Review Article **ISSN: 2395 – 5597**

International Journal of Engineering and Robot Technology

Journal home page: www.ijerobot.com

A NOVEL APPROACH FOR PREDICTING HEALTH RISK ASSESSMENT BASED ON PATIENT LOG

Sadhana Mani*[1]

[1]*Department of Computer Science and Engineering, Anna University, Chennai, Tamil Nadu, India.

ABSTRACT

Distinguishing and overseeing patients most at hazard inside the social insurance framework is fundamental for governments, doctor's facilities, and wellbeing safety net providers yet they utilize diverse measurements for recognizing the patients they see to be at generally chance. Wellbeing safety net providers are generally worried with protection chance, since they consent to repay wellbeing related administrations in return for a settled month to month premium. Poor hazard measure could bring about surpassing a money related spending plan. "90% of the world's information was created over the most recent couple of years." Due to the coming of new innovations, gadgets, and correspondence implies like interpersonal interaction locales, the measure of information delivered by humankind is developing quickly consistently. The measure of information delivered by us from the earliest starting point of time till 2003 was 5 billion gigabytes. In the event that you heap up the information as circles it might fill a whole football field. A similar sum was made in each two days in 2011, and in at regular intervals in 2013. This rate is as yet developing gigantically. In spite of the fact that this data created is important and can be valuable when handled, it is being dismissed. Enormous Data is a gathering of substantial datasets that can't be handled utilizing conventional figuring systems. It is not a solitary strategy or an instrument rather it includes numerous territories of business and innovation.

KEYWORDS

Health care, Big data, Patient, Insurance claim, Predicting days, Analyzing database.

Author for Correspondence:

Sadhana Mani,
Department of Computer Science and Engineering,
Anna University, Chennai, Tamil Nadu, India.

Email: Sadhusmarty@gmail.com

INTRODUCTION

Medicinal services managers worldwide are endeavoring to bring down the cost of care while enhancing the nature of care given. Hospitalization is the biggest part of wellbeing use. In this manner, prior recognizable proof of those at higher danger of being hospitalized would help medicinal services heads and wellbeing guarantors to grow better

arranges and procedures. In this paper, a strategy was created, utilizing substantial scale medical coverage claims information, to foresee the quantity of hospitalization days in a populace. We used a relapse choice tree calculation, alongside protection assert information from 242,075 people more than three years, to give expectations of number of days in healing facility in the third year, in light of clinic affirmations and system claims information from the initial two years. The proposed technique performs well in the all-inclusive community and in sub-populaces. Comes about demonstrate that the proposed show altogether enhances expectations more than two set up pattern techniques (anticipating a consistent number of days for every client and utilizing the quantity of days in healing center of the earlier year as the conjecture for the next year). A sensible prescient exactness (AUC = 0:843) was accomplished for the entire populace. Investigation of two sub-populaces - to be specific elderly people matured more than 63 years or more established in 2011 and patients hospitalized for no less than one day in the earlier year - uncovered that the restorative data made more commitment to expectations of these two sub-populaces, in contrast with the populace all in all.

EXSISTING TECHNIQUE

In 2009 and 2010, doctor's facilities included by a long shot the biggest part of well-being use in Australia, expending 40% of consistent well-being spending. Moreover, the Australian Productivity Commission, a free research and counseling body of the Australian Government, called attention to (in their give an account of tax payer supported organizations in 2013) that around AUD (Australian dollars) $3 billion were spent on superfluous open healing facility confirmations yearly. Prior recognizable proof of those at hazard would like wise help diminish pointless hospitalizations and conceivably spare citizens billions of dollars consistently. From different points of view, better expectation of hospitalizations will empower prior intercession, lessening costs and enhancing personal satisfaction.

PROPOSED TECHNIQUE

The point of this paper is to build up a model that predicts the aggregate number of days spent in doctor's facility amid a timetable year for people from an all-inclusive community, utilizing substantial scale medical coverage claims information. Since protection claims have solid financial attributes, their energy in anticipating clinical targets, for example, hospitalizations, are from time to time explored. Important late work has been performed by the creators in an information mining rivalry called the 'Legacy Health Prize (HHP)'. This opposition likewise meant to foresee the quantity of days in healing center for a timetable year. It concentrated on diminishing forecast blunder ascertained in a given condition by upgrading the prescient energy of different refined calculations. A last positioning of 40 out of 1300+ groups was accomplished was accomplished by creators, and - as far as the measure for the nature of forecast - the execution (0.467) with respect to the best scores (0.461) was little in total terms. For classification reasons, the informational index was firmly pseudonymized and improved with impressive points of interest missing.

In this paper, we display Healthcare overseers worldwide are endeavoring to bring down the cost of care while enhancing the nature of care given. Hospitalization is the biggest part of wellbeing use. In this paper, a technique was produced, utilizing extensive scale medical coverage Claims information, to anticipate the quantity of hospitalization days in a populace. The proposed technique performs well in the all-inclusive community and in addition in sub-populaces. Comes about demonstrate that the proposed show essentially enhances Predictions more than two set up gauge techniques which may altogether add to decrease the advancement cost of the Domain Modules. The portrays about the statistic insights of age and sexual orientation in year 2015.

METHODS

Patient registering details in Hospital

In this paper, we display Healthcare overseers worldwide are endeavoring to bring down the cost of

care while enhancing the nature of care given. Hospitalization is the biggest part of wellbeing use. In this paper, a technique was produced, utilizing extensive scale medical coverage Claims information, to anticipate the quantity of hospitalization days in a populace. The proposed technique performs well in the all-inclusive community and in addition in sub-populaces. Comes about demonstrate that the proposed show essentially enhances Predictions more than two set up gauge techniques which may altogether add to decrease the advancement cost of the Domain Modules. The portrays about the statistic insights of age and sexual orientation in year 2015.

Admin maintaining the records

This is the second module of our venture in this with the coming of web applications. The administrator just secure keeping up our capacity that is duty of each demand and reaction administrations. Big Data underpins incorporate boundless support and overhauls and is full overseen by administrator. At whatever point ask for get from customer in this module just reaction to them in our information based EHRs frame work keeping up various division subtle elements like surgery Nursing and so forth these are convey in parallel process on condition.

Data Migration Module with Sqoop

The customary application administration framework, that is, the association of uses with social database utilizing RDBMS, is one of the sources that produce Big Data. Such Big Data, produced by RDBMS, is put away in Relational Database Servers in the social database structure. At the point when Big Data stockpiles and analyzers, for example, Map Reduce, Hive, HBase, Cassandra, Pig, and so forth of the Hadoop biological system came into picture, they required a device to connect with the social database servers for bringing in and sending out the Big Data dwelling in them. Here, Sqoop possesses a place in the Hadoop biological system to give practical communication between social database server and Hadoop's HDFS. Sqoop: "SQL to Hadoop and Hadoop to SQL" Sqoop is a device intended to exchange information amongst Hadoop and social database servers. It is utilized to import information from social databases, for example, MySQL, Oracle to Hadoop HDFS, and send out from Hadoop record framework to social databases. It is given by the Apache Software Foundation. Presently we are prepared with dataset. So now our point is move the dataset into hadoop (HDFS), that will be occur in this module. Sqoop is an order line interface application for exchanging information between social databases and Hadoop. In this module we get the dataset into hadoop (HDFS) utilizing sqoop Tool. Utilizing sqoop we need to perform parcel of the capacity, to such an extent that in the event that we need to bring the specific segment or on the off chance that we need to bring the dataset with particular condition that will be support by Sqoop Tool and information will be put away in hadoop (HDFS).

Sqoop Import

The import instrument imports singular tables from RDBMS to HDFS. Each column in a table is dealt with as a record in HDFS. All records are put away as content information in content documents or as parallel information in Avro and Sequence documents.

Sqoop Export

The fare instrument sends out an arrangement of documents from HDFS back to a RDBMS. The documents given as contribution to Sqoop contain records, which are called as lines in table. Those are perused and parsed into an arrangement of records and delimited with client determined delimiter.

Data Analytic Module with Hive

Hive is an information stockroom framework device to handle organized information in Hadoop. It dwells on top of Hadoop to outline Big Data, and makes questioning and breaking down simple. At first Hive was created by Facebook, later the Apache Software Foundation took it up and created it further as an open source under the name Apache Hive. It is utilized by various organizations. For instance, Amazon utilizes it in Amazon Elastic Map Reduce. Hive is not A social database A plan for On line Transaction Processing (OLTP) A dialect for ongoing inquiries and column level updates.

Features of Hive:

1. It stores diagram in a database and prepared information into HDFS.

2. It is intended for OLAP.
3. It gives SQL sort dialect to questioning called Hive QL or HQL.
4. It is natural, quick, versatile, and extensible.

Hive is an information product house framework for Hadoop. It runs SQL like questions called HQL (Hive inquiry dialect) which gets inside changed over to outline occupations. Hive was produced by Facebook. Hive underpins Data definition Language (DDL), Data Manipulation Language (DML) and client characterized capacities. In this module we need to examination the dataset utilizing HIVE instrument which will be put away in hadoop (HDFS).For investigation dataset HIVE utilizing HQL Language. Utilizing hive we perform Tables manifestations, joins, Partition, Bucketing idea. Hive examination the main Structure Language.

Data Analytic Module with Pig

Apache Pig is a reflection over Map Reduce. It is an instrument/stage which is utilized to break down bigger arrangements of information speaking to them as information streams. Pig is by and large utilized with Hadoop; we can play out every one of the information control operations in Hadoop utilizing Apache Pig. To compose information examination programs, Pig gives an abnormal state dialect known as Pig Latin. This dialect gives different administrators utilizing which software engineers can build up their own capacities for perusing, composing, and handling information. To investigate information utilizing Apache Pig, software engineers need to compose scripts utilizing Pig Latin dialect. Every one of these scripts are inside changed over to Map and Reduce undertakings. Utilizing Pig Latin, software engineers can perform Map Reduce undertakings effortlessly without typing complex codes in Java. Apache Pig utilizes multi-inquiry approach, in this manner decreasing the length of codes. For instance, an operation that would oblige you to sort 200 lines of code (LoC) in Java can be effortlessly done by writing as less as only 10 LoC in Apache Pig. Eventually Apache Pig diminishes the advancement time by just about 16 times. Pig Latin is SQL-like dialect and it is anything but difficult to learn Apache Pig when you know about SQL. Apache Pig gives many implicit administrators to bolster information operations like joins, channels, requesting, and so on. Furthermore, it likewise gives settled information sorts like tuples, sacks, and maps that are lost from Map Reduce. Apache Pig is an abnormal state information stream stage for execution Map Reduce projects of Hadoop. The dialect for Pig will be pig Latin. Pig handles both structure and unstructured dialect. It is additionally top of the guide decrease handle running foundation.

ALGORITHM

Map Reduce Technic

Map Reduce is a handling strategy and a program display for conveyed registering in view of java. The Map Reduce calculation contains two imperative assignments, specifically Map and Reduce. Delineate an arrangement of information and proselytes it into another arrangement of information, where singular components are separated into tuples (key/esteem sets). Furthermore, diminish undertaking, which takes the yield from a guide as an info and joins those information tuples into a littler arrangement of tuples. As the arrangement of the name Map Reduce infers, the diminish assignment is constantly performed after the guide work.

Advantage

The significant favorable position of Map Reduce is that it is anything but difficult to scale information handling over various registering hubs. Under the Map Reduce show, the information handling primitives are called mappers and reducers. Disintegrating an information handling application into mappers and reducers is some of the time nontrivial. Be that as it may, once we compose an application in the Map Reduce frame, scaling the application to keep running more than hundreds, thousands, or even a huge number of machines in a bunch is just an arrangement change. This basic versatility is the thing that has pulled in numerous developers to utilize the Map Reduce show.

The Algorithm

For the most part Map Reduce worldview depends on sending the PC to where the information lives. Map Reduce program executes in three phases, in particular guide arrange, rearrange organize, and lessen arrange.

Map arrange

The guide or mapper's employment is to prepare the information. By and large the information is as record or registry and is put away in the Hadoop document framework (HDFS). The info record is passed to the mapper work line by line. The mapper forms the information and makes a few little pieces of information.

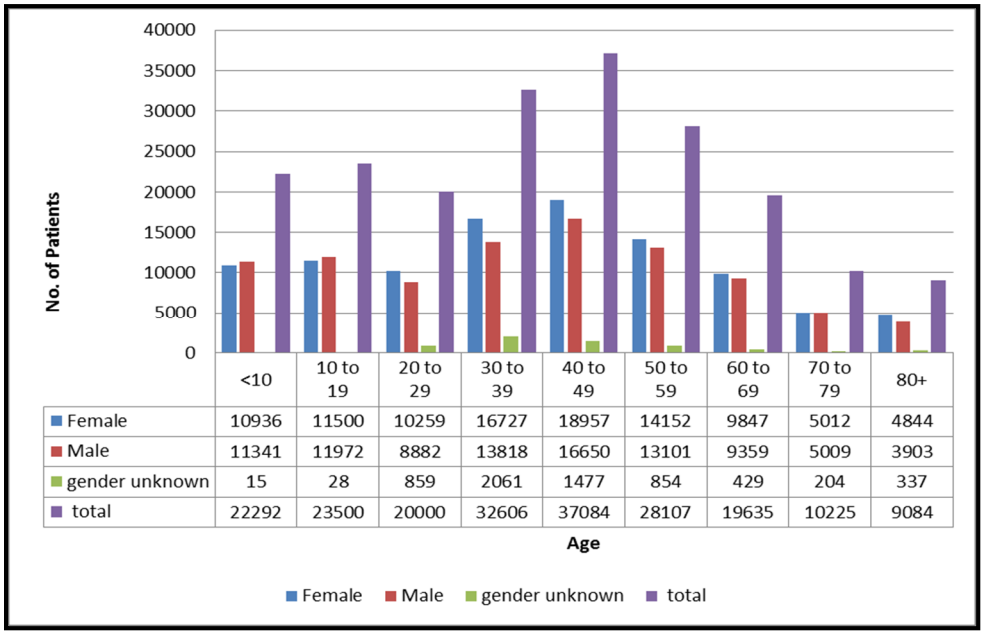
Reduce arrange

This stage is the blend of the Shuffle organize and the Reduce arrange. The Reducer's occupation is to prepare the information that originates from the mapper. Subsequent to preparing, it delivers another arrangement of yield, which will be put away in the HDFS

RESULTS

The execution was measured on four distinctive sub-populaces. Assemble 1 incorporated the entire populace. Gather 2 included clients conceived on or after the year 1948, while Group 3 was made out of clients conceived before the year 1948. These sub-groupings were picked since the normal number of DIH expanded generously between the ages of 61 and 65 years, as Figure No.2 shows, and the middle age of 63 years in year 2011 was taken, relating to a birth year 1948 in this informational index.

In the entire populace, 85% of clients were conceived on or after 1948 and 15% of clients were conceived before 1948. Furthermore, the model was assessed on a sub-populace (Group 4) in which clients had no less than one day (1+ days) in clinic in the year prior to the expectation year (2012). It needs to brought up that, diverse to the next three gatherings, 1+ days gathering was the main gathering in which the partner of clients utilized for preparing and forecast were distinctive. In this gathering, preparing was performed utilizing those clients who had no less than 1 day in healing facility in 2010 and expectation was performed on clients who have no less than 1+ days in the year 2011. These two subsets of clients were not a similar partner, but rather there was cover. In the forecast informational index of 1+ days gathering, 70% of clients were conceived on or after 1948 and 30% of clients were conceived before 1948.



	<10	10 to 19	20 to 29	30 to 39	40 to 49	50 to 59	60 to 69	70 to 79	80+
Female	10936	11500	10259	16727	18957	14152	9847	5012	4844
Male	11341	11972	8882	13818	16650	13101	9359	5009	3903
gender unknown	15	28	859	2061	1477	854	429	204	337
total	22292	23500	20000	32606	37084	28107	19635	10225	9084

Figure No.1: Sample Demographic Statistics of Age and Gender in Year 2015

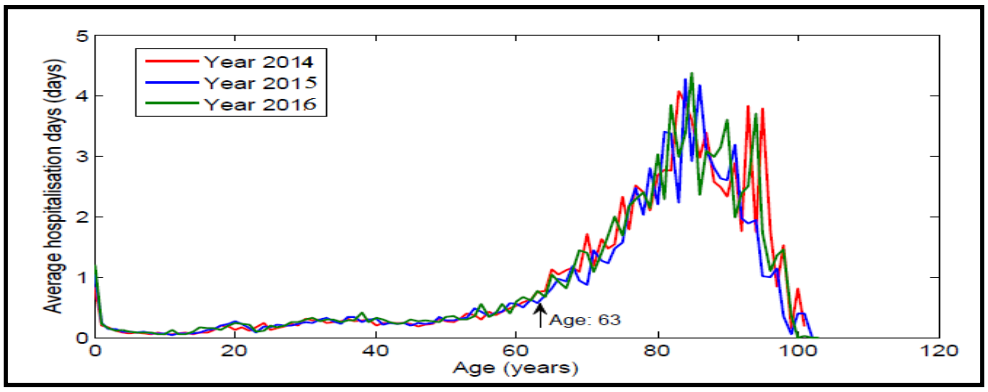


Figure No.2: Average days in healing facility per individual by age for each of the three years of HCF information

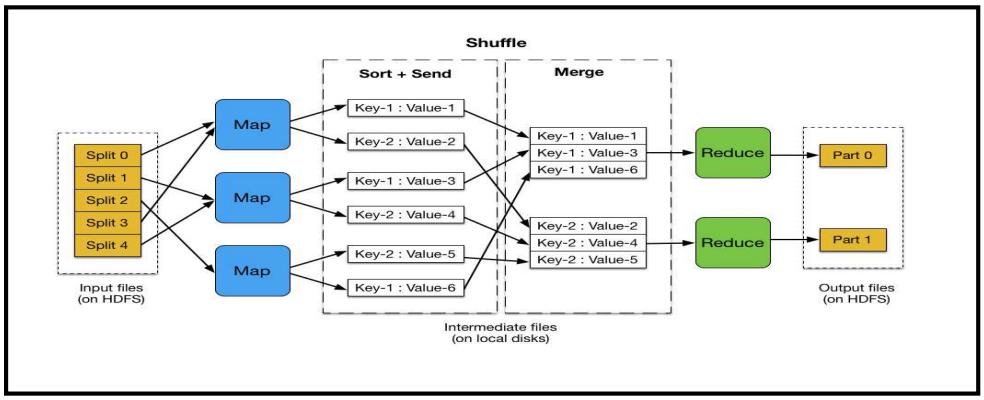


Figure No.3: Map Reduce Data Flow

CONCLUSION

A method for predicting future days in hospital has been developed using features extracted from customer demographics, past hospital admission and hospital procedure claim data. The model was developed using data from an observation period of two years and was later evaluated based on data, we are also interested in forecasting days in hospital on shorter time scales such as by season or month by using big data. The accuracy of forecasting would greatly depend on the density of information available. If the claim information gets too sparse, the prediction accuracy is expected to decrease when doing shorter time scale forecasting, such as weekly predictions. It would also be interesting to know what kind of time resolution of forecasting is able to be supported by claim data sets. From this point of view, tele health data, such as physiological monitoring or self-reported data on a weekly or daily basis, would significantly increase the temporal data density.

ACKNOWLEDGEMENT

The author wish to express their sincere gratitude to Department of Computer Science and Engineering, Anna University, Chennai, Tamil Nadu, India for providing necessary facilities to carry out this review work.

CONFLICT OF INTEREST

We declare that we have no conflict of interest.

BIBLIOGRAPHY

1. Yang Xie, Student Member, IEEE, Gunter Schreier, Senior Member, IEEE, David C W Chang, Sandra Neubauer, Ying Liu, Stephen J. Redmond, Senior Member, IEEE, Nigel H. Lovell, Fellow, *IEEE "Predicting Days in Hospital Using Health Insurance Claims"*, 19(4), 2015, 1224-1233.
2. Donze J, Aujesky D, Williams D, and Schnipper J L. “Potentially Avoidable 30-Day Hospital Readmissions in Medical

Patients: Derivation and Validation of a Prediction Model", *JAMA Internal Medicine,* 173(8), 2013, 632-638.

3. Hasan O, Meltzer D O, Shaykevich S A, Bell C M, Kaboli P J, Auerbach A D, Wetterneck T B, Arora V M, Zhang J, and Schnipper J L. "Hospital readmission in general medicine patients: A prediction model", *Journal of General Internal Medicine,* 25(3), 2010, 211-219.
4. Coiera E, Wang Y, Magrabi F, Concha O P, Gallego B and Runciman W. "Predicting the cumulative risk of death during hospitalization by modeling weekend, weekday and diurnal mortality risks", *BMC Health Services Research,* 14(1), 2014, 226.
5. Cumming R B, Knutson D, Cameron B A and Derrick B. "A comparative analysis of claims-based methods of health risk assessment for commercial populations", *Final Report to the Society of Actuaries,* 2002, 1-85.
6. Zhao Y, Ash A S, Ellis R P, Ayanian J .Z, Pope G C, Bowen B, and Weyuker L. "Predicting pharmacy costs and other medical costs using diagnoses and drug claims", *Medical Care,* 43(1), 2005, 34-43.
7. Bertsimas D, Bjarnadottir M V, Kane M A, Kryder J C, Pandey R, Vempala S and Wang G. "Algorithmic prediction of health-care costs", *Operations Research,* 56(6), 1382-1392, 2008.
8. Pietz K, Ashton C M, McDonell M and Wray N P. "Predicting healthcare costs in a population of veterans affairs beneficiaries using diagnosis-based risk adjustment and self-reported health status", *Medical Care,* 42(10), 2004, 1027-1035.
9. Powers C A, Meyer C M, Roebuck M C and Vaziri B. "Predictive modeling of total healthcare costs using pharmacy claims data - A comparison of alternative econometric cost modeling techniques", *Medical Care,* 43(11), 2005, 1065-1072.
10. How much do we spend on health? Australian Institute of Health and Welfare, Australian Government. [Online], Available: http://www.aihw.gov.au/australias-health/2012/spending-on-health/, 2014.
11. Dove H G, Duncan I and Robb A. "A prediction model for targeting low cost, high-risk members of managed care organizations", *American Journal of Managed Care,* 9(5), 2003, 381-389.
12. Fireman B, Bartlett J and Selby J. "Can disease management reduce health care costs by improving quality?" *Health Affairs,* 23(6), 2004, 63-75.
13. Seto E. "Cost comparison between telemonitoring and usual care of heart failure: A systematic review", *Telemedicine Journal and E-health,* 14(7), 2008, 679-686.
14. Polisena J, Coyle D, Coyle K and McGill S. "Home tele health for chronic disease management: A systematic review and an analysis of economic evaluations," *International Journal of Technology Assessment in Health Care,* 25(3), 2009, 339-349.
15. "Report on government services 2013 volume 2: Health; community services; housing and homelessness", *Steering Committee for the Review of Government Service Provision, Canberra: Productivity Commission,* 2013.
16. Brierley P, Vogel D and Axelrod R. Heritage provider network health prize round 1 milestone prize: How we did it - Team 'Market Makers', [Online], Available: http://www.heritagehealthprize.com/c/hhp/leaderboard/milestone1, 2014.
17. Heritage provider network health prize private leaderboard - heritage health prize, [Online], Available: https://www.heritagehealthprize.com/c/hhp/leaderboard, 2014.
18. "The international statistical classification of diseases and related health problems, 10th revision, australian modification (ICD-10-AM)", *National Centre for Classification in Health,* 1998.

19. "The australian classification of health interventions (ACHI) seventh edition - tabular list of interventions and alphabetic index of interventions," *National Centre for Classification in Health (NCCH),* 2010.
20. Australian refined diagnosis-related groups (AR-DRG) data cubes, Australian Institute of Health and Welfare, Australian Government, [Online], Available: http://www.aihw.gov.au/hospitalsdata/ ar-drg-data-cubes/, 2013.
21. Medicare benefits schedule (MBS) online, Department of Health, Australian Government, [Online], Available: http://www.mbsonline.gov.au/, 2014.
22. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H and Ghali W A. "New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality", *Journal of Clinical Epidemiology,* 57(12), 2004, 1288-1294.
23. Quan H, Li B, Couris C M, Fushimi K, Graham P, Hider P, Januel J M and Sundararajan V. "Updating and validating the Charlson comorbidity index and score for risk adjustment in hospital discharge abstracts using data from 6 countries", *American Journal of Epidemiology,* 173(6), 2011, 676-682.
24. International statistical classification of diseases and related health problems 10th revision, World Health Organization, [Online], Available: http://apps.who.int/classifications/icd10/browse/2010/en, 2010.

Please cite this article in press as: Sadhana Mani. A novel approach for predicting health risk assessment based on patient log, *International Journal of Engineering and Robot Technology,* 4(1), 2017, 13-20.