ISSN NO. 2320–7418 DOI: 10.22270/jmpas.V10I4.1395

International open access journal

Journal of Medical Pharmaceutical and Allied Sciences

Journal homepage: www.jmpas.com

Research article

Comprehensive analysis on comparison of machine learning and deep learning applications on cardiac arrest

Debnath Bhattacharyya[1]*, B. Dinesh Reddy[2], Nakka Marline Joys Kumari[3], N. Thirupathi Rao[2]

1. Koneru Laksmaiah Education Foundation, Greenfield, Vaddeswaram, Andhra Pradesh, India
2. Vignan Institute of Information Technology, Visakhapatnam, Andhra Pradesh, India
3. Vignan Foundation for Science, Technology and Research, Guntur, Andhra Pradesh, India

ABSTRACT

Machine Learning is the technology of having machines to understand and behave as humans do. Refining their learning in supervised manner over time, by feeding them information and data in the form of experiences in the real world. Heart disease has a wide variety of consequences, varying from asymptomatically to extreme arrhythmias, and even premature cardiac failure. A comparative computational analysis was conducted on open-source datasets among the most frequently used classification algorithms in Machine Learning and Neural Networks by randomly splitting data in to test and training and in-depth survey of feature selection is addressed. Our study further concentrates on working with massive datasets from prospective study.

Keywords: Cardio Vascular Diseases, SVM, k-means, Machine learning, Neural Networks.

Received - 19/06/2021, Reviewed - 30/06/2021, Revised/ Accepted- 04/07/2021

Correspondence: Debnath Bhattacharyya* ✉ debnathb@kluniversity.in

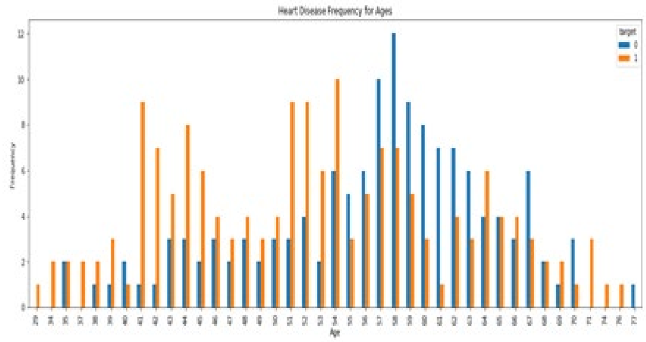
Department of computer science and engineering, Koneru Laksmaiah Education Foundation, Greenfield, Vaddeswaram, Guntur, Andhra Pradesh, India

INTRODUCTION

Cardiovascular disease and Cardiac arrests head the list of Significant [(1)] Concern. They are known globally as a leading cause of death. Most of the Cardio vascular disease regarded as the primary factor in mortality rate among men and Women of all ages and genders. As Heart disease is a non-communicable risk condition includes CAD history of age, ethnicity, age and family, while modifiable risk indicators involve elevated Blood pressure [(2),] Blood cholesterol, Blood triglyceride, diabetes, drug consumption, opium usage and smoking. Changing the lifestyle will control responsive risk factors to some extent. According to estimates from the World Health Organization, count of cardiac arrests contributed to about 17.9 million deaths in 2016 [(3)]. The study of cardiovascular diseases and tracing the symptoms in prior helps to build a good recommendation system for Doctors and Research. As, it is a concern of taking a wise decision with mere response of time is accounted for saving the lives of a patient.

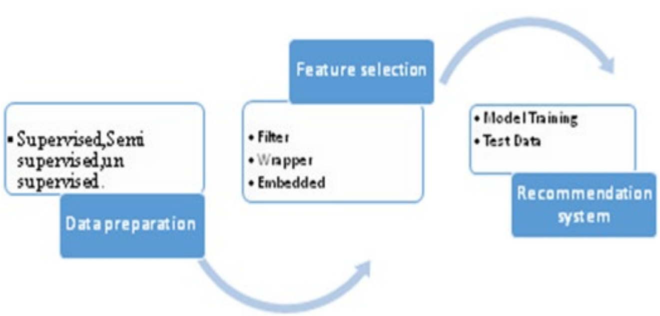
The doctors then require good instructions on when to expect a heart attack, how to assess it, and how to act after a diagnosis is made. In addition, the guidelines will be focused on data gathered from the community of interest representing local patterns [(4)] of local practice. The present studies analyze the predictions using retrospective study. i.e., learning from available and previous health records [(5)]. Doctors not only require specific instructions on when to perform an ECG and when to refer a patient, they need to consider the risk of early intervention mortality as well as the medical-implications of this situation. Machine Learning has become wide spread and a primary tool for promoting clinical study and practice on extracting Electronic Health Records (EHRs) [(6)]. In Comparison to supervised learning, unsupervised machine learning has shown ability to classify novel patterns and associations from EHRs without the need for human generated Labels. Nevertheless, the huge and unprecedented amounts of historic record and the constant streaming of data [(7)] produced in medical services are now an enormous challenge for collecting, storing, and analyzing.

Figure 1: A graph that represents correlation of heart diseases with respective to Frequency of age



Machine Learning holds tremendous promise [8] to reduce product and services prices, automate market operations and better serve consumers. During this period of exponential technical progress, it is regarded as one of the most relevant technology fields and its implementation is gathering popularity across all the industries including Healthcare too. Machine Learning [9] enables to take decisions on their own, based on the past data. It is more likely suitable to low end systems, most of the features to be identified are Hand written and Experts. Most of the features needed to be identified well in advance and code manually. Helpful for Intelligence efficiency in healthcare. Deep Learning enables to take decisions with the help of artificial neural network based actual value and estimated value. Deep learning [10] is suitable for training voluminous data, not suitable for small datasets. Its high end working and testing time give a quality support for massive data in healthcare. Deep Learning algorithms includes Recurrent Neural networks, Convolution neural networks, Generative Adversial Networks, Perceptron**.**

Figure 2: Steps for Feature Selection

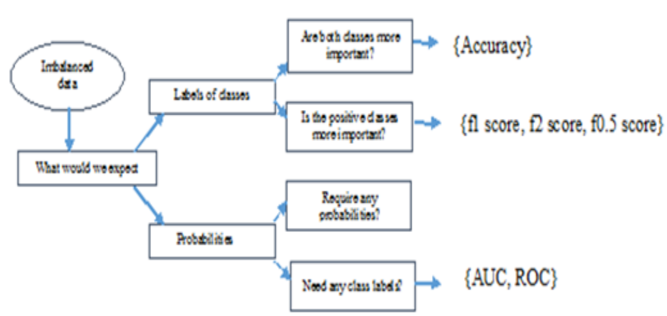


Related works

Suggests clinicians can use Machine learning models to predict the prognosis of patients of HF with preserved ejection fraction. Machine learning methods and patient features that are readily accessible, the existing models estimate mortality risk and HF hospitalization in HFpEF patients.[11] In addition, the findings highlight the fact that statistics on standard of living and health status data that are not regularly obtained in a clinical experience have a significant effect on outcomes of HFpEF patients. explores a process called ensemble classification, which is used by integrating several classifiers to increase the performance of weak algorithms.[12] A comparative theoretical methodology was developed to decide how to use the ensemble method to improve predictive accuracy of heart disease. Ensemble approach for boosting, bagging, majority voting is used for comparison. Out of all majority voting algorithms improves an accuracy to 7.26%.[13] Hyper trophic cardiomyopathy is perhaps the most common hereditary cardiovascular disease in young adults, and a significant cause of heart attack and stroke referred as sudden cardiac arrest.

Throughout the study, Machine Learning algorithms are used to build and test a predictive approach that fixes data discrepancy, using a collection of clinical variables to classify patients with hyper trophic cardiomyopathy (HC) and defined as prolonged ventricular fibrillation. [14] Usage of these algorithms gives the ability to alter the model as more clinical data are accessible as well dealing with imbalanced data. Data acquired from clinical records and suggestive for the impact of heart risk on Low exercise. Discuss the findings of benchmarking for multiple clinical data task of prediction such as death prediction, duration of stay prediction, and cohorts using deep learning algorithms. [15] Deep learning algorithms tend to outperform all the other methods, especially from large amount of data from the series raw clinical time as input features on publicly available datasets medical information mart for intensive care [16] III (MIMIC-III) For extracting appropriate features from the dataset, two types of feature selection algorithms such as relief and univariate feature selection are used. Four traditional feature selection methods [17] were used. Their study improved the accuracy with the help of applying hyper parameters tuning and cross validations. The application of this data is used for managing streams of data from twitter and complete experimentation is achieved by integrating Apache spark with Apache Kafka. Accuracy obtained has accuracy of 94%.

Figure 3: Dealing with imbalanced data for enhancing better performances



Combined work teams, which include medical professionals and mathematicians or computer scientists, are suitable for conceptualizing and designing machine learning programs. implemented the work on around 4000 patients(18) using two different tools R-STUDIO and RAPID MINER suggests that R-studio is a valuable resource for performing extremely accurate, dynamic Machine Learning analytics in building a database of all changes. Rapid Miner uses a very basic and straightforward graphical interface to run and stimulate Machine Learning algorithms, but its ability to manage the parameters can be smaller and less accurate in case of complex analyses. Suggests that recurrent neural networks have been used for HER (19) data to assess disease onset risks. Although these models showed positive effects on relatively limited datasets, there was no assessment of the availability and generalizability of such models and their applicability across hospitals to specific patient populations. So as a further study of improvement Recurrent Neural Networks is implemented on massive datasets of electronic healthcare data. Gated units and long-term recurrent memory were used on EHR by considering, surgery reports, age and number of visits of patients.

Further study enhances better performances ranges across various patient populations, and through research is required before RNN models are applied in practice. No unique data form can encompass the complexities of all the related factors to interpreting a phenomenon such as a disease. Hence, integrative approaches which integrate data from different technologies have developed as essential computational and statistical strategies. Considering variability in only one form of data will skip several critical trends that can only be found when taking in to account different types of biomedical data. Possible example of using diagnostics of the disease as an informative point. The Machine Learning algorithm estimates a likelihood that a disease may develop in the near future. To make reasonable Prediction, the machine learning systems needs to use many distinct types of data. Accurate estimations can only be achieved by the study of data from different categories of patient. With respective to age increases, the frequency and chances of having a disease, Clinical features and outcomes of the elderly patients hospitalized with acute myocardial infractions are evaluated.

Figure 4: Process of Machine Learning

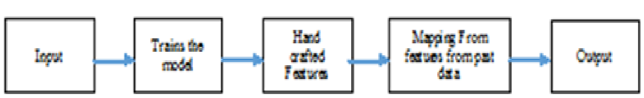
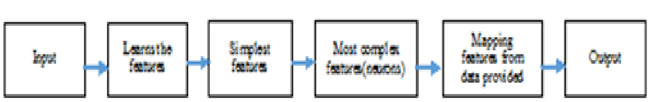


Figure 5: Process of Deep Learning



Feature selection is key for making simpler, faster and more reliable machine learning algorithms. Feature selection or variable selection is the process of selecting a subset of relevant features from the total features of a level in a data set to build machine learning algorithms. Variety of reasons for feature selection. First simpler models are easier to interpret. It is easier for the users of the model to understand the output of a model that uses 10 variables than the output of a model that uses 100 variables. A second reason is, shorter training times. Reducing the number of variables used to build the machine learning models reduces the computational cost and therefore speeds up model building but more importantly simpler models, also score faster the applications, which is particularly important if the model is in a live environment where decisions need to be made sub second. Third, is enhanced generalization by reducing over fitting?

Very often, many of the variables are the noise with little if any predictive value. The machine learning models learn, however from this noise causing over fitting and reducing generalization. By eliminating irrelevant noisy features, we can improve substantially the generalization of a machine learning model. Fourth, it is easier to implement by software developers. When the machine learning model is deployed, often the software developers need to write code to call the variables that need to be fed into the model to produce the output. Another reason to select features is reduced risk of data errors, variable redundancy during model use. In Address on types of feature selections. Filter methods are feature selection procedures that rely on the characteristics of the data. Filter methods do not involve machine learning algorithms at all at the time of screening the features. These methods evaluate the features only and make a selection based on the feature characteristics. Filter methods tend to be less computationally expensive than any other feature selection procedures.

However, they usually give a lower prediction performance compared to wrapper methods or embedded methods. On the other hand, filter methods are very well suited for a quick screen and fast removal of the most irrelevant features from a data set. Wrapper methods instead use a predictive machine learning algorithm to consider the optimal feature subset. Wrapper methods, in essence, will build a machine learning algorithm for each feature subset they evaluate and then select the subset of variables that produces the highest performing algorithm. (Luu and Luu 2018) There will therefore build simpler machine learning models, at each round of feature selection. These make these methods very computationally expensive and sometimes they are even not possible to run with the available computing resources.

MATERIALS AND METHODS

Data Description

The data used in this study is the Cleveland Clinic Foundation.

Figure 6: Depicts the correlations of features

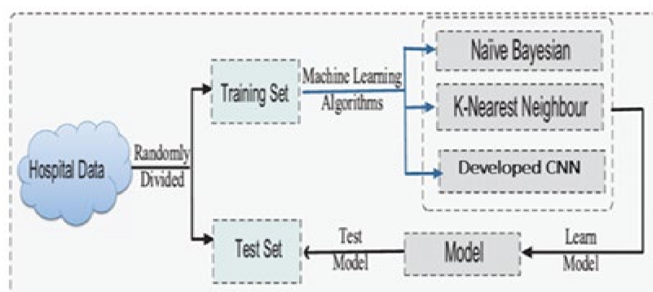
Implementation and experimental evaluation

Experimentation Setup

It is important to know what kind of data you're dealing with. Because different techniques might have different difficulties depending on the kind of data, we are handling. There are numerical, categorical and ordinal data. The numerical data is possibly the most common type of data. It can also include people heights, page load times, and stock prices. Now there are basically two kinds of numerical data. There's discrete data, which is integer based. For example, that could be counts of some sort of event like how many purchases did a customer make in a year? The discrete value has an integer restriction to it. The other type of numerical data is continuous data, which has an infinite range of possibilities. For example, going back to the height of people. There is an infinite number of possible heights with people. The second type of data is categorical data. And this is data that has no inherent numeric meaning. We can't really compare one category to another directly. Examples like categorizing gender, yes, no questions, product category, and political party. We can assign numbers to these categories. But those numbers have no

Inherent meaning. There's no real numerical, quantifiable meaning to them. Categorical data does not have any intrinsic numerical meaning, Ordinal Data is a mixture of categorical and numerical, that has mathematical meaning. For example, Rating of movie, based on user interest he can refer his choice as either 1 or 5 for good and bad.

Figure 7: Process for Building Machine Learning Model



Experimentation 1

Comparison of six different classifiers, with train split as 70 and test split as 30 to enhance its performance.

K-Nearest Neighbors

KNN specifies the solving of both classification and regression issues using K-nearest Points. Classification based on neighbors is a form of instance-based Learning or non-generalized learning: it does not seek to create a general internal model but merely stores instances of training data.

$$d(a,b) = d(b,a)$$
$$= \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + ------ + (b_n - a_n)^2}$$

Naïve Bayes

$$p(q|r) = \frac{p(q \mid r)\, p(q)}{p(q)} \quad (2)$$

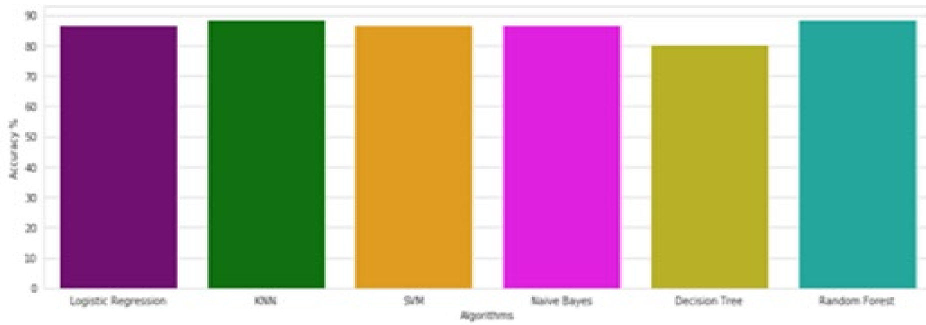
Multiple Linear regressions

X= a0+a1x1+a2x2+ ----- +an*xn (3)

SVM

A Support vector machine builds a hyper plane in a high dimensional space and can used for both classifications and regressions. The hyper plane which has the largest distance to the nearest training points of any class achieve fine separation.

Figure 8: Comparisons of Traditional Machine Learning Classifier



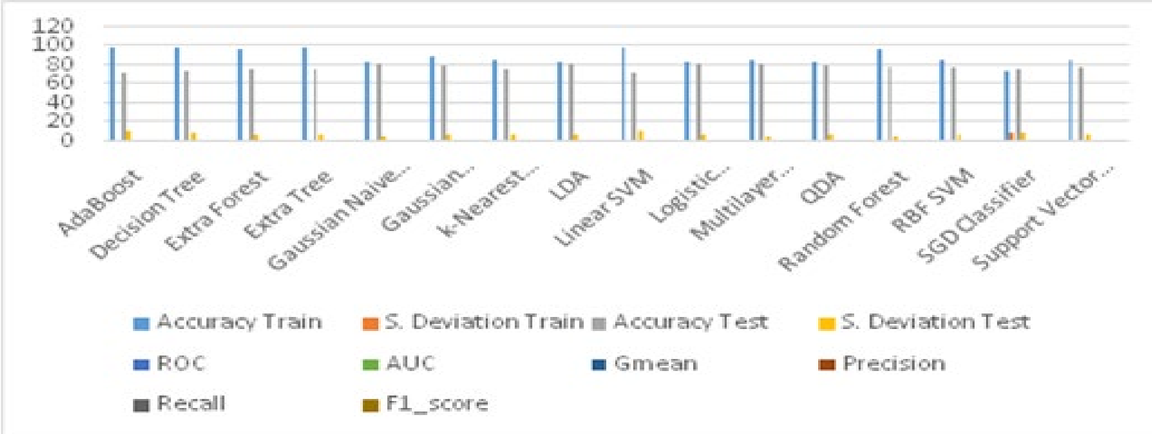
Experimentation 2

Computing the performance of twelve classifiers with respective to Performance measures

Table 1: Comparing all Twelve Traditional Machine Learning classifiers with respective to their Performance Measures

Classifier	Accuracy Train	S.Deviation Train	Accuracy Test	S. Deviation Test	ROC	AUC	Gmean	Precision	Recall	F1_score
AdaBoost	99.528245	0.157256	72.467532	11.743417	0.783188	0.5	0.650935	0.734694	0.837209	0.782609
Decision Tree	99.528245	0.157256	74.891775	9.185245	0.793605	0.5	0.66322	0.75	0.837209	0.791209
Extra Forest	97.640948	0.633105	75.865801	7.905176	0.806444	0.5	0.678426	0.745098	0.883721	0.808511
Extra Tree	99.528245	0.157256	75.38961	6.788432	0.836483	0.5	0.718936	0.804348	0.860465	0.831461
Gaussian Naive Bayes	83.80518	0.635471	82.554113	5.585254	0.888566	0.5	0.80128	0.902439	0.860465	0.880952
Gaussian Process	88.678699	1.059879	79.177489	6.815123	0.86531	0.5	0.768697	0.897436	0.813953	0.853659
k-Nearest Neighbors	86.111325	0.816833	75.4329	7.242031	0.844477	0.5	0.735529	0.853659	0.813953	0.833333
LDA	83.961697	1.06127	82.077922	8.35666	0.822432	0.5	0.705598	0.829268	0.790698	0.809524
Linear SVM	99.528245	0.157256	72.467532	11.743417	0.783188	0.5	0.650935	0.734694	0.837209	0.782609
Logistic Regression	83.489942	1.338543	82.554113	7.256863	0.854893	0.5	0.751849	0.875	0.813953	0.843373
Multilayer Perceptron	85.272802	0.675243	82.554113	5.977476	0.887355	0.5	0.802374	0.923077	0.837209	0.878049
QDA	83.489942	0.762968	80.17316	6.915034	0.856105	0.5	0.750738	0.857143	0.837209	0.847059
Random Forest	97.64205	0.628973	77.748918	6.186691	0.80281	0.5	0.676036	0.777778	0.813953	0.795455
RBF SVM	85.4814	0.860542	78.246753	7.490085	0.886143	0.5	0.803882	0.945946	0.813953	0.875
SGD Classifier	75.031965	9.986706	77.012987	10.260197	0.686289	0.5	0.616939	0.591549	0.976744	0.736842
Support Vector Machine	85.4814	0.860542	78.246753	7.490085	0.886143	0.5	0.803882	0.945946	0.813953	0.875

Figure 9: Comparisons of Traditional Machine Learning Classifier



Experimentation 3

Normalization is the process of scaling individual samples to have unit norm. This process can be useful if we plan to use a quadratic form such as the dot-product or any other kernel to quantify the similarity of any pair of samples. For polynomial kernels in SVM if we don't have normalized data it takes forever to train. Normalization makes algorithms execute quickly and can boost accuracy by huge amounts.

A changed=A−Amin*Amax−Amin → (5)

Role of Hyper parameters: Kernel: The kernel to use (linear, polynomial, rbf, sigmoid), Gamma: Kernel coefficient for rbf, poly and sigmoid. Degree: Degree of the Polynomial kernel function ('poly').

Polynomial Kernel: k (a, b) = a*b (6)

RBF Kernel: k (a, b) = (a*b+1) d (7)

Sigmoid Kernel: k (a, b) = e-b|a-b||a-b| (8)

Performance of each Kernels in SVM Classifier with a test and train split of 50% each.

Figure 10: Comparisons of Different Kernels in SVM, by a very distance the linear kernel outperforms other two kernels

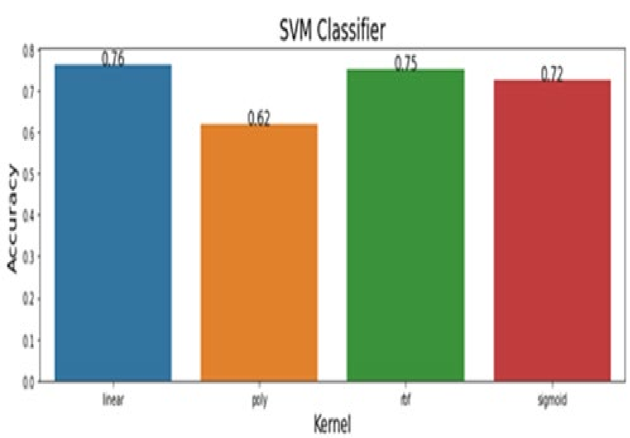
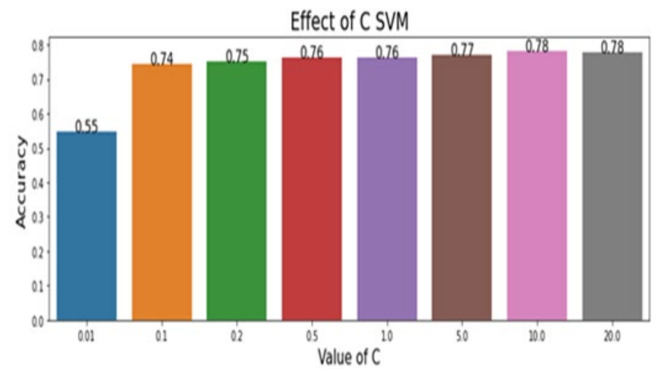


Figure 11: C is the inverse of the force of regularization; must be a positive float. Lesser Value provide for stronger regularization



As Voting Strategy improves the accuracy and better performance of a classifier in fig.11 shown below.

Figure 12: Performance of voting classifier to improve the weaker data

We may never know which range of hyper parameters would fit better, but we can certainly say that those are worth tuning. Many hyper parameters are probably less powerful than others. Knowing this we will give some time to ourselves. We cannot apply same analysis for large datasets, but we can adjust a few hyper parameters that normally make a lot of difference.

Experimentation 4

Neural Networks Performance with 30% test set and 70% training set.

Algorithm	Time	Correctly Classified	Incorrectly Classified
SVM	0.09 sec	77.34%	22.66%
Naïve Bayes	0.03 sec	76.30%	23.70%
Neural Networks (Perceptron)	0.81 sec	75.13%	24.87%
Random Forest	0.55 sec	74.74%	25.26%
JRip	0.19 sec	74.48%	25.52%
Decision Tree (J48)	0.14 sec	73.83%	26.17%
Decision Table	0.23 sec	72.40%	27.60%

CONCLUSION

Our work contributes understanding of Machine learning and Deep Learning Applications in assisting Recommendation system for quick decisions. Survey on importance of feature selection and tuning parameters were discussed. The further work can be contributed by applying all the computational applications on massive data collected from all domain experts, as it is a care of concern for livelihood, keen study is required.

REFERENCES

1. Ahmed, Hager, Eman MG Younis, Abdeltawab Hendawi, and Abdelmgeid A Ali 2019. “Heart Disease Identification from Patients’ Social Posts, Machine Learning Solution on Spark” Future Generation Computer Systems
2. Angraal, Suveen, 2020. “Machine Learning Prediction of Mortality and Hospitalization in Heart Failure With Preserved Ejection Fraction” JACC: Heart Failure 8(1): 12–21
3. Beunza, Juan Jose, 2019. “Comparison of Machine Learning Algorithms for Clinical Event Prediction (Risk of Coronary Heart Disease)” Journal of Biomedical Informatics 97(May): 103257
4. Bhattacharya, Moumita, 2019. “Identifying Ventricular Arrhythmias and Their Predictors by Applying Machine Learning Methods to Electronic Health Records in Patients With Hypertrophic Cardiomyopathy (HCM-VAr-Risk Model)” American Journal of Cardiology 123(10): 1681–89
5. Force Thomas, 2011. “NIH Public Access” 122 (11): 1130–33
6. Kourou, Konstantina, 2015. “Machine Learning Applications in Cancer Prognosis and Prediction” Computational and Structural Biotechnology Journal 13: 8–17
7. Kubat, Miroslav 2017. “An Introduction to Machine Learning” An Introduction to Machine Learning (June): 1–348
8. Latha, C Beulah Christalin, S Carolin Jeeva, 2019. “Improving the Accuracy of Prediction of Heart Disease Risk Based on Ensemble Classification Techniques.” Informatics in Medicine Unlocked 16
9. Purushotham, Sanjay, Chuizheng Meng, Zhengping Che, Yan Liu, 2018 “Benchmarking Deep Learning Models on Large Healthcare Datasets” Journal of Biomedical Informatics 83: 112–34.
10. Rasmy, Laila, 2018 “A Study of Generalizability of Recurrent Neural Network-Based Predictive Models for Heart Failure Onset Risk Using a Large and Heterogeneous EHR Data Set” Journal of Biomedical Informatics 84: 11–16
11. Tsai, Chun Wei, Chin Feng Lai, Han Chieh Chao, Athanasios V Vasilakos 2015. “Big Data Analytics: A Survey” Journal of Big Data 2(1): 1–32
12. Tseng, Chih-jen, Chi-jie Lu, Chi-chang Chang, and Gin-den Chen, 2017. “Artificial Intelligence in Medicine Integration of Data Mining Classification Techniques and Ensemble Learning to Identify Risk Factors and Diagnose Ovarian Cancer Recurrence” 78: 47–54
13. Zitnik, Marinka, 2019. “Machine Learning for Integrating Data in Biology and Medicine: Principles, Practice, and Opportunities” Information Fusion
14. Sk Meeravali, Debnath Bhattacharyya, N.Thirupathi Rao & Yu-Chen Hu, 2021. Performance analysis of an improved forked communication network model, Connection Science
15. Bhattacharyya, D, Kumari, N M J, Joshua, E S N, & Rao, N T, 2020. Advanced Empirical Studies on Group Governance of the Novel Corona Virus, MERS, SARS and EBOLA: A Systematic Study Int J Cur Res Rev| Vol, 12(18), 35
16. Kumari, N M J, Krishna, K K, 2018. Prognosis of Diseases Using Machine Learning Algorithms: A Survey In 2018 International Conference on Current Trends towards Converging Technologies (ICCTCT) (pp. 1-9) IEEE
17. Joshua, ESN, Chakkravarthy, M, Bhattacharyya, D, 2020. An extensive review on lung cancer detection using machine learning techniques: A systematic study Revue d'Intelligence Artificielle, Vol 34, No 3, pp 351-359.
19. Eali Stephen Neal Joshua, Debnath Bhattacharyya, Midhun Chakkravarthy, Yung-Cheol Byun, 2021. "3D CNN with Visual Insights for Early Detection of Lung Cancer Using Gradient-Weighted Class Activation", Journal of Healthcare Engineering, vo6 5, ID 6695518, 11 pages.

How to cite this article

Debnath Bhattacharyya, B. Dinesh Reddy, Nakka Marline Joys Kumari, N Thirupathi Rao, 2021. “Comprehensive analysis on comparison of machine learning and deep learning applications on cardiac arrest” Jour. of Med. P’ceutical & Alli. Sci. V 10 – I4, 1395 P-3125-3131. DOI: 10.22270/jmpas. V10I4.1395.