that the proposed system is better than the existing system in terms of higher performance. An experimental result shows that the proposed method achieves superior performance.

4.1. Accuracy

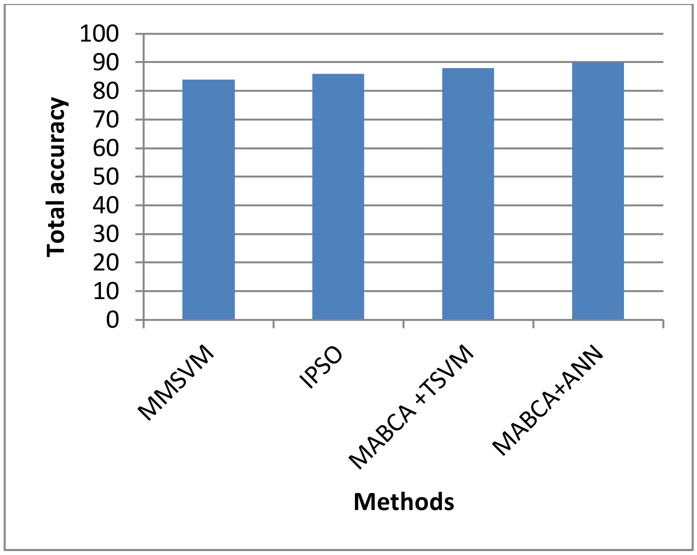
The accuracy is the proportion of true results (both true positives and true negatives) among the total number of cases examined.

Accuracycan be calculated from formula given as follows

Accuracy =

$$\frac{\text{True positive} + \text{True negative}}{\text{True positive} + \text{True negative} + \text{False positive} + \text{False negative}}$$

An accuracy of 100% means that the measured values are exactly the same as the given values.



From the above graph we can observe that the comparison of existing and proposed system in terms of accuracy metric. In x axis we plot the types and in y axis we plot the total accuracy values. The total accuracy values are lower by using existing algorithm of MMSVM and IPSO algorithm. The accuracy value of MMSVM and IPSO is 84% and 86% respectively. The accuracy value is higher by using the proposed of MABCA with TSVM and ANN. The accuracy value of MABCA with TSVM is 88% and MABCA with ANN is 90%. From the result, we conclude that proposed system is superior in performance.

Table 1

Performance metric	MMSVM	IPSO	MABCA +TSVM	MABCA +ANN
Accuracy	82%	85%	89%	90%

From the graph, the values are tabulated in the table 1. It shows that the proposed system is shown higher accuracy values compare than previous algorithm. Thus the result concludes that the proposed system is used to provide efficient classification results.

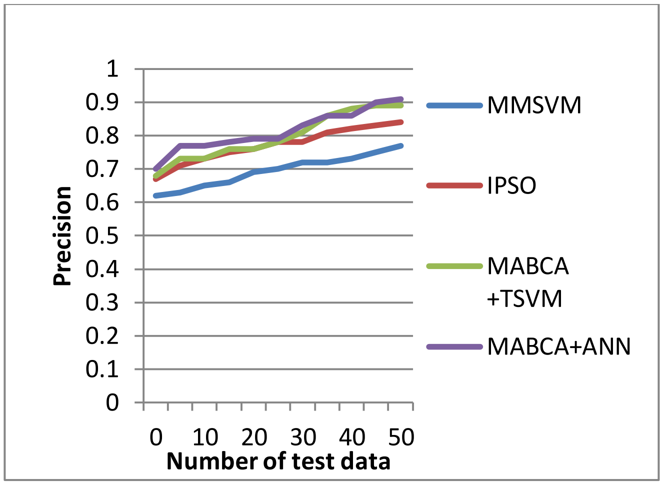
4.2. Precision

The precision is calculated as follows:

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive+False positive}}$$

Precision can be seen as a measure of exactness or quality, whereas recall is a measure of completeness or quantity. In simple terms, high precision means that an algorithm returned substantially more relevant results than irrelevant. In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class).

Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 **ISSN: 2321-8363**
Impact Factor: 4.123



From the above graph we can observe that the comparison of existing and proposed system in terms of recall metric. The Intermediate values can be calculate let f(x) be a continuous function on the interval (0.7,0.86).If d->[f(a),f(b)],then there is a c->[0.7,0.86] such that d=f(c) i.e ., f(c)=0.79.In x axis we plot the types and in y axis we plot the recall values. The recall values are lower by using existing algorithm of MMSVM and IPSO algorithm. The recall value is higher by using the proposed of MABCA with TSVM and ANN algorithm. From the result, we conclude that proposed system is superior in performance.

Table 2

Performance metric	MMSVM	IPSO	MABCA+TSVM	MABCA +ANN
Precision				
0	0.62	0.67	0.68	0.7
5	0.63	0.71	0.73	0.77
10	0.65	0.73	0.73	0.77
15	0.66	0.75	0.76	0.78
20	0.69	0.76	0.76	0.79
25	0.7	0.78	0.78	0.79
30	0.72	0.78	0.81	0.83
35	0.72	0.81	0.86	0.86
40	0.73	0.82	0.88	0.86

From the graph, the values are tabulated in the table 2. It shows that the proposed system is shown higher precision values compare than previous algorithm. Thus the result concludes that the proposed system is used to provide efficient classification results.

Priyanga M *et al*, International Journal of Computer Science and Mobile Applications,
Vol.4 Issue. 6, June- 2016, pg. 54-69 ISSN: 2321-8363
Impact Factor: 4.123

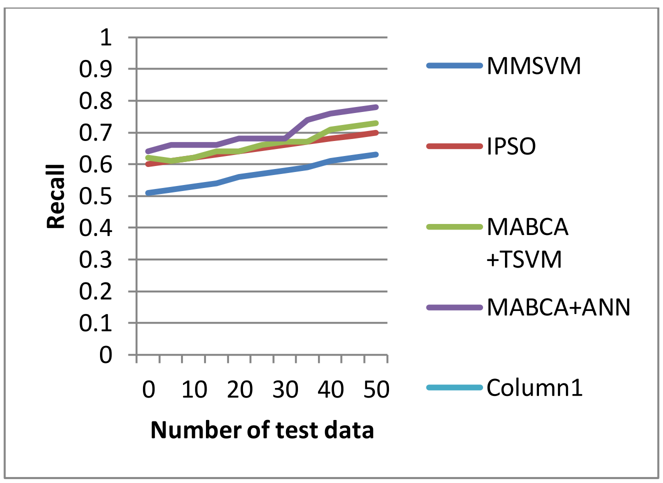
4.3. Recall

The calculation of the recall value is done as follows:

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive+False negative}}$$

The comparison graph is depicted as follows:

Recall is defined as the number of relevant documents retrieved by a search divided by the total number of existing relevant documents, while precision is defined as the number of relevant documents retrieved by a search divided by the total number of documents retrieved by that search. Recall in this context is defined as the number of true positives divided by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been).



From the above graph we can observe that the comparison of existing and proposed system in terms of recall metric. The Intermediate values can be calculate let f(x) be a continuous function on the interval (0.64,0.77).If d->[f(a),f(b)],then there is a c->[0.64,0.77] such that d=f(c) i.e ., f(c)=0.68. In x axis we plot the types and in y axis we plot the recall values. The recall values are lower by using existing algorithm of MMSVM and IPSO algorithm. The recall value is higher by using the proposed of MABCA with TSVM and ANN algorithm. From the result, we conclude that proposed system is superior in performance.

Table 3

Performance metric	MMSVM	IPSO	MABCA+TSVM	MABCA+ANN
Recall				
0	0.51	0.6	0.62	0.64
5	0.52	0.61	0.61	0.66
10	0.53	0.62	0.62	0.66
15	0.54	0.63	0.64	0.66
20	0.56	0.64	0.64	0.68
25	0.57	0.65	0.66	0.68
30	0.58	0.66	0.67	0.68
35	0.59	0.67	0.67	0.74
40	0.61	0.68	0.71	0.76
45	0.62	0.69	0.72	0.77

From the graph, the values are tabulated in the table 3. It shows that the proposed system is shown higher recall values compare than previous algorithm. Thus the result concludes that the proposed system is used to provide efficient classification results.

5. Conclusion

In this section, the conclusion decides that the proposed system is increased the classification performance using modified artificial bee colony algorithm. The various time series data is implemented and the methods are focused on the classification of more accurate results. The existing RF, SVM, MMSVM and IPSO algorithms are used to handle the multiple classification time series data and also dealt with the unbalanced dataset. The proposed MABCA with TSVM and ANN algorithm is used to improve the classification performance and reduce the time complexity issues significantly by using global optimal features. Thus, the experimental result proved that the proposed system is better than the existing system.

6. Acknowledgement

We the authors assure you that, this is our own work and also assure you there is no conflict of interest.

References

1. *Anwar N, Huntz E, Kolch W, Pitti A (2010),'Semantic Data Integration for Francisellatularensisnovicida Proteomic and Genomic Data'; Vol. 21, No.5, pp.321-345.*
2. *Batal I, Valizadegan H, Cooper G F, and Hauskrecht M (2011), 'A pattern mining approach for classifying multivariate temporal data,' in Proc. IEEE Int. Conf. Bioinformat. Biomed., Vol. 20, No.3, pp. 358–365.*
3. *Breiman L, 'Random forests' (2001), Mach. Learning, Vol. 45, No. 1, pp. 5–32.*
4. *Damian Bargiel and Herrmann S (2011), 'Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight terraSAR-X data'., Vol.3, No.5, pp. 859–877.*
5. *Campos M, Palma J, and Mar´ın R (2007), 'Temporal data mining with temporal constraints artificial intelligence in medicine', Vol.2, No.4, pp. 67–76.*
6. *Dowdy S, Wearden S, and Chilko D (2004), 'Statistics for Research', 3rd ed. New York, NY, USA: Wiley, Vol.42, No.6, pp. 625-640.*

7. *Exarchos T, Tsipouras M, Papaloukas C, and Fotiadis D (2009), 'An optimized sequential pattern matching methodology for sequence classification,' Vol. 19, No. 2, pp. 249–264.*
8. *Gardner S P (2005) 'Ontologies and Semantic Data Integration', Drug Discovery Today, Vol.10, No.14, pp.1001-1007.*
9. *Gupta S, Kumar D, Sharma A (2011), 'Data Mining Classification Techniques Applied For Breast Cancer Diagnosis And Prognosis', Vol.11, No.7, pp.125-135.*
10. *Han J and Kamber M (2006), 'Data Mining: Concepts and Techniques', 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, Vol.20, No.12, pp.325-365.*
11. *Hayder m, Albehadili, Abdurrahman and E. Islam, An algorithm for time series prediction using particle swarm optimization (PSO), In IJSK, vol.4, 2014.*
12. *Joseph Sexton, Urban D L, Donohue M J, and Song C (2005), 'Long-term land cover dynamics by multi-temporal classification across the Landsat- 5 record,' Remote Sens. Environ., Vol. 128, No.4, pp. 246–258.*
13. *Lee C H, Chen J C, and Tseng V S (2011), 'A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring,' Comput.Methods Programs Biomed., Vol. 101, No. 1, pp. 44– 61.*
14. *Lee M, Lu H, Ling T, and Ko Y(1999), 'Cleansing data for mining and warehousing,' in Database and Expert Systems Applications, Vol.46, No. 6, pp.750-807.*
15. *Nhat-Duc Hoang, Anh-Duc Pham and Minh-TuCao,'A Novel Time Series Prediction Approach Based on a Hybridization of Least Squares Support Vector Regression and Swarm Intelligence',In Applied Computational Intelligence and Soft Computing 2014 (2014): 15.*
16. *Sarkar, Chandrima, Sarah Cooley, and JaideepSrivastava, Improved feature selection for hematopoietic cell transplantation outcome prediction using rank aggregation, Computer Science and Information Systems (FedCSIS), 2012 Federated Conference on, IEEE, 2012.*
17. *Soni J, Ansari D,Sharma S (2011),'Predictive Data Mining for Medical Diagnosis:An Overview of Heart Disease Prediction', Vol.47, No.2, pp.640-650.*
18. *Stacey M and McGregor C (2007), 'Temporal abstraction in intelligent clinical data analysis: A survey,' Artif. Intell. Med., Vol. 39, No. 1, pp. 1–24.*
19. *Tseng, Yuh-Min, et.al (2014), 'Multiple Time Series Clinical Data Processing for Classification with Merging Algorithm and Statistical Measures', Vol.56, No.4, pp.5-35.*
20. *Yin Z and Zhang J (2014), 'Identification of temporal variations in mental workload using locally-linear-embedding-based EEG feature reduction and classification techniques,' Comput. Methods Programs Biomed., Vol.115, No.3, pp. 119–134.*