Hybrid Technique for Medical Data Classification using Multi-Layer Perceptron with NB Classifier

Thalakola Syam Sundara Rao , Bhanu Prakash Battula

***Abstract:** Medical data analysis gains more interest from the last decade due to its significance advantages. Medical data is a heterogeneous data, which is the combination of text data, numeric data and image data. For to analyze such heterogeneous data traditional data analysis mechanisms are inefficient. To handle this heterogeneous data deep learning is obvious choice. Deep learning is able to handle text, numeric and image data more efficiently than traditional data mining techniques. In this paper we proposed a deep learning based multilayer perceptron to analysis medical data. This method independently address the text data, image data and numerical data and combinable made medical data classification.*

***Keywords :** Medical data, Multi-layer perception, deep learning, medical text, medical images.*

I. INTRODUCTION

Modern medical organizations generates great amount of data which is stored in medical databases[1]. Medical data is data which contains all varitie of datas like text data, image data and numericl data. medical data might cover MRI data, ECG signals, numerical values of blood sugar, cholostral levels and blood pressure as well as thephysician's interpretation. By processing all these kinds of data we can attain great deal of knolowedge to solve all human decesis[2]. And also we can able to improve the human life time. But for to process these kinds of data previously the peoples uses traditional data mining techniques[3]. To process these hetrogenious data, data mining techniques not able to deal with all the kinds of datas. Traditional techniques will able to address with numerical data but not with text and imge datas effectively[4]. Data mining techniquees face processing speed and accuracy while dealing with text and image data. Deep learning is the solution to address such problems[7].

The capacity to learn is a standout amongst the most fundamental attributes of insightful conduct. Deep learning is a subfield of AI concentrating computational strategies that can improve execution on some errand by learning [8].

Revised Manuscript Received on October 30, 2019.
* Correspondence Author

Thalakola Syam Sundara Rao*, Research Scholar, Department of CSE, Acharya Nagarjuna University, Guntur-522508, Andhra Pradesh.
Email: syamsundar.mtech71@gmail.com

Dr Bhanu Prakash Battula, Professor & Head, Department of CSE, Tirumala Engineering College, Guntur,Andhra Pradesh, India.

The point of deep learning exploration might be intellectual, specialized or hypothetical [9]. Psychological points look to display human learning at some dimension. Computerizing the procedure of learning securing for information based frameworks is a case of a specialized point. Hypothetical investigation considers, for instance, attributes of learning techniques, for example, their extension and impediments. Like AI, deep learning is an intrinsically interdisciplinary recorded. Measurement, for instance, is generally used in the field of DL. Deep learning techniques can be grouped based on different criteria, for example, hidden learning procedure, portrayal of information, or application area. Information securing and information mining are imperative application zones of Deep learning. Deep learning techniques have been used in a wide assortment of utilization spaces, for example, Visa misrepresentation location, manually written character acknowledgment, discourse acknowledgment, showcasing investigation, quality control in assembling, aircraft seating designation, sustenance and substance recipe enhancement and programmed order of divine objects[11].

In this paper initially we take medical data bases may contains text data, image data as well as numeric data as input. And apply pre processing of the data for text data we use one approch and for image data we use image pre-procesing mechanisms. After words we use multi layer perceptron mechanism to process medicl data. And we use classifier to classifiy medical data. The rest of the paper is organized as follows section-2 describes about state of the art, section-3 describes proposed mechanism, section-4 presents experimental analysis and finally section-5 concludes the paper.

II. LITERATURE WORK

A few works connected Machine figuring out how to foresee illnesses from the patient clinical status. Lyman GH et al. [2] utilized a four-layer CNN to foresee congestive heart disappointment and interminable obstructive aspiratory infection and indicated noteworthy points of interest over the baselines. RNNs with long momentary memory (LSTM) shrouded units, pooling and word installing were utilized in DeepCare [4], an start to finish Machine unique system that construes current disease states and predicts future therapeutic results. The creators moreover proposed to direct the LSTM unit with a rot impact to deal with sporadic coordinated occasions (which are run of the mill in longitudinal EHRs).

Besides, they consolidated therapeutic mediations in the model to progressively shape the expectations. DeepCare was assessed for illness movement demonstrating, mediation suggestion and future hazard forecast on diabetes and emotional well-being understanding companions[21].
RNNs with gated repetitive unit (GRU) were utilized by Tatonetti et al. [7] to create Doctor AI, a start to finish demonstrate that utilizes persistent history to foresee determinations and drugs for resulting experiences[21]. The assessment demonstrated fundamentally higher review than shallow baselines and great generalizability by adjusting the subsequent model starting with one organization then onto the next without losing significant exactness. In an unexpected way, Miotto et al. [8] proposed to take in Machine patient portrayals from the EHRs utilizing a three-layer Stacked Denoising Autoencoder (SDA).
They connected this novel portrayal on ailment hazard forecast utilizing arbitrary woodland as classifiers. The assessment was performed on 76 214 patients involving 78 ailments from various clinical areas and worldly windows (up to a 1 year). The outcomes demonstrated that the Machine portrayal prompts essentially better forecasts than utilizing crude EHRs or ordinary portrayal learning calculations (for example Essential Component Analysis (PCA), k-implies). In addition, they additionally demonstrated that outcomes altogether improve while including a strategic relapse layer top of the last AE to tweak the whole administered system [10]. So also, Libbrecht et al. [10] utilized RBMs to take in portrayals from EHRs that uncovered novel ideas and exhibited better forecast exactness on various illnesses.
Machine learning was likewise connected to display ceaseless time signals, for example, research facility results, close to the programmed distinguishing proof of explicit phenotypes. For instance, Wang F et al. [11] utilized RNNs with LSTM to perceive designs in multivariate time arrangement of clinical estimations. In particular, they prepared a model to characterize 128 analyses from 13 much of the time however sporadically inspected clinical estimations from patients in pediatric escalated unit care. The outcomes demonstrated huge enhancements with deference to a few in number baselines, including multilayer perceptron prepared available designed highlights. Bellazzi R et al. [12] utilized SDAs regularized with an earlier information dependent on ICD-9s for recognizing trademark examples of physiology in clinical time arrangement. Hripcsak et al. [13] utilized a two-layer stacked AE (without regularization) to display longitudinal successions of serum uric corrosive estimations to recognize the uric-corrosive marks of gout and intense leukemia. Razavian et al. [14] assessed CNNs and RNNs with LSTM units to foresee illness beginning from research facility test estimates alone, indicating preferable exhibitions over strategic relapse with hand-built, clinically applicable highlights.
Neural language Machine models were likewise connected to EHRs, specifically to learn inserted portrayals of restorative ideas, for example, illnesses, prescriptions and research center tests, that could be utilized for investigation and forecast [9]. For instance, Luo J et al. [15] utilized RBMs to learn reflections of ICD-10 codes on a companion of 7578 emotional wellness patients to foresee suicide chance. A Machine design dependent on RNNs likewise acquired promising outcomes in expelling shielded wellbeing data from clinical notes to use the programmed de-recognizable proof of free-content patient synopses [8]. The forecast of impromptu patient readmissions after release as of late gotten consideration too. In this area, Gottlieb et al. [20] proposed Deepr, a start to finish engineering in light of CNNs, which recognizes and joins clinical themes in the longitudinal patient EHRs to stratify restorative dangers[24]. Deepr performed well in foreseeing readmission inside a half year and was ready to recognize significant and interpretable clinical trails[24].

III. PROPOSED WORK

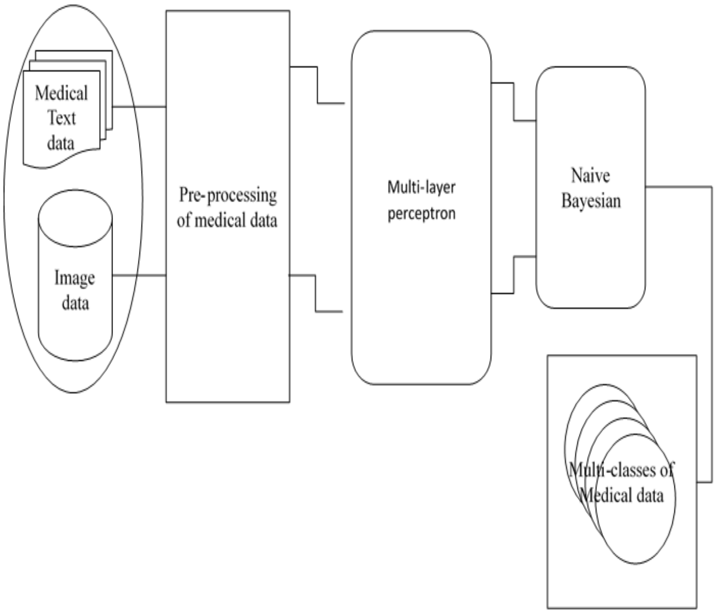

Fig-1:Proposed Framework

Input: medical data it includes medical text and medical images
Output: multiple medical classes
Pre-processing:
Step-1: Text pre processing
Tokenization—transform each and every sentences to words
Eradicating gratuitous punctuation, tags
Stop wordsremoval—recurrent confrontations such "is", "as", "the", etc. that do not have precise semantic
Stemming—reduced words are get as output by reducing avoidable characterizes like suffixes.
Lemmatization—alternative tactic to confiscatevariation by seminal the part of speech and employing comprehensive database of the semantic.
Step-2: Image pre processing
Read image
Resize image
Remove noise (De-noise)
Segmentation
Morphology(smoothing edges)

Multi-layer perceptron

Describe t_i = objective output of item *i*, and a_i = the definite output. This outcome is intended from the total response using a sigmoid equation:

$$a_i = f(u_i) = \frac{1}{1 + e^{-u_i}} \quad (1)$$

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-12, October 2019

In perceptron with one layer, the net input is calculated based on weighted quantity

$$u_i = \sum_j W_{ij} a_j + \theta_i$$

The amount of squared error is intended from the ``target activation", t_i:

$$E = \sum_i (t_i - a_i)^2$$

The whole is over the yield units, since we don't have a clue what the objective actuations ought to be for the shrouded units.

We need to alter the loads such that will limit E. This is bend fitting in a high dimensional space. How might we make E - > 0? E is certainly a component of every one of these loads and predispositions. One approach to do it is to modify the loads toward the negative slope of E, with the goal that we roll out an improvement in each weight:

$$\Delta W_{ij} \propto -\left(\frac{\partial E}{\partial W_{ij}}\right)$$

We can use the chain rule to show that the weight change is:

$$\Delta W_{ij} \propto -\left(\frac{\partial E}{\partial W_{ij}}\right) = -\left(\frac{\partial E}{\partial a_i}\right)\left(\frac{\partial a_i}{\partial W_{ij}}\right) = 2(t_i - a_i)a_i(1 - a_i)a_j$$

The articulation on the correct hand side emerges from a pleasant element of the sigmoidal actuation work. From Eq. (1) you ought to have the capacity to confirm that:

df_i / du_i = f_i (1 - f_i) = a_i (1 - a_i)

Putting in a steady of proportionality that assimilates the factor of two, and including another term that I will clarify in a moment, we have:

$$\Delta W_{ij}(n+1) = \epsilon \delta_i a_j + \alpha \Delta W_{ij}(n),$$

Where *n* is the iteration number, and

$$\delta_i \equiv (t_i - a_i)a_i(1 - a_i).$$

Aside from the last term, Eq. (3) resembles the perceptron learning rule, however with an alternate articulation for delta. Here is the learning rate. On the off chance that it is excessively expansive, we may bounce forward and backward over the way along the inclination, following way b, and may not achieve the base. The last term in the condition above is an additional variety in the calculation that anticipates radical changes in the loads because of the utilization of slope plunge. This term gives our direction in weight space some "energy"as to protect some memory of the bearing it remained going.

It is to utilize an alternate articulation for the delta when part I is a concealed unit rather than a yield unit:

$$\delta_i = \left(\sum_k \delta_k W_{ki}\right) a_i (1 - a_i)$$

The key factor in bracket including the whole over k is an estimation to (ti - ai) for the concealed coatings when we don't know ti. It exploits the outlets that have been determined for the layer above. Memo that the whole over k is the aggregate over the units that get contribution from the ith unit.

The strategy for modifying the weights is:

1. Contemporary contributions for the primary example to the information layer
2. Summation the biased contributions to the following layer and figure their actuations [Eq-2 and 1]
3. Contemporary enactments (2) to the following layer, rehashing (2) until the actuations of the yield layer are known
4. Relate yield enactments to the objective qualities for the example and ascertain deltas for the yield layer [Eq. (4)]
5. Broadcast mistake in reverse by utilizing the yield layer deltas to compute the deltas for the past layer [Eq. (5)]
6. Use these deltas to ascertain those of the past layer, rehashing until the main layer is come to
7. Compute the weightiness vagaries for all loads and inclinations (treat predispositions as loads from a unit having an enactment of 1) [Eq-3]
8. If preparing by example, update every one of the loads and inclinations, else rehash the cycle for all examples, summing the progressions and applying toward the finish of the age
9. Repeat the whole technique until the all-out entirety of squared mistakes is not exactly a predefined basis

Presently we see why it is gotten back to engendering. We begin with a forward pass, displaying an info design, and ascertain the initiations of each layer from those of the first layer, utilizing the present estimations of the loads. When we get to the yield layer, we can contrast the yield actuations with the objective qualities for the given example, and ascertain the delta esteems for the yield layer. Presently we proliferate the mistake in reverse by utilizing these delta esteems to ascertain the deltas for the former layer. In the event that we are preparing by age, we present another example and entirety the delta-W over the arrangement of examples, refreshing the loads toward the finish of every age. At that point we repeat the entire system until the blunder is decreased to an "adequate" esteem. As we did with the single layer perceptron, we alter the predisposition terms by treating them simply like the loads from a (3) unit that dependably has an enactment of 1.

(4)

IV. EXPERIMENTAL ANALYSIS

Here we performed the experiments with python with numpy, matplotlib and keeras packages. Here we used the hardware configurations of 16 GB RAM, 500 GB HDD and we use Ubuntu 16.04 LTS.

Here initially we measure computation cost of the model with respect to the input. That is for training model for images data and also training model with image data. And find the accuracy of the model with respect to the existing solutions.

Sensitivity, specificity and accuracy are described in terms of TP, TN, FN and FP.

Sensitivity = TP/(TP + FN)

Specificity = TN/(TN + FP)

Accuracy = (TN + TP)/(TN+TP+FN+FP)

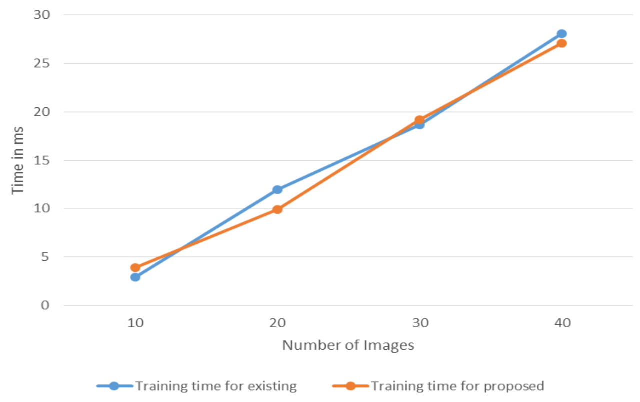

Figure-2: Computation time for training image data to deep learning model

Here figure-2 shows the training time taken for an algorithm with respect to varying the input data that is number of images from 10 to 40 in number. Here figure shows that the time variations. And the proposed methodology takes slightly less time than compared existing.

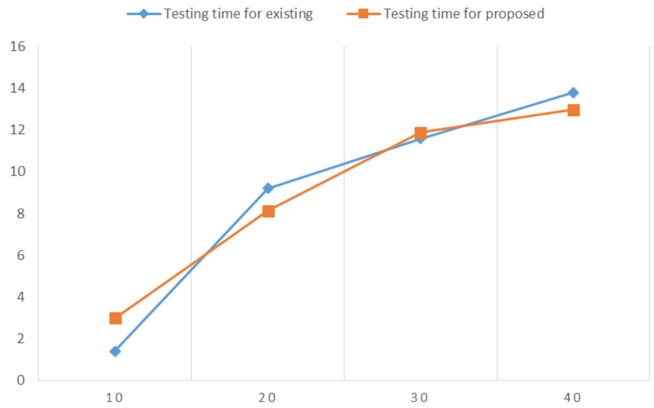

Figure-3: Computation time for Image data testing

Here figure-2 shows the testing time taken for an algorithm with respect to varying the input data that is number of images from 10 to 40 in number. Here figure shows that the time variations. And the proposed methodology takes slightly less time than compared existing.

Figure-4: Computation time for training deep learning model EHR data

Here figure-4 shows the training time taken for an algorithm with respect to varying the input data that is size of the text document varying from 10 to 25 KB size. Here figure shows that the time variations. And the proposed methodology takes slightly less time than compared existing.

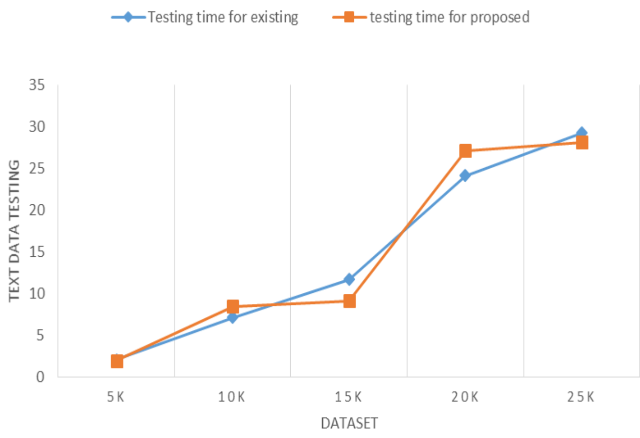

Figure-5: Computation time for training deep learning model EHR data

Here figure-5 shows the testing time taken for an algorithm with respect to varying the input data that is size of the text document varying from 10 to 25 KB size. Here figure shows that the time variations. And the proposed methodology takes slightly less time than compared existing.

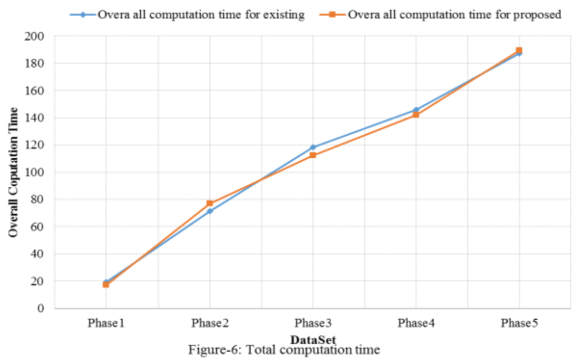

Figure-6: Total computation time

Here figure-6 shows the total computation time for image training, text training, image data testing and text data testing. Here we varying the input data size that was shown in phases wise. In phase one stands for training the model with 10 images and 5 K of text notes, in second phase we can do it with 20 images and 10 K text notes, third phase takes input as 30 images 15K text docs like that we can form the phases of input data to make training and testing. And here we show the computation time overall. Proposed method performs with less computation time because of the organizing of input data.

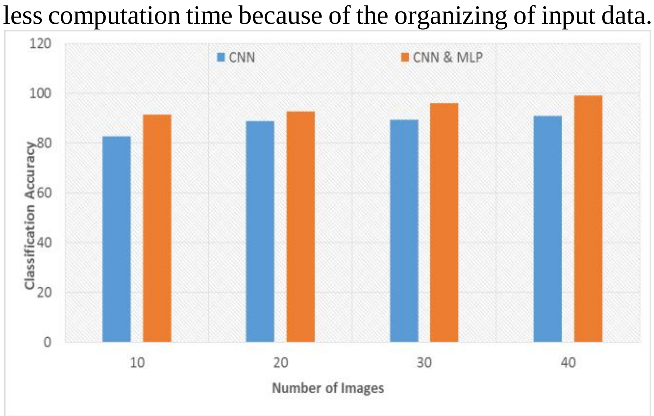

Figure-7: Classification accuracy for medical image data

Here figure-7 shows the classification accuracy of medical image data. Accuracy means number of exact classification with respect to total number of classes. Here this figure shows the classification accuracy of different image data combinations trained by the model. Here existing CNN classification model underperformed compared to proposed MLP and CNN.

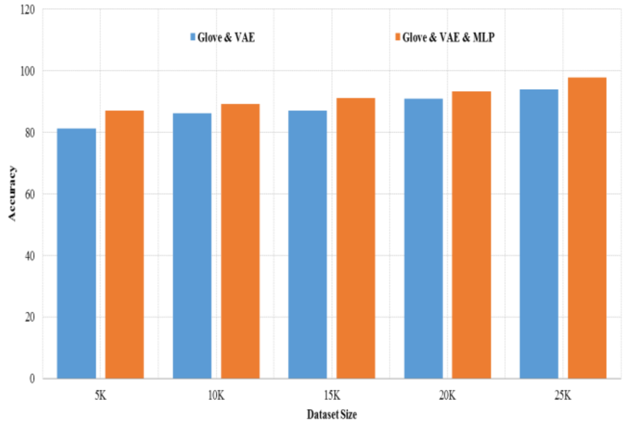

Figure-8: Classification accuracy for EHR data

Here figure-8 shows the classification accuracy of medical text data. Accuracy means number of exact classification with respect to total number of classes. Here this figure shows the classification accuracy of different image data combinations trained by the model. Here existing GloVE classification model underperformed compared to proposed GloVE, VAE and MLP.

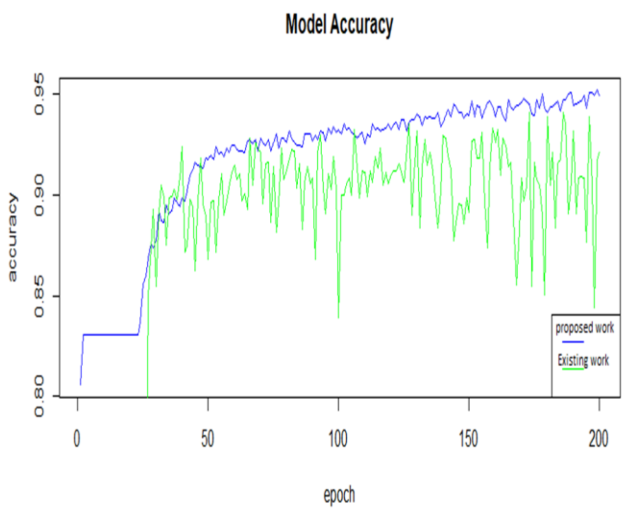

Figure-9: Model classification accuracy

Figure-9 shows the accuracy of overall model which is the hybrid model contains text and image data classification model.

V. CONCLUSION

In this paper initially we take medical data bases may contains text data, image data as well as numeric data as input. And apply pre processing of the data for text data we use one approch and for image data we use image pre-procesing mechanisms. After words we use multi layer perceptron mechanism to process medicl data. And we use classifier to classifiy medical data. The performance results shows that the trainig time of classifier is lower in case of text, image and the accuracy of the classification also increases. The model is shows the better performance when the size of the data increases.

REFERENCES

1. Precision Medicine Initiative (NIH). https://www.nih.gov/precision-medicine-initiative-cohort-program (12 November2016, date last accessed).
2. Lyman GH, Moses HL. Biomarker tests for molecularly targetedtherapies — the key to unlocking precision medicine.N Engl J Med 2016;375:4–6.
3. Collins FS, Varmus H. A new initiative on precision medicine.N Engl J Med 2015;372:793–5.
4. Xu R, Li L, Wang Q. dRiskKB: a large-scale disease-diseaserisk relationship knowledge base constructed from biomedicaltext. BMC Bioinformatics 2014;15:105.
5. Chen Y, Li L, Zhang G-Q, et al. Phenome-driven diseasegenetics prediction toward drug discovery. Bioinformatics2015;31:i276–83.
6. Wang B, Mezlini AM, Demir F, et al. Similarity network fusionfor aggregating data types on a genomic scale. Nat Methods2014;11:333–7.
7. Tatonetti NP, Ye PP, Daneshjou R, et al. Data-driven predictionof drug effects and interactions. Sci Transl Med2012;4:125ra31.
8. Miotto R, Weng C. Case-based reasoning using electronichealth records efficiently identifies eligible patients for clinicaltrials. J Am Med Inform Assoc 2015;22:e141–50.
9. Li L, Cheng W-Y, Glicksberg BS, et al. Identification of type 2diabetes subgroups through topological analysis of patientsimilarity. Sci Transl Med 2015;7:311ra174.
10. Libbrecht MW, Noble WS. Machine learning applications ingenetics and genomics. Nat Rev Genet 2015;16:321–32.
11. Wang F, Zhang P, Wang X, et al. Clinical risk prediction by exploring high-order feature correlations. AMIA AnnualSymposium 2014;2014:1170–9.
12. Bellazzi R, Zupan B. Predictive data mining in clinical medicine:current issues and guidelines. Int J Med Inform2008;77:81–97.
13. Hripcsak G, Albers DJ. Next-generation phenotyping of electronichealth records. J Am Med Inform Assoc 2013;20:117–21.
14. Jensen PB, Jensen LJ, Brunak S. Mining electronic healthrecords: towards better research applications and clinicalcare. Nat Rev Genet 2012;13:395–405.
15. Luo J, Wu M, Gopukumar D, et al. Big data application in biomedicalresearch and health care: a literature review. BiomedInform Insights 2016;8:1–10.
16. SNOMED CT. https://www.nlm.nih.gov/healthit/snomedct/index.html (15 August 2016, date last accessed).
17. Unified Medical Language System (UMLS). https://www.nlm.nih.gov/research/umls/ (15 August 2016, date last accessed).
18. ICD-9 Code. https://www.cms.gov/medicare-coverage-database/staticpages/icd-9-code-lookup.aspx (15 October 2016,date last accessed).
19. Mohan A, Blough DM, Kurc T, et al. Detection of conflicts andinconsistencies in taxonomy-based authorization policies.In: 2011 IEEE International Conference on Bioinformatics andBiomedicine, Atlanta, GA, USA, 2011, 590–4.
20. Gottlieb A, Stein GY, Ruppin E, et al. A method for inferringmedical diagnoses from patient similarities. BMC Med2013;11:194.
21. Bikku, T., Nandam, S. R., &Akepogu, A. R. (2017). An Iterative Hadoop-Based Ensemble Data Classification Model on Distributed Medical Databases. In Proceedings of the First International Conference on Computational Intelligence and Informatics (pp. 341-351). Springer, Singapore.
22. Bengio Y, Courville A, Vincent P. Representation learning:a review and new perspectives. IEEE Trans Pattern Anal MachIntell 2013;35:1798–828.
23. Farhan W, Wang Z, Huang Y, et al. A predictive model formedical events based on contextual embedding of temporalsequences. J Med Internet Res 2016;4:e39.

24. Lakshman Narayana Vejendla , A Peda Gopi and N.Ashok Kumar,(2018),“ Different techniques for hiding the text information using text steganography techniques: A survey”, Ingénierie des Systèmes d'Information, Vol.23, Issue.6,pp.115-125.
25. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature2015;521:436–44.
26. Abdel-Hamid O, Mohamed A-R, Jiang H, et al. Convolutionalneural networks for speech recognition. IEEE/ACM TransAudio Speech Lang Process 2014;22:1533–45.
27. Deng L, Li X. Machine learning paradigms for speech recognition:an overview. IEEE Trans Audio Speech Lang Process2013;21:1060–89.
28. Cho K, Courville A, Bengio Y. Describing multimedia contentusing attention-based encoder–decoder networks. arXiv2015. http://arxiv.org/abs/1507.01053
29. Hannun A, Case C, Casper J, et al. Deep speech: scaling upend-to-end speech recognition. arXiv 2014. https://arxiv.org/abs/1412.5567
30. Google’s DeepMind forms health unit to build medical software.https://www.bloomberg.com/news/articles/2016-02-24/google-s-deepmind-forms-health-unit-to-build-medical-software (4 August 2016, date last accessed).
31. Enlitic uses deep learning to make doctors faster and moreaccurate. http://www.enlitic.com/index.html (29 November2016, date last accessed).
32. Bengio Y, Lamblin P, Popovici D. Greedy layer-wise trainingof deep networks. Adv Neural Inf Process Syst 2007;19:153–60