A Case Study of Pneumonia Prediction utilizing a Big Data Deep Learning Framework Using Keras

Md.Al-Samiul Amin Rishat (id:17-35746-3)[a], Oishi Chowdhury (id:18-36807-1)[a], Sonaly Akter (id:17-33586-1)[a], Raffe,Al Shazid (id:17-33356-1)[a]

[a]*Department of Computer Sciences, American International University-Bangladesh*

Abstract

Huge Information prescient investigation utilizing machine learning methods is right now a much dynamic space of examination in clinical science. With expanding size and intricacy of clinical information like X-beams, profound learning acquired gigantic accomplishment in forecast of numerous deadly illnesses like pneumonia. In this examination work, DCNN (profound convolutional neural organizations) an effective anticipating model for large information, having profound layers is a proposed, which can characterize whether an individual is having a pneumonia or not. The analyses are conveyed after removing the highlights of great X-beam pictures information what's more, accomplished a forecast exactness of 0.84 and AUC of Promising outcomes are discovered, when the consequences of the DCNN structure is contrasted and the ordinary classifiers like SVM, arbitrary timberland, and so forth utilizing diverse assessment measurements like precision, affectability, and so forth With the appearance of expanding instances of pneumonia, thoughtful execution of profound learning can have a major influence in improving the execution of expectation of numerous lethal sicknesses in the future

Keywords: big data, machine learning, prediction, deep learning, pneumonia

1. Introduction

Around 1,000,000 grown-ups are finding with pneumonia furthermore, consistently around 50,000 kick the bucket from this lethal in the US alone[1]. Pneumonia is affecting a great deal kids who are under age of five and furthermore normal reason for death of them overall[2]. Anticipating Pneumonia is significant in the clinical field. Different tests can take some time however anticipating by X-beam of chest will assist the specialist with getting a thought of the illness and steps can be taken likewise. Distinguishing pneumonia by noticing X-beam of chest is an intricate undertaking and is a functioning space of examination. Profound learning just as Large Information are two mainstream fields in the quickly becoming computerized world [3]. While Huge Information has various definitions, this examination work allude it to the veracity for example unstructured information as introduced in the Figure 1, characterizing significant Versus of large information[4]. The clinical information is huge, complex, and hard to examine utilizing traditional information investigation methods. Consequently, profound learning offers an incredible Arrangement in reaping important

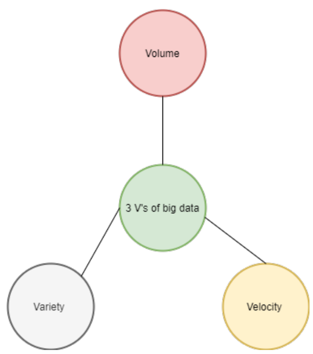

Figure 1: e 1 3 V's of Big Data **?**

information from such unpredictable clinical information. In this exploration work, the proposed forecast model is executed by convolutional profound neural organizations (CNN) utilizing Python programming language. CNNS, otherwise called ConvNets in profound learning. After pre-handling of information, distinctive AI calculations are prepared to quantify the presentation of CNN with well known and current classifiers. Promising outcomes are accomplished, when the aftereffects of the proposed system is contrasted and the customary classifiers like SVM, arbitrary backwoods, adaboost, and so forth utilizing unique assessing measurements like exactness, particularity, region under the bend, and affectability, and so forth.

Research questions

[1]How does DCNN(deep convolutional neural networks) classify pneumonia?

[2]How much convenient big data deep learning framework for predicting pneumonia??

[3]Is big data predictive analytics using machine learning is helpful for medical science?

In case a piece of the paper is coordinated as follows: Area 2 examines about the connected work. Area 3 presents brief conversation of grouping techniques that are utilized in the proposed structure. Area 4 gives insight regarding the information, its highlights and exploratory arrangement. Segment 5 clarifies rundown of the investigation results, charts, and execution measures. Segment 6 presents end and about future scope.

2. Literature Review

Specialists are using the consequences of machine learning forecasts for tackling issues of clinical science [12, 13, 14, 17, 18]. Clinical pictures have enormous volume of data which can be removed and utilized for future counteraction of perilous infections [19]. Numerous specialists have executed AI calculations utilizing Python and R language for separating data from the clinical pictures [17]. Utilization of troupe techniques for advancing the aftereffects of forecast exactness is much in pattern today. Group classifiers centers around hybridization for improving the consequences of AI expectation model [20]. As of late, profound learning is a lot of dynamic space of examination in clinical science. Greenspan et al. has audited the present furthermore, future points of view of profound learning in clinical science [21]. Forecast model utilizing Convolutional Neural Organizations (CNN) helps in giving much better exploratory outcomes for high dimensional picture information [21]. High dimensional information comprises of the clinical pictures which has enormous number of highlight descriptors. Highlight extraction procedures are applied on acceptable quality X beam pictures to remove the various element descriptors. Profound learning neural organizations are prepared with the removed information to construct the expectation model [21]. Exploration is too conveyed for expectation of pneumonia utilizing machine learning classifiers [22].

3. Proposed Method

The primary motivation behind investigating the field of AI is get a prepared model for the arrangement and forecast of pneumonia patients, thinking about accessible X-beam information. The result of DCNN proposed system assists with anticipating if an individual has pneumonia. A. Proposed system The result of proposed DCNN structure assists with foreseeing if an individual has pneumonia dependent on the X-beam picture of chest. Typically, in genuine situation issues, there is less authority over the nature of pictures. Some customary pre processing like expulsion of ruined pictures, cleaning, and so forth are continuously required[6]. AI point is to receive productive strategies to handle enormous and complex information likewise considering cost. The theoretical and itemized perspective on DCNN structure is shown in the Figure 2 and Figure 3 separately. Picture is first preprocessed in required organization for taking care of to neural network and furthermore checked for any adulterated picture and eliminated it. Then, at that point, the changed over information goes through the DCNN where different highlights are removed at each level. Toward the end of the DCNN there is completely associated layer and afterward the last layer which is yield layer which anticipate '1' or '0' , '1' for pneumonia and '0' for ordinary. With the assistance of back propagation the organization learns the right loads. After the model is prepared it very well may be utilized for foresee yield on information which it has not seen before.

The information on which we need to anticipate ought to be in same arrangement as is preparing information. We tried our model on premise of different measurements (Exactness, AUC , blunder rate, affectability, and so forth) Think about this load of execution measurements, models can measure up well since it is nice to realize what is TP rate as it shows in the number of cases the will do the right characterization.

The individual having pneumonia ought to be named positive however, on the off chance that an individual not having pneumonia is delegated positive is anything but a major issue as it tends to be additionally redressed.

3.1. AI Classifiers

I. Neural Organization: The thought depends on human cerebrum working, similar to neuron imparts in human mind a similar idea is applied here. There are unique layer of neurons and they initiate different neurons, as this is learns right weight for expectation [8].

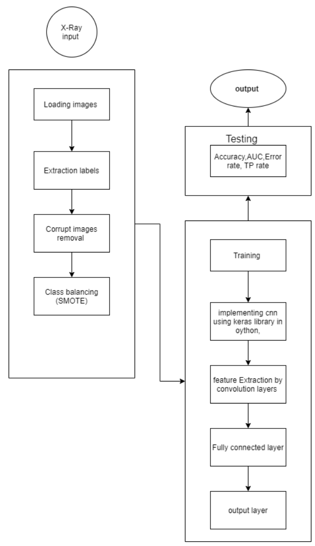

Figure 2: Proposed Prediction Framework

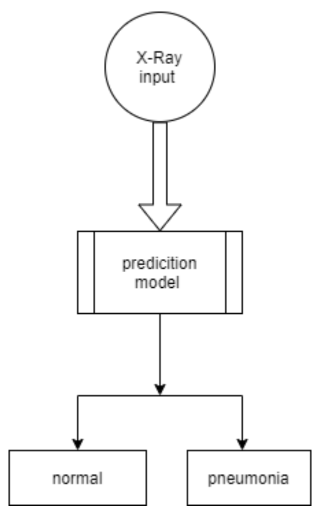

Figure 3: DCNN Framework

ii. Arbitrary Timberland : It outfits aftereffects of various choice trees and take their normal, thusly further develops exactness and furthermore try not to over-fit [7].

iii. Backing Vector Machine: Utilizing vector, this strategy finds a hyperplane between the datasets. The hyperplane behaves like a divider between the unique classes. Checked the classification of new inconspicuous information (in which bunch it falls as all are isolated by hyper plane) likewise and results are additionally displayed here. The measurement relies upon the quantity of highlights [9].

iv. Adaboost : It's anything but a gathering based strategy in which the yield of one become contribution of next tree after a few changes. Doing so works on the exactness and over-fitting [10].

v. Strategic Relapse : This is a grouping technique which become familiar with some connection in the dependant variable (mark) and autonomous factors (highlights) by considering the likelihood [11,15].

vi. Choice Tree: It's anything but a diagram based AI classifier [16]

4. Results

This segment examines boundary assessment measurements to measure the presentation of different AI calculations. The outcomes are talked about much exhaustively and are additionally introduced graphically.

4.1. Execution Assessment

The outcomes and execution of the proposed structure is assessed with various boundaries displayed in disarray network Table 1. The different assessment measurements determined from the Table 1 are introduced in Table 2.

Table 1 Confusion Matrix

Predicted Condition	Condition Positive	Condition Negative
Pneumonia Positive	T P (A)	F P (C)
Pneumonia Negative	F N (D)	T N (B)

Table 2 Performance Metric Formula

Sensitivity	A/(A + B)
Specificity	B/(D + B)
Accuracy	(A + B)/(A + C + D + B)
F Score	(2 *A)/((2 *A) + (D + C))
MCC	(A * B)(D * C)/SQRT((A + D)*(A + C) * (B + D) * (B + C)

In light of different measurements DCNN performed better compared to different models as displayed in Figure 4. As it was imbalanced information, we can't thoroughly rely upon precision so contrasting different measurements results DCNN gives great outcomes. Neural Organization and Arbitrary Woodland likewise very great and are solid contenders. Looking at TP rate and FP rate DCNN keeping up with its stand. Generally speaking DCNN is giving productive outcome on concealed information.

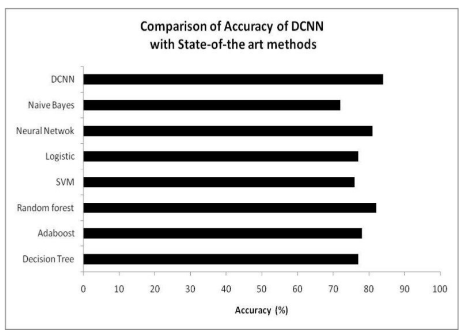

Figure 4: DCNN Framework

5. Discussion

After experimentation and testing of different models,it can easily be observed that accuracy and other parameters of the DCNN are the best among all other models.The performance of DCNN is better than the state-of-the art methods. False positive rate is low for DCNN and True positive rate is high which is good as they show that model is working good on unseen data. The patient having pneumonia is more likely be detected. The chance of cases of patients having pneumonia but classified as normal is low which is shown by False positive rate. Many models have accuracy close to one another but when we consider other metrics then we can compare the models easily. TP rate is also an important metric to consider when comparing the models.

6. Conclusion

Pneumonia is hazardous in the event that it's anything but analyzed appropriately in patients. Around two third of the worldwide populace needs admittance to radiology diagnostics in India, as indicated by an gauge by the World Well being Association. In this examination, chest X beam picture reports are used to prepare a proficient profound AI based expectation model for foreseeing pneumonia in patients. Profound learning makes this assignment more successful as profound learning is effective if there should arise an occurrence of picture information handling. A productive model is constructed utilizing profound learning calculations in Python language which will assist specialists with identifying this dangerous infection. The proposed system is contrasted and cutting edge strategies for AI and discovered to be more effective in expectation with a normal exactness of 0.84, which is discovered to be superior to any remaining classifiers. For future work, enhancement of results will be accomplished for working on the exhibition of forecast. Further, more volume of picture information will be gathered and information preparing is done on the highest point of Hadoop framwork

7. References

1. P. Rajpurkar et al."Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning."arXiv preprint arXiv:1711.0522,2017
2. WHO. Pneumonia, 2016 [Online] Available:http://www.who.int/news-room/fact-sheets/detail/pneumonia [Accessed: May 24, 2018]
3. X. Chen. "Big data deep learning: challenges and perspectives."IEEE access", pp. 514-525, 2014.

4. A. Gandomi et al. "Beyond the hype: Big data concepts, methods, and analytics."International Journal of Information Management vol 35(2),pp.137-144, 2014.

5. Chest XRay data, 2018, [ONLINE] Available:http://dx.doi.org/10.17632/2file-41d542e7-7f91-47f6-9ff2- dd8e5a5a7861 [Accessed: May 24, 2018]

6. H. Nishtha et al. "B2FSE framework for high dimensional imbalanced data: A case study for drug toxicity prediction."Neurocomputing vol. 276, pp.31-41, 2018.

7. Liaw, Andy, and Matthew Wiener. "Classification and regression by randomForest."R news vol. 2(3), pp. 18-22, 2002.

8. A. Rowley et al. . "Neural network-based face detection."IEEE Transactions on pattern analysis and machine intelligence vol. 20(1), pp. 23-38, 1998.

9. M. Hearst, et al. "Support vector machines."IEEE Intelligent Systems and their applications vol. 13(4), pp. 18-28, 1998.

10. R.. Takashi Onoda, and K-R. Müller. "Soft margins for AdaBoost."Machine learning, vol. 42(3), pp. 287-320, 2001.

11. Hosmer Jr, David W., Stanley Lemeshow, and Rodney X. Sturdivant. Applied logistic regression. Vol. 398. John Wiley Sons, 2013.

12. P. Pedro et al. . Community-acquired pneumonia: identification and evaluation of non responders. Therapeutic advances in infectious disease, 1(1), pp. 5-17, 2013.

13. M. Aydogdu et al. Mortality prediction in community-acquired pneumonia requiring mechanical ventilation; values of pneumonia and intensive care unit severity scores. Tuberk Toraks, vol. 58(1), pp. 25–34, 2010.

14. D. Mollura et al. White paper report of the rad-aid conference on international radiology for developing countries: identifying challenges, opportunities, and strategies for imaging services in the developing world. Journal of the American College of Radiology, vol. 7(7), pp. 495– 500, 2010.

15. Press, S. James, and Sandra Wilson. "Choosing between logistic regression and discriminant analysis." Journal of the American Statistical Association 73.364 (1978): 699-705

16. Safavian, S. Rasoul, and David Landgrebe. "A survey of decision tree classifier methodology." IEEE transactions on systems, man, and cybernetics 21.3 (1991): 660-674.

17. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. Artificial Intelligence in medicine. 2001 Aug 1;23(1):89-109.

    Magoulas GD, Prentza A. Machine learning in medical applications. InAdvanced Course on Artificial Intelligence 1999 Jul 5 (pp. 300-307). Springer, Berlin, Heidelberg.pneumonia pattern using RNA-Seq and machine learning: challenges and solutions. BMC genomics. 2018 May;19(2):101.

18. Wernick MN, Yang Y, Brankov JG, Yourganov G, Strother SC. Machine learning in medical imaging. IEEE signal processing magazine. 2010 Jul;27(4):25-38.

19. Dietterich, Thomas G. "Ensemble methods in machine learning." International workshop on multiple classifier systems. Springer, Berlin,

Heidelberg, 2000.
20. Greenspan H, Van Ginneken B, Summers RM. Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique. IEEE Transactions on Medical Imaging. 2016 May;35(5):1153-9.
21. Choi Y, Liu TT, Pankratz DG, Colby TV, Barth NM, Lynch DA, Walsh PS, Raghu G, Kennedy GC, Huang J. Identification of usual interstitial pneumonia pattern using RNA-Seq and machine learning: challenges and solutions. BMC genomics. 2018 May;19(2):101

8. Contribution Table

ID	Name	Contribution
Md.Al-Samiul Amin Rishat	17-35746-3	Literature Review,Proposed Method
Oishi Chowdhury	18-36807-1	Introduction,Results
Raffe,Al Shazid	17-33356-1	Abstract,Discussion
Sonaly Akter	17-33586-1	Conclusion