IOSR Journal of Engineering (IOSRJEN) **www.iosrjen.org**
ISSN (e): 2250-3021, ISSN (p): 2278-8719
Vol. 08, Issue 12 (December. 2018), ||V (II) || PP 04-11

A Machine Learning Approach to Effective Patient Flow Prediction Using IMCP and DCM Framework

Basmath.P[1], Ramesh Kumar.B[2]
M.Phil Scholar, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu, India[1]
Assistant Professor, Department of Computer Science, Sree Narayana Guru College, Coimbatore, Tamil Nadu India[2]
Corresponding Author: Basmath.P

Abstract: Over the past decade due to increasing the population Overcrowding in emergency care unit (CU) departments is a problem in many countries. This overcrowding which leads reduces the quality of care and increases medical errors. Generally predicting the transition processes of patients (PFP) is more difficult and crucial process. The field of data mining has involved in those domains to predict Patient Flow effectively. The previous studies have used several features to predict Patient Flow, which has been collected from patient's hospital data. The main drawbacks of the previous studies are that need accurate and more number of features to effectively predict the patient flow and this system can only apply in discrete time series.

The system implements a new framework Improvised mutually correcting process (IMCP) along with Discriminative Conditional Model (DCM) model to predicting and managing patient flow with more accurately. Our process model is able to describe time-varying transition processes in continuous time. Random over sampling process has been applied for overcoming the data imbalance problem in patient flow data. DCM has been applied for improving the prediction of transition events in the case of sparse data. From the experimental results, patient flow prediction accuracy of our proposed algorithm reaches 97.8% with a convergence speed which is faster than the existing system.

Keywords: Patient Flow, Machine Learning, DCM.

Date of Submission: 29-11-2018 Date of acceptance: 13-12-2018

I. INTRODUCTION

In the medical worlds, information is increasing every day. The number of peoples can have used medical services through some technologies like ***e-medicine***. During this field cannot use technology, so it meets more issues such as less data security, memory consumption problems, risky updating, and time-consuming problems. But in the technology improving the world is to use some technical terms and improving the field. The growing term of data mining is to can improve the medical data so easily maintain the ***e-care data***, securely. The data mining is including more effective terminologies. The medical data is very important data. In which includes data types are hospital data, doctors data, patient data, medical data, staffs data, scan reports, medical reports, service reports, scheming reports, health reports, awareness reports, disease reports, these are more and lots of reports generated every day. So in this developed field is holds some issues, but in this proposed system concept is to focus only the term, "Patient flow". The patient flow means patient related all information is included. ***Flows*** means continue processes but these tasks are finished its give the effective result, also. Then the patient flow is available in each hospital ***Care Unit (CU)***. In this term patient flow organize the details are improves the patient outcomes, allocating critical care resources and reducing the preventable re-admissions. The records of a patient are submitted and added in the ***Electronic Health Records (EHRs)***. If to maintain these records effectively follows the special data mining technique namely, machine learning. It also gave the best results and a clear way of problem solve terms. So the trend and optimal records to be maintained in the patient flow use some correct terms. Then the developing novel framework is modeling this patient flow it by using various care units and joined it, then predicting the patient destination care units. Increased, data is followed the ***generative point process model***. It is via achieves maximum likelihood estimation. The ***discriminative learning algorithms*** main aim is to improve the prediction of transition events in spare data. So it is using the parameter zing model that is a ***mutually-correcting process***. The problem term is to estimates by using the generalized linear model namely, "***Alternating Direction Method of Multipliers (ADMM)***". In this model is efficient learning basis. The combination of this concept achieves the features like simultaneous

feature selection, then learning after adding a ***group-lasso regularize*** to the efficient based algorithm. It controls the negative influence of imbalanced data by using a learning model. It takes the sample from the training data and then improving the robustness follows learning methods.

II. PROBLEM DEFINITION

Generally, the emergency department is classified into several divisions but admits all category patients, also. The ED is a basic treatment section. If the ED patient is discharged is based on only patient condition. In that department is first gives the first aid. Then after giving some another treatment. The emergency department is classified into ED care needed and primary care treatable. But the system mainly focuses the ED care needed section it only provides two steps, (a) preventable/avoidable, (b) not preventable/avoidable.

In this paper [1] Dr. Kasthuri, Sreejith has presented the concept namely, "***Prediction and Classification of Patient Data Using Mutually-Correcting Nonparametric Model***". In the system of patient care is widely increased which the reason for the large population and huge health records. These health records and the health-related details are maintained properly by using some technologies, like data mining. These records can be accessed fastly and quickly for data analysis and retrieval. The concept is to apply the machine learning techniques because it is very effective to predict and manages the patient flow. The technique is performed ***(i) feature selection, (ii) prediction, (iii) estimation via a non-linear model***. It enhanced the ADMM by considering the class imbalance issues. In the concept introduces the technique namely, "***Non-Parametric Self Sampling (NPSS) technique***". This technique includes three types of algorithm are, OSCI-Over Sampling Class Imbalance with Non-Parametric model and the autoregressive hidden Markov model. The data mining mines the patient records and then predicts the patient flow in various care units. It handles large, dynamic and patient sequential data with the high prediction accuracy. Using the autoregressive technique with a hidden Markov model and performs strong and hidden feature selection, then performs the learning. The class imbalance on the learning and prediction model is mixing the training data and regenerates the dataset. So the data mining techniques and methodologies are to predict the data with accuracy for the critical care unit.

Use Auto-regressive model with a hidden Markov model and improve and develops the non-parametric model. Then the sampling technique is to solve the class imbalance issues and it to improve the accurate prediction. It handles the event sequences by adopting new methods and techniques for to handles the temporal dynamic events with the high dimensional view. The combination of point process model (using discriminative learning-based algorithms) and autoregressive and hidden markov model is called new adopting techniques. These techniques are solving the feature selection problem.

In this paper [2] Díez-Pastor, José F., Juan has presents "Identifying Overcrowding Indicators by using Patient Pathway Workflow Model in Emergency Department". The random balance technique is combines or ensembles the approaches included with imbalanced data that handles the existing classifiers. It grouping the features of various different classifiers and then performs the random selection. The ensemble method is to combine the RB-boost from the previous Adaboost algorithm. So far this technique is randomly choosing the data proportion for making the new one, although it follows the combining and generating. It has collected many datasets and then performs the imbalance data classification. But the main drawback of the RB-Boost technique is not considered the intrinsic characters and noise elimination process. So this limitation of the reason is inside the RB-Boost, therefore cannot maintain the dynamic electronic patient health records.

In this paper [3] Faten Ajmi, Sarah Ben Othman have recently presented the concept namely, "*Patient Pathway Workflow Model Identifying Overcrowding Indicators in the Emergency Department*". The main concept is focused on a patient pathway in an Adult Emergency Departments (AED). The data mining technique helps solves the overcrowding problem in the ED by using the ***workflow approach*** and ***BPMN tool***. BPMN means, ***Business Process Model and Notation***. It is one of the graphics tools and it which allows to verifies and processes the validation. The approach of workflow includes the graphical model then translates the patient flow into services. The BPMN has includes four types of categories are, workflow, organization, readability and specifies behaviors. The workflow model with the '***Bonita soft***' tool is to helps better which means it integrates the engine, automatic generation of the interface, the model of simulation is put in the place and connecting the possibility by the connector. The patient pathway is modeling by using and computerized system using the tool.

In the concept [4] Maratea, Antonio, Alfredo Petrosino, and Mario Manzo has presents "Imbalanced data learning by using Adjusted F-measure and kernel scaling**"** is follows the classification procedure based on Support Vector Machine (SVM). These data mining technique solves the class imbalance problem by using and then adjusting the f-measure and kernel scaling. It reduces the issue of misclassification. But it does not consider the sequential datasets and cannot handle the multiple health features.

In this paper [5] Wang, Yichen, Robert Chen, Joydeep Ghosh, Joshua C. Denny, Abel Kho, You Chen, Bradley A has presents "Health data analytics is using the Knowledge guided tensor factorization and completion: Rubik". In ***Alternative Direction Method of Multipliers (ADMM)*** is an algorithm. It solves the convex optimization problems. In the ADMM is uses the ***Discriminative Learning of Mutually-Correcting (DLMC) algorithm*** with some parametric approaches. In this algorithm is reducing the problem by segmenting into small chunks and it makes the easiest way to the process. These approaches provide the detailed analysis and mine the electronic health records with dynamic flow records.

In this paper [6] Xu, Hongteng, Weichang Wu has presents "Prophecy of Patient flow passing through discriminative learning of mutually-correcting processes." The main concept of patient flow detection is collected and predicts the data from ***Electronic Health Records (EHRs)***. In these records to predicted by using several different types of data mining techniques, but additionally introduced the ***NPSS***. Then, it handles the patient health records and predicts these records and predicts the resources. It has used the ***maximum likelihood estimation process via a discriminative learning algorithm***. In the combinational term is gives the efficient learning method for effective EHRs. It contains the high accuracy in the given data set samples and it holds a certain number of parameters. But not properly performed the time duration factor.

III. PROPOSED SYSTEM

The chapter completely discusses about the proposed system methodology and the steps by step process involved in that proposed system. The system successfully proposes a new novel framework (IMCP) Improvised mutually correcting process along with Discriminative Conditional Model (DCM) for Prediction of Patient Flow effectively.

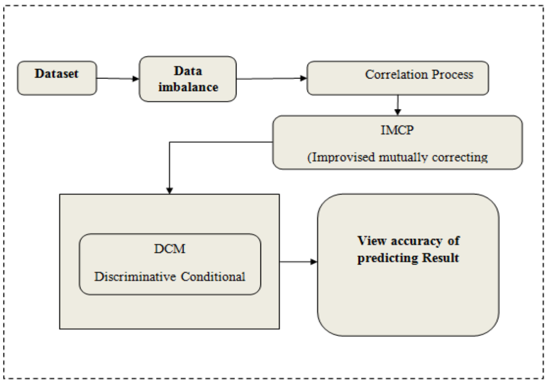

Fig 1.0 Proposed system architecture

3.1Contribution of the Proposed Work

The followings are the contributions of the proposed system.

- To obtain effective patient flow prediction in many hospital environments is critical and complication process which leads reduces the quality of care and increases medical errors and most of the existing research failed to obtain effective flow prediction. So proposed novel framework (IMCP) along with DCM model to predicting and managing patient flow with more accurately.
- Our process model is able to describe time-varying transition processes in continuous time.
- A ROS (Random over Sampling) process has been applied for overcoming the data imbalance problem in patient flow data and it reducing a number of observations from majority class to make the data set balanced. This ROS method is best to use when the data set is huge data and reducing the number of training samples.
- IMCP model has been used for to find and predict what types of CUs they need and how long will they stay at different CUs. This to capture the positive and negative influences among unit types and durations, respectively.
- DCM has been applied for improving the prediction of transition events in the case of sparse data.

IV. RESEARCH METHODOLOGY

The proposed research work concerned with 4 major parts is preprocessing, ROS, IMCP and DCM. The first process is the patient data collection from websites and uploads in to our proposed model. Each step is discussed in the following chapters.

4.1 Data upload and preprocessing

This module is the first level of the process, where the users are going to upload the data in to database server. Next step is Preprocessing this is the process of elimination, which eliminates duplicate and incomplete and redundant data from the dataset before processing in to the proposed model. It does not include redundant records in the train set as well, so the classifiers will not be biased towards more frequent records.

4.2 Random over Sampling (ROS)

Random oversampling is a simple and easy yet effective approach to data set resembling. This technique has been applied for overcoming the data imbalance problem. It reduces the number of observations from majority class to make the data set balanced. First, initially select data set documents randomly from the original training data set. Second, one always randomly oversamples with the replacement. This ROS method is best to use when the data set is huge data and reducing the number of training samples. This Random over Sampling implementation steps are list out below.

ROS Algorithm steps
Input:
X: the original training data set {x1, x2… xN}
S: majority class or negative class
S+: minority class or positive class
l: the reduced dimension
Output: S: the oversampled training set
Step 1: Start the process
Step 2: W=0 then // Initialize weight matrix W
Step 3: For each vector x in X
Step 4: J ← ----- is 1:0;
Step 5: Y←----- is 0
Step 6: X SNN s Find x's k number of nearest neighbors according to distance.
Step 7: For each v ( ) (X kNN(x) S
Step 8: κ compute linear spanning weights of T.
Step9: For each v ( ) (X kNN(x) S
Step 10: κ compute linear spanning weights of T.
End
Step11: XkNN (v) κ XkNN (v) {v}
Step 12: Y κ Compute the embedding data according to W
End

4.3 Improvised mutually correcting process

This model proposed to capture the properties of patient flow. This is more flexible, which can capture both the increase and decrease of intensity function between adjacent events and predicts the Probabilities. To describe the point process model, time-varying transition processes in continuous time.

IMCP Pseudo code
Step 1: load the dataset
Step 2: In order to map predicted values to probabilities use the sigmoid function.
S (z1) = frac {1} {1 + e1^ {-z1}}
S (z1) = output between 0 and 1 (probability estimate values).
e = base of natural log.
Step 3: Initiate Decision boundary process
Begin
p> threshold, class=1
p < threshold, class=0
End
Step 4: process gradient ascent to maximize the log.
S'(z)&= s(z)(1 – s(z))
Predictions = predict (features, weights)
Gradient /= N //derivative for each feature
Return;
Step 5: Mapping probabilities to class labels.
Return decision boundary (predictions).
Finally assign the class label to these predicted probabilities.
Step 6: End of the process.

4.4 Discriminative Conditional Model

DCM has been applied for improving the prediction of transition events in the case of sparse data. This DCM aim to estimate the joint probability of all sequence events via a maximum probability estimator.

DCM Pseudo code

Step 1: Scan the dataset
Step 2: Initialize randomly, X (0) = (0), Y (0) = 0 //Assign the values
Outer iteration number k = 0 //assign k values
Step 3: Repeat process
Inner iteration number l = 0, M (k, l) = M (k).
Step 4: Update process using this equation
Min L (k+1) + k1; 2;
l = L + 1.
Step 5: Update X (k+1) And Y (k+1)
k = k + 1.
Step 6: predicts of Cui and Dui
Cui= arg maxc2Cp (cjtuiHuti-1);
Dui= arg maxD2Dp (tuiHuti-1);
Step 7: Until
X (k) + Y (k) 2k
Update Y: Y (k+1) = Y (k)+ ((k+1) X(k+1)):
Step 8: End of the process.

V. RESULTS AND ANALYSIS

The performance of this proposed research work model Scheme was compared and effectively evaluated with the existing Model based on the some important following parameters such as Execution Time, accuracy and Time taken for patient flow prediction.

5.1 Data Sets

A typical patient electronic health data set consists of a patient's profile such as age gender and name along with certain diseases code and treatment process this include medications, the transitions, nursing information, and durations in various care units. Initially these patient electronic health data collect from web resource and store in server for processing this data. The system used a dynamic dataset, which can be any number of electronic health data. Data collection is the first step this dataset will be preprocessed before starting the implementation. Preprocessing steps eliminates the duplicate and missing items in the dataset.

Name	Age	Gender	Add Date	Disease	general ward	Medical ICU
Leo	45	male	15/11/2018	Coronary heart disease	Yes	Yes
Steepen	34	male	16/11/2018	Heart surgery	Yes	Yes
Wasim	34	male	13/11/2018	Fever/Illness	Yes	No
Malar	34	Female	15/11/2018	Caesarean section surgery	Yes	Yes
John	45	male	16/11/2018	Heart surgery	Yes	Yes
jasmine	52	female	14/11/2018	Fever/Illness	Yes	No

Table 5.1 Dataset Sample For our experiment

This dynamic data set has contains 8 attributes such as Name, Age, Gender, Ad Date, Disease, Anesthesia Services, general ward and Medical ICU. There are a total more than 100 patient records in the database.

5.2 Experimental Results

In this section this present the details of the experiment conducted in order to test the proposed technique. More intricate the system is being implemented, the more involved will be the systems analysis and design effort required just for accomplishment. The implementation phase comprises of several activities. Carried outs the required hardware and software acquisition. System is requires some software's to be developed. The research application is implemented in java as front end MySQL is as back end. This proposed work was implemented and the performance of this proposed work is compared with two existing approaches methods .The figure below shows the configuration of the system requirement for take a Experimental Results.

Software Requirements	
Front End	Java
Code Behind	Swing
Back End	MySQL SERVER
Operating System	Windows XP.

Hardware Requirements	
Hard disk	500 GB
RAM	2GB
Processor	Core2Dual
Monitor	17” Color monitor
Key board, Mouse	Multi media.

Table 5.2 Comparison table

Performance Analysis
The performance of the proposed methods is compared in general data set evaluation presents the performance on the first data sets, which exhibit a variety of properties and have been used in previous work by others.
Execution Time
This is a process for execute complete data set to complete entire process.
Accuracy
The system finally performs the analysis to show the accuracy of the proposed system. Accuracy refers to the proportion of data classified an accurate type in total data, namely the situation TP and TN, thus the accuracy is.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

Time taken
Total of time taken for the extract effective and important features from patient health care dataset.The figure below shows the results and comparison of the proposed system.

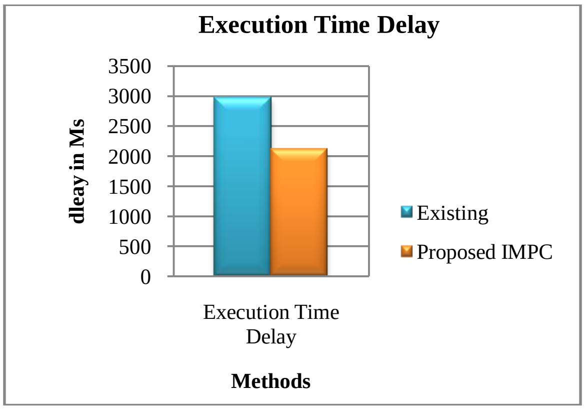

Table 5.3 Comparison table
Performance comparison of proposed IMPC with existing approaches based on Execution Time Delay.

Type	Mutually-Correcting Process	Proposed IMPC
Execution Time	2984.06	2108.08
Result Accuracy	90	97
Time taken	70.68	49.69

From the chart it shows the performance measure based on the Execution Time delay and the proposed approach IMPC took less time while comparing the other methods and the worst time complexity is Existing Method.
Performance comparison of research proposed IMPC with existing mutually-correcting process approaches based on Result patient flow prediction accuracy

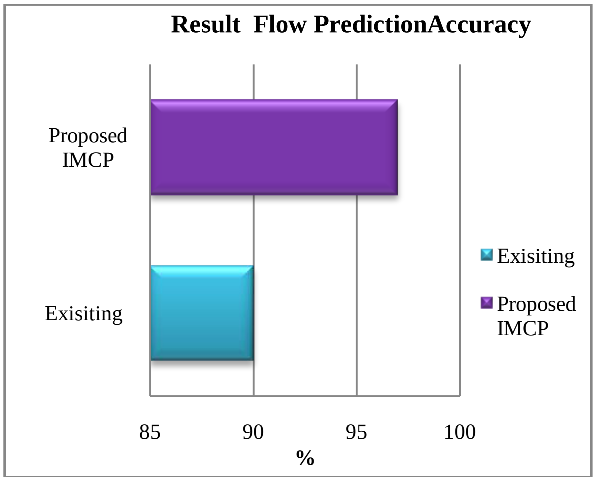

Table 5.4 Comparison table

From The graph chart it shows the performance measure based on the accuracy of patient flow detected proposed IMCP approach took more prediction accuracy while comparing the other methods and the worst based on the accuracy is existing method .

Time comparison between existing model and proposed Model

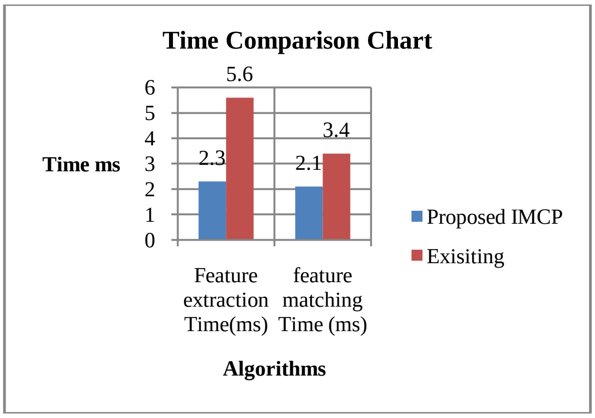

Fig: 5.5 Time comparison between existing and proposed Model

The above figure represents the comparison between existing and proposed system based on the feature extraction Training time.

VI. CONCLUSION

The main purpose of the study was to investigate the major problem of overcrowding which means predicting the transition processes of patients in hospital. . In the developed system is rapid, accurately and effectively handles the patient flow prediction. The mutually-correcting processes improve the flexibility and it considers it is one of the properties of the patient flow. Generally, the mutually means it is one of combinational and combined action that is called the mutual action. The combinational behavior of discriminative learning algorithm includes the multinomial logistic regression with group-lasso and the last one is achieves feature selection for the period of the learning model. The novel pre-processing method solves the data imbalance problem. It facilitates the modeling a patient needs that is care teams within the Care Units, then to improves the care management and the coordination for the patients with more constraints. The investigational results are evaluated. The experimental result shows that integrated proposed model shows better and effective Patient Flow Prediction assessment compared to existing techniques.

REFERENCES

[1]. Dr. Kasthuri, Sreejith. R. “Using Mutually-Correcting Nonparametric Model to Predict and Classify of Patient Data". JNCET, February 2018.

[2]. Díez-Pastor, José F., Juan J. Rodríguez, César García-Osorio, and Ludmila I. Kuncheva. "Imbalanced data using Ensembles of variable prior’s classifiers: Random balance." *Knowledge-Based Systems* 85 (2015):

[3]. Faten Ajmi, Sarah Ben Othman, Hayfa Zgaya Biau, Slim Hammadi. “Identifying Overcrowding Indicators by using Patient Pathway Workflow Model in Emergency Department”. ICCS (2): 540-552, Jun 2018.
[4]. Maratea, Antonio, Alfredo Petrosino, and Mario Manzo. "Imbalanced data learning by using Adjusted F-measure and kernel scaling" *Information Sciences* 257: 331-341, (2014).
[5]. Wang, Yichen, Robert Chen, Joydeep Ghosh, Joshua C. Denny, Abel Kho, You Chen, Bradley A. Malin, and Jimeng Sun. "Health data analytics is using the Knowledge guided tensor factorization and completion: Rubik." In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1265-1274. ACM, 2015.
[6]. Xu, Hongteng, Weichang Wu, Shamim Nemati, and Hongyuan Zha. "Prophecy of Patient flow passing through discriminative learning of mutually-correcting processes." *IEEE Transactions on Knowledge and Data Engineering* 29, no. 1: 157-171, (2017).

Basmath.P. " A Machine Learning Approach to Effective Patient Flow Prediction Using IMCP and DCM Framework."IOSR Journal of Engineering (IOSRJEN), vol. 08, no. 12, 2018, pp. 04-11.