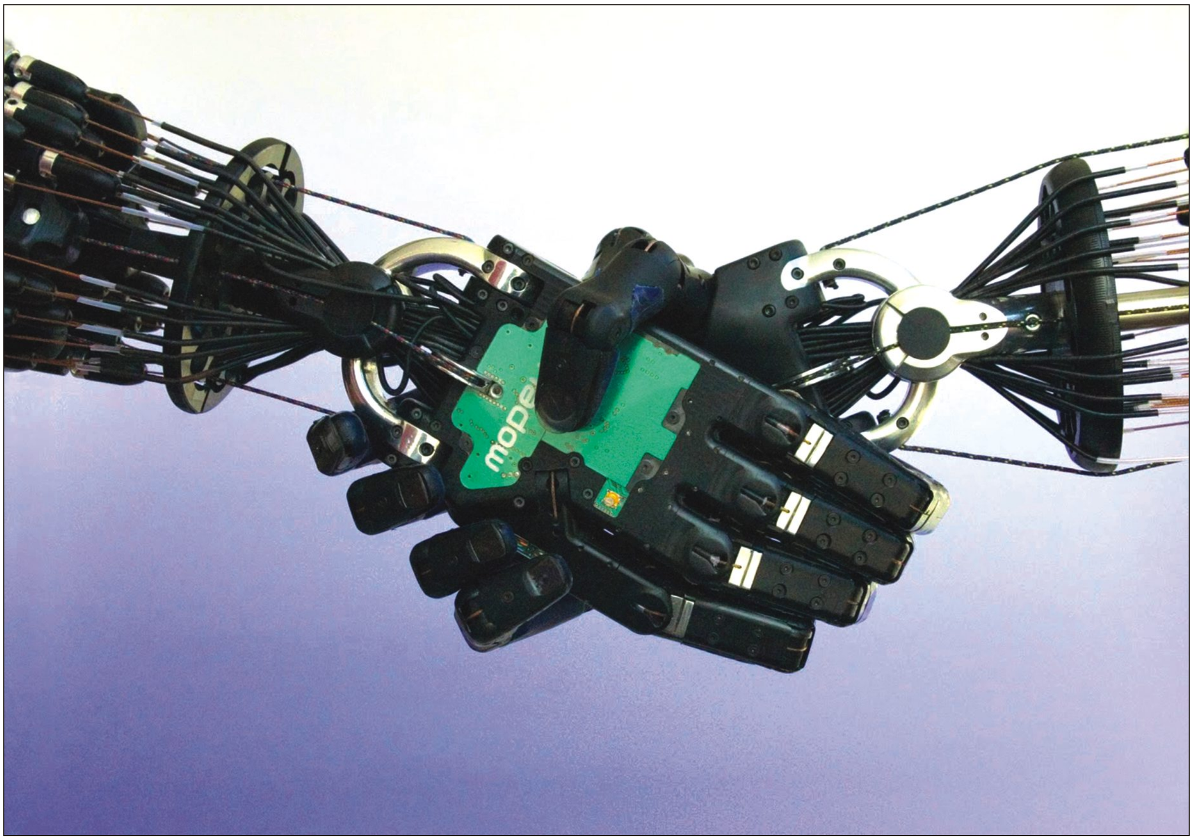


Journal of Artificial Intelligence

ISSN 1994-5450

science alert

ANSInet
an open access publisher
http://ansinet.com

OPEN ACCESS
Journal of Artificial Intelligence

ISSN 1994-5450
DOI: 10.3923/jai.2018.39.47

Research Article
Machine Learning in Biomedical Mining for Disease Detection

Niharika and Baij Nath Kaushik

School of Computer Science and Engineering, Shri Mata Vaishno Devi University, 182320 Katra, Jammu and Kashmir, India

Abstract

Background and Objective: Medicinal service providing industries creates a lot of complex information about patients, healing facilities assets, illnesses, diagnoses strategies, electronic patient's records and so forth. Text identification of handwritten medical transcripts is extremely difficult task in order to diagnose illness. Text mining is an adaptable innovation that can be connected to various distinctive assignments in medical domain. The objective of this review is to extract novel information from scientific text. **Materials and Methods:** In this paper, the authors have methodologically surveyed recent trends in text analytics with regard to developing application realm in the biomedical sciences. The materials and methods used are different types of machine learning classifiers and their respective variants. **Results:** In this, a study of results from past years have been investigated wherein, their approaches and their outcomes are compared through various evaluation measures. **Conclusion:** The survey provides a brief explanation of the stages which are involved in text analytics of medical records. Also it describes up-to-date machine learning techniques with their relevant parameters highlighting the recent trends which are followed by various researchers.

Key words: Biomedical mining, electronic medical records, health care management system, machine learning, text analytics

Citation: Niharika and Baij Nath Kaushik, 2018. Machine learning in biomedical mining for disease detection. J. Artif. Intel., 11: 39-47.

Corresponding Author: Baij Nath Kaushik, School of Computer Science and Engineering, Shri Mata Vaishno Devi University, 182320 Katra, Jammu and Kashmir, India

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

In recent years, the area of Biomedical domain has witnessed significant development and its applications have grown manifold. The data has been rapidly growing exponentially from terabytes to exabytes[1]. The correct scrutiny of many data sources in the biomedical field like patient's demography, background, treatments, symptoms, procedures and medications, etc. needs to be carried out. Significant advances in this discipline lead to large data available in internet sources, various publications, blogs, etc., which then is a primary concern in the area of big data[2,3]. Not just the data size but, the complex domains, unstructured textual health records in medical sciences are also amongst the major concerns now-a-days.

For extracting knowledge, the key way to deal with using vast volumes of accessible chronic diseases-related information is applying machine learning and text mining techniques in different electronic medical records (EMRs) or electronic health records (EHRs) and hospital information systems (HIS)[4]. The serious social effect of the particular infection renders these systems one of the primary needs in medical science examine, which unavoidably produces enormous measures of information. Without a doubt, subsequently, machine learning and text mining approaches in EMRs and HIS are of awesome concern when it comes to diagnosis, administration and other related clinical organization facets[1]. Subsequently, in the framework of this report, vigorous attempts were made to survey the present literature on machine learning and text mining approaches in medical research for different chronic diseases.

Machine learning is a sub-discipline of computer science and a branch of artificial intelligence (AI) for inspiring systems to copy without being doubtlessly customized. In simple terms it can say, it can be elucidated as machines learn from data without depending on rules-based programming[5]. According to Arthur Samuel, machine learning is the sub-field of computer science that, gives "computers the ability to learn without being explicitly programmed." Samuel is an American pioneer in the field of computer gaming and artificial intelligence, who coined the term "machine learning" in 1959[6].

People utilize machine learning innovation in most of the applications like self-driving cars, playing video games automatically, practical speech recognition, effective web search and an infinitely enhanced understanding without even knowing it[7]. Researchers accept this is the most superb approach to make progression towards human-level[8] AI. Utilizing some Profound Learning systems, specialists are getting prepared to make snappy and quick choice along these lines making a move in biomedical field[9].

Text analytics, basically refers to mining of text from a bunch of information to extract meaningful information. In the field of biomedical sciences, text analytics has a major role to play as there is a huge need to convert the data from unstructured format to explicit knowledge. Different tasks and phases involved in text analytics in medical domain are depicted in Fig. 1. The initial step of text analytics of medical data is Information Retrieval. Information retrieval is the tracing and recovery of specific information from stored data.

Information retrieval, as the name infers, concerns bringing of important information from databases[10]. Beginning with the information side of things, the principle issue here is to get a representation of each document and query which is reasonable for a system to utilize it further. It is fundamentally related to ease the client's ingress to a lot of (transcendentally textual) information[11]. The procedure for information retrieval includes the stages given below:

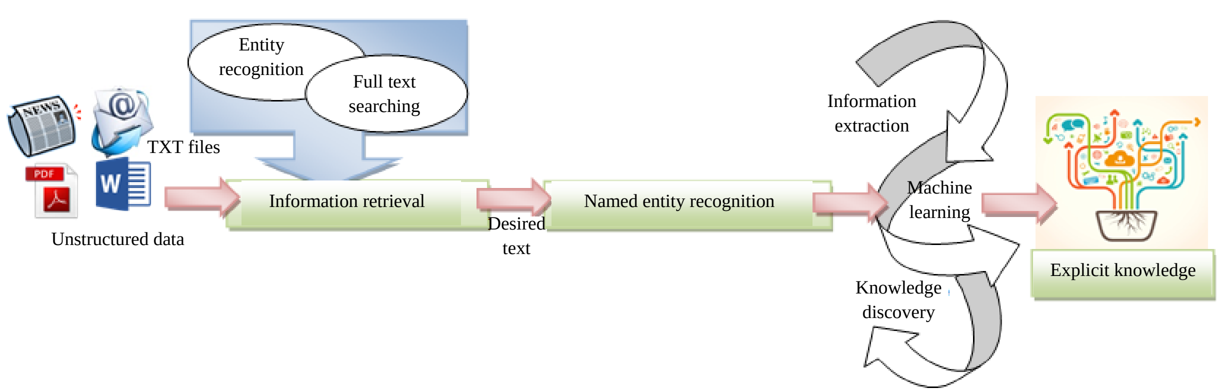


Fig. 1: Stages and errands engaged with medical text analytics

- **Document processing:** In this stage how the documents are represented, collected, retrieved and processed is taken care of
- **User's querying:** In this stage, how a specific user interacts and responds to the queries is taken care of
- **Suitable document:** The selection of an appropriate document, its ranking, order and modeling is taken care of
- **Result:** In this stage, presenting the search result is performed

A retrieval system stores units of information[11]. In information retrieval, entity recognition and full text searching is performed to gather and model all the relevant information of biomedical study. Perhaps a standout amongst the most widely recognized and surely understood use of information retrieval is the recovery of text records from the web. With its current development, the web is turning fast into the fundamental media of transmission for business and scholastic information. Subsequently it is basic to have the capacity to tap the correct document from this immense sea of data. This is indeed, one of the fundamental pushing power for the improvement of information retrieval. Till now, a lot of successful systems have already been developed. Along these lines is characterized here as any device which helps access to records determined by the subject and the operations related to it. This is one of the most important steps in the Medical Text Analytics process. The undesired unstructured information is converted to some useful data.

The next step is Entity Recognition. Also referred to as named entity recognition. Named entities are "atomic elements in text" which belongs to "predetermined classes such as the names of persons, their company names, their residence, expressions of times, quantities, monetary values, percentages, etc.". Named entity recognition (NER) is the task of identifying such named entities. In information extraction, the following goals are kept in mind[7,12].

- Re-arrange the information to make it useful for users
- Information is put in a semantic precise form that permits many results and conclusions to be given by various algorithms

For that machine learning classifiers are incorporated. Now next, with information extraction, extracting the relevant information from a bunch of information to get the desired output, machine learning classifiers and knowledge discovery is used. Machine learning classifiers can be artificial neural network (ANN), support vector machines (SVM), Naïve Bayesian, decision tree methods, deep learning, convolution methods and so on by Santos *et al.*[13] and Kaushik and Banka[8]. By using such techniques the unstructured data is converted to explicit knowledge.

Machine learning is a technique for data investigation that automates explanatory model building[13]. It is a branch of artificial intelligence in light of machines ought to have the capacity to learn and adjust accordingly, keeping in mind their past experiences. The machine learning applied to any software allows it to become more accurate in predicting outcomes without being directly programmed. The essential start of machine learning is to manufacture calculations that can get input data and utilize the factual investigation to foresee a result within an adequate range. The process of machine learning revolves around four different tasks, that are, train model, apply model, capture feedback and prepare data[14]. A different kind of machine learning calculation resembles getting the chances to see the big picture view of AI and what the objective is of the considerable number of things that are being done in the field and place you in a superior position to separate a real time issue and then outline a machine learning framework. Different types of machine learning tasks and their applications are depicted in Fig. 2. Machine learning functions are ordinarily characterized into three general classifications, which are as follows[15]:

- **Supervised learning:** In supervised learning, the training data set is responsible for the machine learning task of inferring a function. The labeled training data has a joint of training examples. There are two types of variables in it, that are, input variables and output variables. The objective function is used to estimate the value of a variable, i.e., output variable also known as dependent variable from a pool of variables, that are input variables, also known as independent variables[16]. Classification and regression are the two categories in supervised learning which includes techniques such as decision tree (DT), k-nearest neighbor (KNN), support vector machine (SVM), genetic algorithms, etc.[17]
- **Unsupervised learning:** In unsupervised learning, machine learning algorithm draws inferences from datasets which consists of input data with no labeled acknowledgments. Clustering is one of the most common unsupervised learning, which is used to find hidden patterns or grouping in data[18]. Another rule learning of this type is association based learning which helps to find out the relations among the objects of a database

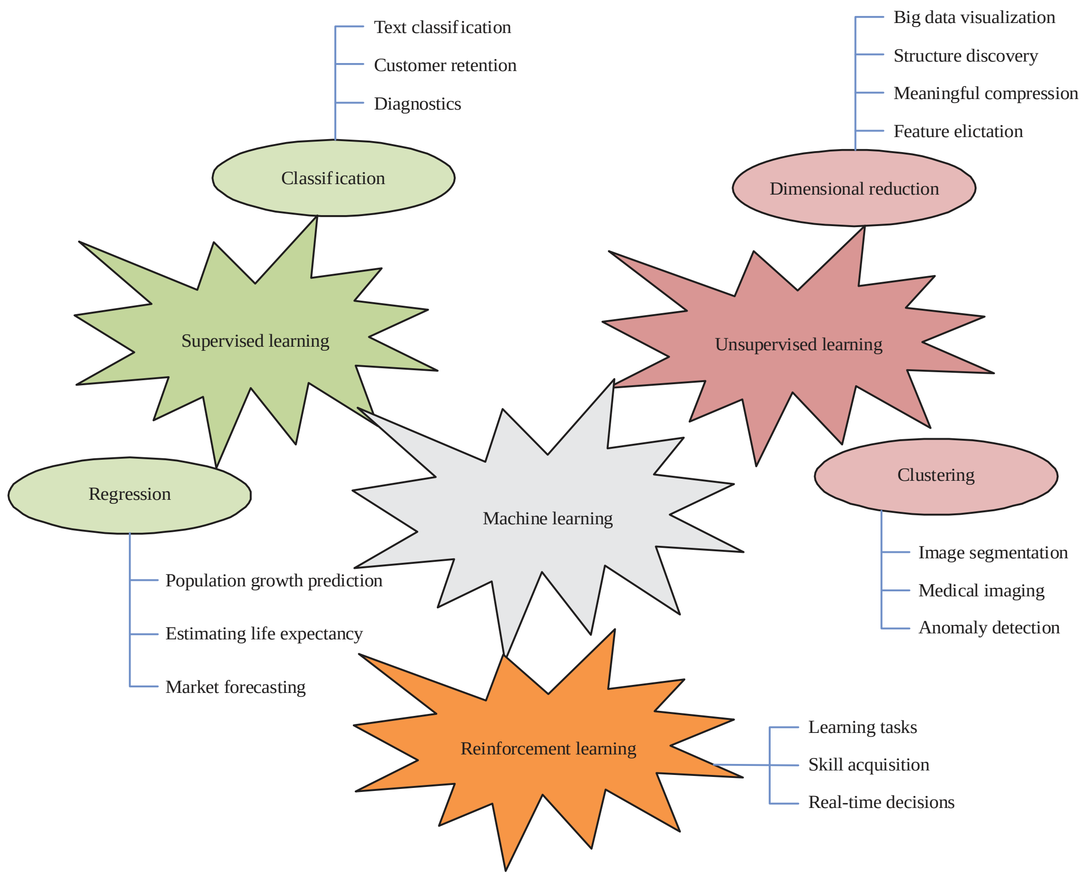


Fig. 2: Types of Machine Learning and their various applications

- **Reinforcement learning:** Reinforcement learning is a kind of machine learning and also a subdivision of artificial intelligence. It enables machines and programming operators to consequently decide the perfect conduct inside a particular setting, keeping in mind the end goal to boost its execution[19,20]. The issue, because of its abstraction, is examined in numerous different controls, like game theory, operational research, statistics and genetic algorithms, swarm intelligence, simulation-based optimization, control theory, multi-agent systems and information theory[21,22]

Knowledge including certainties, data or portrayals, implicit or explicit, alludes to the hypothetical or functional comprehension of an area or a subject. It is of interest to researchers in machine learning, pattern recognition, databases, statistics, artificial intelligence, knowledge acquisition for expert systems and data visualization[23]. The bringing together the objective of the knowledge discovery in databases (KDD) procedure is to separate knowledge from information with regards to huge information bases[18]. Knowledge discovery is able to merge medical information with other n-number of sources of data to generate a futuristic interpretive framework[24]. Discovering knowledge from the biomedical text is a procedure with the intends to discover answers for biomedical inquiries, for example, distinguishing new medication targets or novel malignancy symptomatic bio-markers. Contingent upon how broad the information field is, knowledge discovery can require a computerized program or not[25]. The biggest data field out there recently and unquestionably the most dig capable for knowledge is the World Wide Web.

Therefore, the objective of this study is to identify the optimal classifier from approximate results that can be used for further research in medicinal prescriptions. Additionally, the researchers will find a detailed approach in applying supervised learning onto handwritten as well as online medical transcriptions.

MATERIALS AND METHODS

While using any machine learning classifier or any of its variant, certain measures are adopted to check the efficiency of the algorithm. Hence, for understanding the importance of different chronic disease classification, we need evaluation measures. These parameters are computed from a confusion matrix hat which contains all the information of the original class. All these measures shows the performance of the machine learning classifiers.

Few evaluation parameters that better focuses on the performance a classifier has on the minority class are as follows:

- **Accuracy:** The accuracy of a classifier on a given set of tests is the percentage of test set tuples that are accurately characterized by the classifier. Accuracy is one-fourth the no. of accurately classified instances
- Recall or the true positive (TP) rate is: TP/(TP+FN) of the classifier, where FN is False Negative
- **Precision:** Both of precision and recall tend to trade off each other. Precision, which is the proportion of positives that are classified correctly: TP/(TP+FN). Precision is one-fourth the ratio of accurately classified +ve instances to the overall no. of classifies as +ve
- **F1 Score:** The F1 measure or score takes both recall and precision into consideration. It can be defined as one fourth the harmonic mean between recall and precision. It is used to assess the performance of the acknowledgment
- **AUC:** It is the area under the curve (AUC). One can examine the TP rate vs. the FN rate in the receiver operating characteristic (ROC) curve and the relative AUC value

To date, the significant recent materials in brief and the optimal or average of the methods (evaluation parameters) used are also explained in the Table 1.

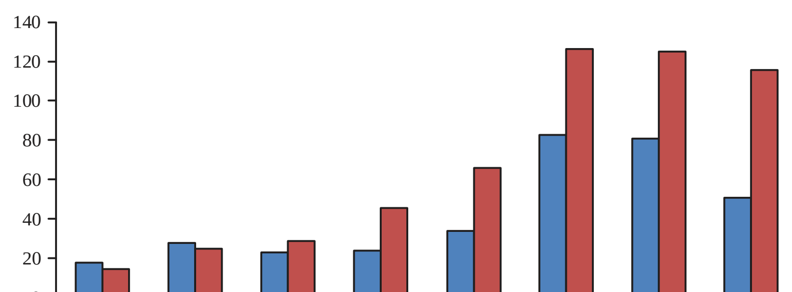
RESULTS AND DISCUSSION

As calculated from PubMed, Fig. 3 shows the yearly distribution of biomedical mining and electronic health records (EHRs) articles onto which machine learning is applied.

The results of ongoing papers from past two years has been shown in the Fig. 4, the values of which are illustrated in Table 2.

Kavakiotis, *et al.*[11] presented a systematic review on machine learning applications, techniques in data mining and the requisite tools in diabetes research. They considered few parameters which include prediction and diagnosis, complication in a diabetic patient, genetic background and health care and management. Also the authors employed different machine learning classifiers for optimal results.

Lucini *et al.*[26] highlighted some text mining to approximate the value of urgent bed demands. Also the future bed demands are approximated by utilizing text data only.The authors have given a robust and strong tool for using text analytics in medical records. Using Nu-Support Vector Classification (Nu-SVC) an average F1 Score was calculated.



	2010	2011	2012	2013	2014	2015	2016	2017
Biomedical mining	18	28	23	24	34	83	81	51
EHRs	18	25	29	46	66	126	125	116

Publications

Fig. 3: Articles per year in the collection employed while using the query word "Biomedical mining/EHRs" and "Machine Learning" in the title or abstract

Table 1: Recent Research Trends for various machine learning techniques used in medical sciences

Publications	Title of the paper	Technologies used	Overview	Evaluation measures (best)
Computational and Structural Biotechnology Journal, Elsevier[11]	"Machine Learning and Data Mining Methods in Diabetes Research"	Gaussian naïve bayes (NB), Random forests (RF) Support vector machine (SVM), K-nearest neighbour (KNN)	Authors have taken online data sets to perform a successful review on various data mining and machine learning algorithms for diabetes disease. Also the review included the best accuracy among different papers	SVM accuracy = 84.09 % RF accuracy = 81.3% SVM AUC = 97.9% RF AUC = 80.3%
International Journal of Medical Informatics, Elsevier[26]	"Text mining approach to predict hospital admission using early medical records from the emergency department"	Decision tree, random forest, multinomial naïve bayes, extremely randomized tree, support vector machine AdaBoost, nu-support vector classification and logistic regression	Authors in this study have considered 8 text mining techniques for prediction method. For feature selection, information measures such as χ^2 and F-score metrics are considered. For all text mining methods analyzed, precision and recall values have been enlightened	F_1 score = 77.70% Std deviation = 0.66%
Journal of Healthcare Informatics Research, Springer[27]	"Association Rule Mining in Multiple, multi- dimensional time series medical data"	Association rules, Pattern Mining, Clustering	In this study, the authors have proposed a technique in which they successfully completed in 2 stages, 1 is of frequent pattern matching and other is of time series attributes which were developed in the 1st stage. They also used time series pattern mining (TSPM) technique	Confidence = 100% Support = 4.5% J-measure = 0.0084
Health Informatics Journal, SSCI[28]	"Medical informatics research trend analysis: A text mining approach"	Singular Value Decomposition (SVD), term-by-document matrix (TDM)	Authors have reviewed various research areas in medical sciences. The data from different medical corpus has been taken, clusters are then created for pre processing step and SVD and TDM is applied onto them for extracting useful knowledge to obtain optimal results. Also term frequency (TF) and inverse document frequency (IDF) for tissue and bone is calculated	TF (tissue) = 1.37 TF (bone) = 1.61 IDF (tissue) = 0.67 IDF (bone) = 0.97
Methods, Elsevier[29]	"DISEASES: Text mining and data integration of disease–gene associations"	Named entity recognition (NER), information extraction (IE), Natural language processing (NLP)	Authors in this study explains the extraction and scoring of disease-gene associations. Also a new resource has been developed, which is called as, DISEASE resource	False positive rate (FPR) = 0.16%
International Journal of Medical Informatics, Elsevier[30]	"Text mining of cancer-related information: Review of current status and future directions"	Named entity recognition (NER), information extraction (IE), natural language processing (NLP)	Review emphasised on robust unfairness between representative methods like information extracton and named entity recognition considering the electronic medical records (EMRs). They also showed the shift from rule based to machine learning for better results	F Measure of NER = 80-90% F Measure of IE = Above 90% Avg Recall = 93% Avg Precision = 94%
Journal of Cardiac Failure, SCI[31]	"Prevalence of heart failure signs and symptoms in a large primary care population through the use of text and data mining of the electronic health record"	Natural language processing (NLP), EPIC electronic health record (EHR), chi-square, t-tests	In this study, authors have explained that applying automated text and data mining of electronic health records (EHRs) for Heart Failure (HF) signs and symptoms is achievable and also they are recorded in case subjects on a timely basic before a clinical diagnosis can be made	Case subjects = 52.22% Control subjects = 24.23%
Journal of Biomedical Informatics, Elsevier[32]	"Biomedical Text Mining and its applications in Cancer Research"	Logistic regression (LR), Support vector machine (SVM)	In this study a wide range of corpus has been explained briefly. Various datasets for entity recognition systems, relationship extraction and for text mining are also elucidated in this methodological review	SVM F_1 Rate = 78.4% LR Recall Rate = 84.64% LR Precision Rate = 95.86%
IEEE Xplore Digital Library[20]	"Mining Medical Data to Identify Frequent Disease using Apriori Algorithm"	Association rules, Apriori Algorithm	In this study, authors have collected the data from hospital information system (HIS) and the requisite data mining tools have been applied using WEKA tool for better diagnosis of any disease. Existing data sets are used to obtain the results by the authors	Accuracy = above 70%

Table 2: Current papers with their evaluating parameters on respective datasets

References	Year	Technologies used	Dataset used	Evaluation parameters (%)
Bannach-Brown *et al.*[33]	2018	Machine learning approaches from SLIM collaboration	Depression training dataset	Accuracy = 87.60 AUC = 93.55
Reddy *et al.*[34]	2018	Logistic regression, regularized regression and gradient boosting machines	EMR patient level dataset	Accuracy = 92.82 AUC = 82.7
Arabasadi *et al.*[35]	2017	Hybrid neural network genetic algorithm	Z-Alizadeh sani dataset	Accuracy = 93.85 AUC = 91
Weng *et al.*[36]	2017	Convolutional recurrent neural network	MGH dataset	Accuracy = 92.50 AUC = 99.10
Weng *et al.*[37]	2017	Convolution neural network	Chest X-ray8 database	Accuracy = 90 AUC = 81
Sarihan and Hanbay[38]	2017	Least-squares support vector machine	Stroke disease dataset	Accuracy = 94 AUC = 74.4

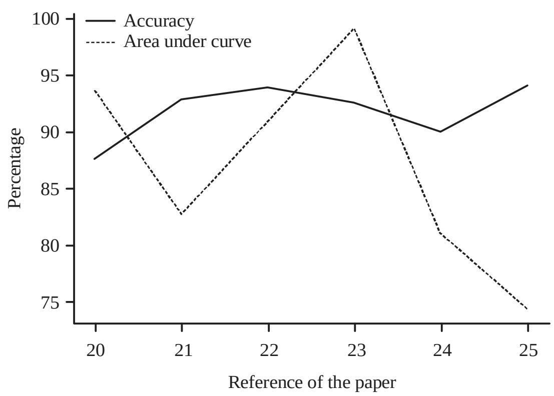


Fig. 4: Comparison of results of resent papers

Pradhan *et al.*[27] introduced hybrid many dimensional associative rules for medical corpus, where most of the time series are linked with many forms of data in related attributes form. They have taken real-time series information of workouts of people which are obtained from multiple electromyogram (EMG) sensors. From this, the authors have shown that their work is collective to know the association rules in medical domain.

In the paper, Kim and Delen[28] distinguished significant branches of medicinal information and investigate the time-variation changes in that. All the publications from PubMed corpus are taken and then a text mining approach is employed to it. Clustering is used to extract the information from various sources. According to the authors in their study, health information technology (HIT) and electronic medical records (EMRs) are growing rapidly so as to discover novel medical insight.

Frankild, *et al.*[29] reported that text mining shall not be employed alone, rather it should be merged with other kinds of affirmation. For that, they have created a novel resource called DISEASE resource in which the results of cancer data from already prevailing databases and the text mining methods created by them are aggregated. The classifiers named Entity Recognition and Information Extraction are also applied along with their DISEASE resource to get optimal results.

Spasic *et al.*[30] showed that for extracting the data from clinical records of cancer ontology text mining techniques are used. The evaluation parameters while using NER lie between 80 and 90%. To improve this performance, rule based methods to machine learning have been incorporated which deals with grammatical mistakes, misspelled words, etc.

Zhu, *et al.*[32] in their study have discussed the latest text mining applications in research of cancer. They have provided some of the resources which are used in cancer text mining. The overview of current work in biomedical text mining is also shown in this review. Types of data sets and mining tools have also been explained very clearly by the authors.

CONCLUSION

At present, there is an immense collection of biomedical text and their fast development which makes it unimaginable for analysts to address the data physically. Researchers can utilize biomedical text mining to find new learning. The authors have looked into the essential research issues identified with text mining in the biomedical domain. The study will help visualize the emerging health hazards/problems and their feature extraction thereof would be useful for public health practice and their solutions to be made accessible for state functionaries at large. The mining and extraction of medical transcripts might not be effortlessly recognized by applying just one machine learning algorithm so, different machine learning classifiers should be applied to build a unified hybrid framework. Also through this analysis, it is comprehended that for mining of handwritten medical records support vector machines (SVM), extreme learning (EL) and various variations of Swarm Techniques are accommodated to get optimal results.

SIGNIFICANCE STATEMENT

In the recent years, there is significant increase in the of machine learning for solving biomedical mining and EMRs or EHRs. But still there are some major areas such as use of text identification from handwritten medical transcripts to identify the disease. Furthermore, feature extraction techniques that are specific to handwritten character recognition can also be developed. In future, the focus will be on data transformation techniques that feed relevant features in deep learning algorithms like convolution neural network (CNN), extreme learning machine (ELM) and particle swarm optimization (PSO).

REFERENCES

1. Kitchin, R., 2014. The real-time city? Big data and smart urbanism. Geo J., 79: 1-14.
2. Lin, W., W. Dou, Z. Zhou and C. Liu, 2015. A cloud-based framework for home-diagnosis service over big medical data. J. Syst. Software, 102: 192-206.
3. Kaushik, B.N. and N.K. Gondhi, 2017. Recent trends of workflow scheduling algorithms in cloud computing under Qos constraints. Proceedings of the 4th International Conference on Signal Processing, Computing and Control, September 21-23, 2017, Solan, India, pp: 396-401.
4. Baker, S., I. Silins, Y. Guo, I. Ali, J. Högberg, U. Stenius and A. Korhonen, 2015. Automatic semantic classification of scientific literature according to the hallmarks of cancer. Bioinformatics, 32: 432-440.
5. Aleven, V.A.W.M.M. and K.R. Koedinger, 2002. An effective metacognitive strategy: Learning by doing and explaining with a computer-based cognitive tutor. Cognitive Sci., 26: 147-179.
6. Kotseruba, I., O.J.A. Gonzalez and J.K. Tsotsos, 2016. A review of 40 years of cognitive architecture research: Focus on perception, attention, learning and applications. CoRR Abs/1610.08602.
7. Meyfroidt, G., F. Guiza, J. Ramon and M. Bruynooghe, 2009. Machine learning techniques to examine large patient databases. Best Pract. Res. Clin. Anaesthesiol., 23: 127-143.
8. Kaushik, B. and H. Banka, 2015. Performance evaluation of Approximated Artificial Neural Network (AANN) algorithm for reliability improvement. Applied Soft Comput., 26: 303-314.
9. Knudsen, P., H. Herborg, A.R. Mortensen, M. Knudsen and A. Hellebek, 2007. Preventing medication errors in community pharmacy: Root-cause analysis of transcription errors. BMJ Qual. Saf., 16: 285-290.
10. Nerkar, B.E. and S.S. Gharde, 2014. Best treatment identification for disease using machine learning approach in relation to short text. IOSR J. Comput. Eng., 16: 5-12.
11. Kavakiotis, I., O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas and I. Chouvarda, 2017. Machine learning and data mining methods in diabetes research. Comput. Struct. Biotechnol. J., 15: 104-116.
12. Raghupathi, W. and V. Raghupathi, 2014. Big data analytics in healthcare: Promise and potential. Health Inform. Sci. Syst., Vol. 2. 10.1186/2047-2501-2-3
13. Santos, R.S., S.M.F. Malheiros, S. Cavalheiro and J.M.P. de Oliveira, 2013. A data mining system for providing analytical information on brain tumors to public health decision makers. Comput. Methods Programs Biomed., 109: 269-282.
14. Hu, X. and H. Liu, 2012. Text Analytics in Social Media. In: Mining Text Data, Aggarwal, C.C. and C.X. Zhai (Eds.)., Springer, Boston, MA., pp: 385-414.
15. Han, J., J. Pei and M. Kamber, 2011. Data Mining: Concepts and Techniques. Elsevier Publishing Company, Amsterdam, Netherlands, ISBN:978-0-12-381479-1, Pages: 702.
16. Pereira, L., R. Rijo, C. Silva and R. Martinho, 2015. Text mining applied to electronic medical records: A literature review. Int. J. E-Health Med. Commun., 6: 1-18.
17. Kushima, M., K. Araki, M. Suzuki, S. Araki and T. Nikama, 2012. Text data mining of the electronic medical record of the chronic hepatitis patient. Proceedings of the International Multi Conference of Engineers and Computer Scientists, Vol. 1, March 14-16, 2012, Hong Kong, pp: 569-573.
18. Massey, J.G., 2015. In context: Extracting relevance from unstructured medical data. Patient Safety & Quality Health.
19. Denny, J.C., 2012. Mining electronic health records in the genomics era. PLoS Comput. Biol., Vol. 8, No. 12. 10.1371/journal.pcbi.1002823.
20. Ilayaraja, M. and T. Meyyappan, 2013. Mining medical data to identify frequent diseases using Apriori algorithm. Proceedings of the International Conference on Pattern Recognition, Informatics and Mobile Engineering, February 21-22, 2013, Karaikudi, India, pp: 194-199.
21. Mladenic, D. and M. Grobelnik, 2003. Text and Web Mining. In: Data Mining and Decision Support, Mladenic, D., N. Lavrac, M. Bohanec and S. Moyle (Eds.). The Springer International Series in Engineering and Computer Science, Vol. 745. Springer, Boston, MA., pp: 15-22.
22. Kaushik, B., N. Kaur and A.K. Kohli, 2011. A new ANN method for measuring overall reliability and performance in growing computer networks with static and variable connections. Int. J. Comput. Sci. Comm., 2: 79-87.
23. Zhu, X., 2007. Knowledge Discovery and Data Mining: Challenges and Realities: Challenges and Realities. IGI Global, UK.
24. Cao, Y.N., P. Zhang, J. Guo and L. Guo, 2014. Mining large-scale event knowledge from web text. Procedia Comput. Sci., 29: 478-487.

25. Kaushik, B., N. Kaur and A.K. Kohli, 2013. Achieving maximum reliability in fault tolerant network design for variable networks. Applied Soft Comput., 13: 3211-3224.
26. Lucini, F.R., F.S. Fogliatto, G.J. da Silveira, J.L. Neyeloff, M.J. Anzanello, R.D.S. Kuchenbecker and B.D. Schaan, 2017. Text mining approach to predict hospital admissions using early medical records from the emergency department. Int. J. Med. Inform., 100: 1-8.
27. Pradhan, G.N. and B. Prabhakaran, 2009. Association rule mining in multiple, multidimensional time series medical data. Proceedings of the IEEE International Conference on Multimedia and Expo, June 28-July 2, 2009, New York, USA., pp: 1720-1723.
28. Kim, Y.M. and D. Delen, 2016. Medical informatics research trend analysis: A text mining approach. Health Inform. J. 10.1177/1460458216678443.
29. Pletscher-Frankild, S., A. Palleja, K. Tsafou, J.X. Binder and L.J. Jensen, 2015. DISEASES: Text mining and data integration of disease-gene associations. Methods, 74: 83-89.
30. Spasic, I., J. Livsey, J.A. Keane and G. Nenadic, 2014. Text mining of cancer-related information: Review of current status and future directions. Int. J. Med. Inform., 83: 605-623.
31. Vijayakrishnan, R., S.R. Steinhubl, K. Ng, J. Sun and R.J. Byrd *et al*., 2014. Prevalence of heart failure signs and symptoms in a large primary care population identified through the use of text and data mining of the electronic health record. J. Cardiac Failure, 20: 459-464.
32. Zhu, F., P. Patumcharoenpol, C. Zhang, Y. Yang and J. Chan *et al*., 2013. Biomedical text mining and its applications in cancer research. J. Biomed. Inform., 46: 200-211.
33. Bannach-Brown, A., P. PrzybyLa, J. Thomas, A.S. Rice, S. Ananiadou, J. Liao and M.R. Macleod, 2018. The use of text-mining and machine learning algorithms in systematic reviews: Reducing workload in preclinical biomedical sciences and reducing human screening error. BioRxiv, 255760. 10.1101/255760.
34. Reddy, B.K., D. Delen and R.K. Agrawal, 2018. Predicting and explaining inflammation in Crohn's disease patients using predictive analytics methods and electronic medical record data. Health Inform. J. 10.1177/1460458217751015.
35. Arabasadi, Z., R. Alizadehsani, M. Roshanzamir, H. Moosaei and A.A. Yarifard, 2017. Computer aided decision making for heart disease detection using hybrid neural network-genetic algorithm. Comput. Methods Programs Biomed., 141: 19-26.
36. Weng, W.H., K.B. Wagholikar, A.T. McCray, P. Szolovits and H.C. Chueh, 2017. Medical subdomain classification of clinical notes using a machine learning-based natural language processing approach. BMC Med. Inform. Decis. Making, Vol. 17, No. 1. 10.1186/s12911-017-0556-8
37. Wang, X., Y. Peng, L. Lu, Z. Lu, M. Bagheri and R.M. Summers, 2017. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA., pp: 3462-3471.
38. Sarihan, M.E. and D. Hanbay, 2017. An expert system for the prediction of stroke disease by different least squares support vector machines models. Biomed. Res., 28: 8760-8764.