International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2017 IJSRCSEIT | Volume 2 | Issue 4 | ISSN : 2456-3307

Multiple-Time-Series Clinical Data Processing for Classification Using Merging Algorithm

Kokila Ikhar*[1], Prof. Gurudev B. Sawarkar*[2]

*Department of Computer Science & Engineering
[1]M.Tech Student, V.M. Institute of Engineering & Technology, Nagpur, Madhya Pradesh, India
[2]Assistant Professor, V. M. Institute of Engineering & Technology, Nagpur, Madhya Pradesh, India

ABSTRACT

A depiction of patient conditions ought to comprise of the progressions in and mix of clinical measures. Customary data-preparing technique and classification calculations may make clinical data vanish and lessen forecast execution. To enhance the precision of clinical-result forecast by utilizing numerous estimations, another various time-arrangement data preparing calculation with period combining is proposed. Clinical data from 83 hepatocellular carcinoma (HCC) patients were utilized as a part of this exploration. Their clinical reports from a characterized period were combined utilizing the proposed blending calculation, and factual measures were likewise ascertained. After data handling, numerous estimations bolster vector machine (MMSVM) with outspread premise work (RBF) parts was utilized as a classification technique to foresee HCC repeat. A numerous estimations arbitrary backwoods relapse (MMRF) was likewise utilized as an extra assessment/classification method. To assess the data-combining calculation, the execution of forecast utilizing handled different estimations was contrasted with expectation utilizing single estimations. The aftereffects of repeat expectation by MMSVM with RBF utilizing different estimations and a time of 120 days (precision 0.771, adjusted exactness 0.603) were ideal, and their prevalence over the outcomes acquired utilizing single estimations was factually noteworthy (exactness 0.626, adjusted exactness 0.459, $P < 0.01$). In the instances of MMRF, the forecast comes about acquired in the wake of applying the proposed combining calculations were additionally superior to anything single measurement comes about ($P < 0.05$). The outcomes demonstrate that the execution of HCC-repeat forecast was fundamentally enhanced when the proposed data-handling calculation was utilized, and that various estimations could be of more noteworthy incentive than single.

Keywords: Data Mining, Data Processing, Multiple Measurements, Support Vector Machine (SVM), Time-Series Analysis.

I. INTRODUCTION

THE assortments of data qualities are one of the real issues of data preparing [1]. There are two sorts of data: time-arrangement data and cross-sectional data. Time-arrangement data are a succession of perceptions of a specific element, which are requested in time, while cross-segment data are gathered by watching many components in the meantime. In the main sort, the elements change after some time, and these progressions contain critical data. For instance, the time arrangement of blood glucose levels and pulse are considered as imperative wellbeing markers [2], [3]. In the second sort, various clinical data that are created in the meantime ought to be viewed as together to get a total photo of a patient's condition in a particular day and age. For example, in a normal wellbeing examination, the wellbeing status of people is portrayed by the aftereffects of a progression of research facility tests and physical examinations, for example, stature, weight, white platelet number, and red platelet tally [4], [5]. Outlining a data-preparing technique that can deal with crosssectional and time-arrangement data in the meantime would in this way appear to be basic for clinical data examination.

Utilizing data-preparing methods before data examination can significantly enhance the nature of the data, lessen the time required for the investigation, and enhance the nature of the investigation [6]. There are various data-preprocessing procedures, including data cleaning, data mix, data change, and data decrease. Data cleaning is the way toward identifying wrong records [7]–[9].

Data coordination consolidates data from various sources [10]. Data change alludes to the use of a deterministic scientific capacity to each point in the data, which may enhance the precision and proficiency of data mining [11]. Data diminishment totals or wipes out excess elements of the data, in this way lessening it to a more sensible size [12]. Changing data from a low-level quantitative frame to an abnormal state subjective depiction is known as transient reflection (TA) [13]. The procedure of TA takes either crude or pre-processed data as info and produces setting delicate and subjective interim-based portrayals. Because of patients' wellbeing data, subjective depiction is nearer to the dialect of clinicians [14]. Worldly classification of time-related clinical data has certain properties that recognize it from other classification strategies [15], [16], and utilizing the attributes of fleeting data could, in principle, enhance the execution of transient classification. Be that as it may, not all the pertinent data or perceptions are recorded in the meantime. For instance, unique research facility tests may have distinctive perception frequencies, because the requests for them may be begun by various divisions or doctors, and may be additionally influenced by the patient's status. Data with an unequal gathering recurrence ought to be prepared before examination. For joining elements of various sorts seen at various circumstances, a consolidating calculation for time-arrangement and multiple variables data is proposed.

The fundamental thought is to consolidation all components that happen inside a characterized day and age; if a specific element has more than one esteem, just a single of these qualities will be spoken to it. This consolidating calculation for various estimations is a strategy for data lessening; that is, data in the first data could be evacuated by the blending calculation. For saving the data, factual measures of the first data in a particular day and age are taken, and these can remain in for the inclination and the dissemination of the first data. After numerous estimations data handling, the data that incorporate chose components are coordinated from their different databases. In this review, the examination targets were patients who had hepatocellular carcinoma (HCC) and were being dealt with by radio recurrence removal (RFA). Their clinical reports that were gathered before treatment were utilized for assessing the data-handling strategy. After data preparing, the single estimation and various estimations data were characterized into two classes—repeat and non-repeat—to foresee patients' HCC result after RFA treatment. A correlation of the classification comes about acquired with single estimations against those gotten with numerous estimations could speak to the execution of the data-handling technique, and show that the strategy could enhance the adequacy of expectation of RFA-treated HCC repeat.

II. LITERATURE REVIEW

In [17], Chao-Hui Lee et.al (2011) proposed a novel data mining strategy to enhance the productivity and adequacy of patient observing. This strategy is utilized for perceiving assaults of endless illnesses through considering of both patients bio-signals and ecological variables. The example based choice tree and affiliation run mining components are utilized to coordinate consecutive example mining calculation to mine asthma assaults elements and constructs the classifiers adequately.

In [18], Themis P. Exarchos et.al (2009) displayed an upgraded successive example coordinating method for grouping classification. The strategy presented in this situation is utilized to concentrate naturally delivers a grouping order demonstrate, relies on upon the two stage strategies. However, this situation improves comes about and not got excessively mindfulness.

In [19], Damian Bargiel et.al (2011) proposed the multi fleeting classification of farming area utilize in light of high determination spotlight TerraSAR-X pictures. This situation opens a few chances to acquire learning about effects caused through rural land use and its varieties. In any case, this situation needs to enhance the classification consequences of single classes rather than class gatherings.

In [20], IyadBatal et.al (2012) recommended an example mining approach for ordering multivariate fleeting data. The joining of classification and example

extraction approach is presently attracted the data mining exploration and furthermore is effectively used in static data, chart data and succession data. In this exploration situation, the negligible prognostic fleeting examples approach and effective calculation is displayed and enhanced to remove these examples.

In [21], Mohamed F Ghalwash et.al (2012) examined about the early classification of multivariate fleeting perceptions utilizing extraction of interpretable shapelets. For the early arrangement assignment, we presented a procedure named as multivariate shapelets location (MSD). It mines the examples from all measurements on the time arrangement datasets. In any case it has issue alongside the running time multifaceted nature while consolidating parallelism in the calculation.

In [22], Zhong Yin et.al (2014) recommended recognizable proof of fleeting varieties in mental workload by utilizing a few data mining calculations. This situation is presented locally straight inserting (LLE) which is utilized for finding the low dimensional complex in the high dimensional complex EEG markers from different cortical areas. To recognize the mental workload (MWL) to discrete levels by utilizing MWL pointers and little measured preparing tests, another EEG approach by combining LLE, bolster vector grouping and bolster vector data depiction techniques are presented additionally evaluated with the assistance of measured data. Nevertheless, the unwavering quality of the situation is lessened fundamentally in this approach.

In [23], Chandrima Sarkaret,al exhibited enhanced element choice instrument to build up the classification precision. In this exploration situation, we utilize rank conglomeration based component choice strategy to pick proper benefactor genotype highlights. The proficient data mining methodology is utilized to choose the vital elements to recognize the ideal giver for patients. It handles the high dimensional dataset even more successfully.

In [24], Yi-Ju Tseng et.al examined numerous time arrangement clinical data handling for classification with blending calculation and factual measures. To advance the exactness of therapeutic result classification utilizing different estimations, a novel various time arrangement data preparing approach with combining calculation is improved. To look at the data combining approach, the classification execution by utilizing handled different estimations is contrasted with classification utilizing single estimations.

In [25], Hayder M et.al proposed molecule swarm enhancement (PSO) to enhance the time arrangement classification data exactness. The proposed PSO advancement calculation is centered on the lessening of number of emphases to achieve ideal arrangement. Gaussiam most extreme probability with PSO is utilized to diminish the mistakes essentially and increment the speed of the calculation.

In [26], Nhat-Duc Hoang et.al introduced a half and half approach called as bolster vector relapse with fake neural system calculation to manage time arrangement classification dataset. The simulated neural system calculation is a looking calculation, which is utilized to recognize the appropriate parameters for expanding the classification exhibitions. Consequently this strategy has been distinguished as the better procedure as the approach gives profound investigation of higher exactness expectation for the predefined datasets. The effectiveness of this approach is superior to anything-alternate systems and furthermore it gives pathway to further change.

III. METHODS AND MATERIAL

1. Pre-processing

In this module, the pre handling method is performed to acquire the more precise classification comes about. Data cleaning is the procedure of finding and amending off base records from the predetermined dataset. Utilized for the most part in databases, the term alludes to recognizing inadequate, erroneous, incorrect, unimportant, and so on data combination is the procedure of consolidate the different data from heterogeneous data sources yet with semantic importance. It is utilized to build the classification exactness comes about for the particular inquiry. Data change is utilized to change the arrangement of data esteems from the source data framework to goal data framework. By utilizing the pre-handling strategy, the exactness of classification execution is expanded as far as diminishment of clamor and missing esteems.

2. Feature selection

In this module, we need to play out the component choice process on the time arrangement dataset. It is utilized to give important element to the preparation and testing process. To evacuate the repetitive and unessential components, the element choice based arbitrary timberland is presented. A group classifier calculation is upgraded which contains stowing and irregular element choice strategies. The recurrence of an element's appearance in the classification trees speaks to the significance of the component. The library arbitrary woodland is used to execute the arbitrary woodland highlight determination handle. Every one of the elements is positioned by the weight relegated to them by irregular timberland.

3. Algorithm for Merging Multiple Features Based on Defined Time Periods

In this module, we need to consolidate the essential components by utilizing the blending calculation all the more effectively. In light of the calculation 1 we assessed the time arrangement data. The calculation is as per the following.

Algorithm 1

Start

Read D_m= m days period

T_{events}= the time of specific event

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

F_B= all features in R_B

Initialize merged records array based on m days period and

FOR each record in , k=1,2…N

T_k= the time of , recorded

i= T_{events} - T_k/ D_m

M_{mi}= the i^{th} merged record based on m days period

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

Set the value W_q of F_q in Mm i as the most recent value of F_q from all the RB k in RB and ith 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 axi= maximum of all the Ft in RB within period i

Ft_ M ini= 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

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

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

Add Ft M axi,Ft M ini, Ft A vg i , Ft SD i , Ft C ori, Ft S lpi as addition features into the ith merged record Mm i

ENDFOR

ENDFOR

OUTPUT Mm

END

The focal thought of this combining calculation is to pick just a single an incentive to remain for an element in one period. Since the season of the objective occasion, for example, treatment for HCC, is set as the key time concerning data handling, the esteem that is nearer to occasion time could be more critical than others could. Hence, the latest esteem is chosen to speak to an element in a period, and along these lines, some profitable data in the first data may be overlooked by the consolidating calculation.

4. Calculation of statistical measure

In this module, factual measure is computed for portraying the data dispersion in every period. There is a likelihood that data in the first data, for example, the inclination and highlight dispersion may vanish after data blending. To secure the data, greatest and least measurements are utilized as a part of this situation. Normal is a strategy for inferring the focal propensity of an element space, and standard deviation is a broadly used estimation of changeability. Pearson's relationship coefficient is appearing, how the element combine is unequivocally related inside the scope of - 1 to +1.

5. Prediction model establishment

In this module, the data mining methodologies are, for example, bolster vector machine (SVM) and irregular woods utilized for single and numerous estimations separately. The SVM assembles the classification show for a twofold class and it utilizes nonlinear mapping to change the data into higher dimensional data. Alongside an appropriate nonlinear mapping, two classes are partitioned through a hyperactive plane. The library SVM is engaged to execute the SVM forecast prepare. The portion work with spiral premise capacity is utilized for SVM display foundation. For different estimations, the expectation results are chosen through voting strategy where more elements had a place with comparable gathering and larger part vote of class is considered as definite forecast result.

Algorithm 2

BEGIN

Sm = 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

P_{Mm}= the predictive model established based on selected features in R_m , 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 S_m i of Pi in period i in S_m

R_m i = prediction result of P S_m i by using P_{Mm}

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

VR i = R_m 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 If the 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 using P_{Mm}

Predict Pi by VR i //regression result

ENDFOR

OUTPUT performance of P_{Mm} based on the prediction results

END

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

6. IPSO classification

The calculation proposed in this work depends on the molecule swarm streamlining system. PSO is a streamlining calculation that enhances a given arrangements through applying numerical tenets and subsequent to registering the wellness of a flow arrangements changes their directions into the pursuit space. Kennedy, Eberhard, and Shi initially present PSO as an enhancement method roused by the social conduct of feathered creature rushes and fish groups. PSO uses a specific number of arrangements, called particles that frame a swarm. Each such molecule has position and speed organizes in the inquiry space. The speed speaks to the change of the molecule position from cycle to emphasis. The change of the molecule's position is managed by the best so far referred to standard title's position too from the best position in the general swarm. This is utilized to enhance the speed of the procedure by utilizing imperative and important data includes in the dataset. It lessens the quantity of cycles by choosing the best answers for time arrangement dataset. The IPSO calculation is as per the following

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

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

End For

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, enhanced PSO approach is utilized for powerful outline handle. Molecule swarm enhancement (PSO) is a computational calculation that advances an issue by iteratively attempting to advance a hopeful arrangement alongside respect to a given measure of quality. The c1 and c2 are psychological parameters; r1 and r2 are arbitrary parameters. It is utilized to pick the best arrangements from the numerous time arrangement data. PSO improves an issue by having a populace of competitor arrangements, here named particles, and moving these particles around in the pursuit space as indicated by basic numerical recipe over the molecule's position and speed. Every molecule's development is affected by its neighbourhood best-known position but on the other hand, is guided toward the best-known positions in the pursuit pace, which are refreshed as better positions are found by different particles. This is relied upon to push the swarm toward the best arrangements.

In the proposed framework, we acquainted enhanced PSO calculation with increment the classification precision. For the given info datasets, the similitudes of numerous components are separated ideally by utilizing PSO parameters. The fundamental point of the PSO calculation is to choose the potential and important elements by producing best wellness work esteem. Likewise, it is successfully utilized for various times includes alongside a few elements. It requires least execution investment via seeking universally and furthermore it refreshes new best closeness esteems rapidly. Henceforth it builds the classification exactness higher for the given indicated datasets and best components are recovered by utilizing enhanced PSO calculation all the more precisely.

IV. RESULTS AND DISCUSSION

The different time-arrangement data-preparing calculation with period blending is a powerful strategy for data handling before classification. The consequences of HCC-repeat forecast in light of: 1) MMSVM with RBF bit and numerous estimations and a time of 120 days; and 2) MMRF with various estimations and factual measures and a time of 120 days were both fundamentally superior to with single estimation. Utilizing the various estimations with factual measures and MMSVM with RBF additionally yielded preferred classification comes about over utilizing the single estimation, however the midpoints of precision and BAC were lower than those of different estimations without measurable measures were. At the end of the day, after data handling, the classification precision and BAC from MMSVM with RBF piece had expanded 23.16% and 31.37%, individually, when contrasted with utilizing a solitary estimation. Consequently, it is recommended that the proposed combining calculation could enhance the expectation execution achievable utilizing clinical data all the more for the most part.

In the ideal MMSVM show gotten from external fivefold cross approval, the most widely recognized chose components were ALT and AST, which were chosen in four folds. The second most normal elements were platelet tally and HBV, which were each utilized

as a part of two folds. ALT and AST are oftentimes utilized for liver capacity testing in routine wellbeing examination. Besides, patients with liver illness as a rule have a diminished platelet check (thrombocytopenia) since platelet generation is managed by throm bopoietin, a hormone delivered in the kidneys and liver [27]. The nearness of the hepatitis B e antigen is related with an expanded danger of HCC [28], [29], and HBV viral load is related with HCC repeat [30]–[32]. These four clinical components are all very identified with liver capacity and HCC, and it is sensible to expect that they may be helpful indicators of HCC repeat. In the ideal MMRF display, the most widely recognized chose highlight was AST, which were chosen in two folds.

The ideal data-handling period was 120 days in the models worked by MMSVM and MMRF, and it fitted the data qualities depicted in Supplementary Data 3. Since the normal number of reports per understanding in the 180 days before RFA treatment extended from 1.72 to 3.07, data-preparing periods that were shorter than 60 days may build the rates of missing esteems. The aftereffects of this paper demonstrate that consolidating various time-arrangement data in a characterized era could change the related time-arrangement data into valuable data. Examining different time-arrangement data independently may disregard the general circumstance in a particular day and age, and consequently diminish the unwavering quality of investigation results.

Our data-handling calculation for numerous estimations gives a general strategy to creating profitable data for classification. It can be utilized with clinical data, as well as with any sort of data that counts with the objectives of the multiple time-arrangement data-preparing calculation, for example, depictions of the state of water and air, meteorological data, and money related data. Through this calculation, incorporated, time-subordinate, and valuable data can be made.

In spite of the fact that our outcomes propose that the different estimations data preparing calculation was useful in HCC-repeat forecast, it was found to have a few impediments. In the first place, despite the fact that the ideal exactness and BAC of classification were gotten from the model utilizing 120 days as a consolidating period, there was no proof that a 120-day time frame would be as valuable when managing other data sets. Additionally, the characterized periods couldn't be set naturally in light of the qualities of a dataset, and it may along these lines be conceivable that we missed the ideal day and age in light of the fact that it was not one of the six time frames we had inspected; at the end of the day, if a day and age of 83 days or 114 days is in actuality superior to a time of 120 days, our trial method would not uncover this. Third, in spite of the fact that the affectability and positive prescient estimation of HCC-repeat expectation utilizing single estimation were expanded by utilizing our proposed calculation; the affectability and PPV were still low. The lopsided data and little research populace may be the explanations behind this. Also, we examined the components by univariate examination (data not appeared), and just tumor size and HBV had noteworthy outcomes. In light of this, one might say that HCC-repeat forecast utilizing research facility reports from before the start of treatment is difficult, and could be enhanced through utilization of the different time-arrangement data-preparing calculation.

V. CONCLUSION

This research introduces a blending calculation for various time series data with various examining rates and data sorts, and assesses the impacts of adding factual measures to it. The outcomes show that the execution of HCC-repeat expectation was fundamentally enhanced through utilization of the calculation, and, as an end product, that numerous estimations could give more helpful data to HCC-repeat forecast than single estimation does.

VI.REFERENCES

[1] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 2nd ed. San Mateo, CA, USA: Morgan Kaufmann, 2005.

[2] J. Blackburn, S. Brumby, S. Willder, and R. McKnight, “Intervening to improve health indicators among Australian farm families,” J. Agromed., vol. 14, no. 3, pp. 345–356, 2009.

[3] E. Sobngwi, J.-C. Mbanya, N. C. Unwin, R. Porcher, A.-P. Kengne, L. Fezeu, E. M. Minkoulou, C. Tournoux, J.-F. Gautier, and T. J. Aspray, “Exposure over the life course to an urban environment and its relation with obesity, diabetes, and hypertension in rural and urban Cameroon,” Int. J. Epidemiol., vol. 33, no. 4, pp. 769–776, 2004.

[4] C. Scheidt-Nave, P. Kamtsiuris, A. G¨oßwald, H. H¨olling, M. Lange, M. A. Busch, S. Dahm, R. D¨olle, U.

Ellert, and J. Fuchs, "German health interview and examination survey for adults (DEGS)— design, objectives and implementation of the first data collection wave," BMC Public Health, vol. 12, art. no. 730, 2012. [Online]. Available http://www.biomedcentral.com/1471-2458/12/730

[5] S. Zhi, W. Sheng, and S. P. Levine, "National occupational health service policies and programs for workers in small-scale industries in China," Amer. Ind. Hygiene Assoc., vol. 61, no. 6, pp. 842–849, 2000.

[6] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2006

[7] M.A.Hern´andez and S. J. Stolfo, "Real-world data is dirty: Data cleansing and the merge/purge problem," Data Mining Knowl. Discovery, vol. 2, no. 1, pp. 9–37, 1998.

[8] M. Lee, H. Lu, T. Ling, and Y. Ko, "Cleansing data for mining and warehousing," in Database and Expert Systems Applications, T. Bench- Capon, G. Soda, and A. Tjoa, Eds. Berlin, Germany: Springer, 1999, p. 807.

[9] W. Lup Low, M. Li Lee, and T. Wang Ling, "A knowledge-based approach for duplicate elimination in data cleaning," Inf. Syst., vol. 26, no. 8, pp. 585–606, 2001.

[10] M. Lenzerini, "Data integration: A theoretical perspective," in Proc. 21st ACMSIGMOD-SIGACT-SIGART Symp. Principles Database Syst.,Madison, WI, USA, 2002, pp. 233–246.

[11] A. A. Hancock, E. N. Bush, D. Stanisic, J. J. Kyncl, and C. T. Lin, "Data normalization before statistical analysis: Keeping the horse before the cart," Trends Pharmacol. Sci., vol. 9, no. 1, pp. 29–32, 1988.

[12] A. S. C. Ehrenberg, Data Reduction: Analysing and Interpreting Statistical Data. New York, NY, USA: Wiley, 1975.

[13] M. Stacey and C. McGregor, "Temporal abstraction in intelligent clinical data analysis: A survey," Artif. Intell. Med., vol. 39, no. 1, pp. 1–24, 2007.

[14] M. Stacey, C. McGregor, and M. Tracy, "An architecture for multidimensional temporal abstraction and its application to support neonatal intensive care," in Conf. Proc. IEEE Eng. Med. Biol. Soc., 2007, pp. 3752–3756.

[15] M. Campos, J. Palma, and R. Mar´ın, "Temporal data mining with temporal constraints artificial intelligence in medicine," in Artificial Intelligence inMedicine, R. Bellazzi, A. Abu-Hanna, and J. Hunter, Eds. Berlin, Germany: Springer, 2007, pp. 67–76.

[16] A. Juan Carlos, "Temporal reasoning for decision support in medicine," Artif. Intell. Med., vol. 33, no. 1, pp. 1–24, 2005.

[17] Breiman L, 'Random forests' (2001), Mach. Learning, Vol. 45, No. 1, pp. 5–32.

[18] 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.

[19] 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.

[20] 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.

[21] 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.

[22] Gardner S P (2005) 'Ontologies and Semantic Data Integration', Drug Discovery Today, Vol.10, No.14, pp.1001-1007.

[23] 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.

[24] 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.

[25] Hayder m, Albehadili, Abdurrahman and E. Islam, An algorithm for time series prediction using particle swarm optimization (PSO), In IJSK, vol.4, 2014.

[26] 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.

[27] K. Kaushansky, "Thrombopoietin: The primary regulator of platelet production," Blood, vol. 86, no. 2, pp. 419–431, 1995.

[28] H.-I. Yang, S.-N. Lu, Y.-F. Liaw, S.-L. You, C.-A. Sun, L.-Y.Wang, C. K. Hsiao, P.-J. Chen, D.-S. Chen, and C.-J. Chen, "Hepatitis B e antigen and the risk of hepatocellular carcinoma," New Engl. J. Med., vol. 347, no. 3, pp. 168–174, 2002.

[29] A. M. Di Bisceglie, "Hepatitis B and hepatocellular carcinoma," Hepatology, vol. 49, no. S5, pp. S56–S60, 2009.

[30] M. Chuma, S. Hige, T.Kamiyama, T. Meguro, A. Nagasaka,K.Nakanishi, Y. Yamamoto, M. Nakanishi, T. Kohara, T. Sho, K. Yamamoto, H. Horimoto, T. Kobayashi, H. Yokoo, M. Matsushita, S. Todo, and M. Asaka, "The influence of hepatitisBDNAlevel and antiviral therapy on recurrence after initial curative treatment in patients with hepatocellular carcinoma," J. Gastroenterol., vol. 44, no. 9, pp. 991–999, 2009.

[31] J.-C. Wu, Y.-H. Huang, G.-Y. Chau, C.-W. Su, C.-R. Lai, P.-C. Lee, T.-I. Huo, I. J. Sheen, S.-D. Lee, andW.-Y. Lui, "Risk factors for early and late recurrence in hepatitis B-related hepatocellular carcinoma," J. Hepatol., vol. 51, no. 5, pp. 890–897, 2009.

[32] I. F. N. Hung, R. T. P. Poon, C.-L. Lai, J. Fung, S.-T. Fan, andM.-F. Yuen, "Recurrence of hepatitis b-related hepatocellular carcinoma is associated with high viral load at the time of resection," Amer. J. Gastroenterol., vol. 103, no. 7, pp. 1663–1673, 2007.