An Imperative Diagnostic Model for Predicting CHD using Deep Learning

Conference Paper · November 2020

DOI: 10.1109/INOCON50539.2020.9298423

An Imperative Diagnostic Model for Predicting CHD using Deep Learning

Nimmala Mangathayaru
Dept. of Information Technology
VNR VJIET
Hyderabad-90, T.S, India
mangathayaru_n@vnrvjiet.in

B.Padmaja Rani
Dept. of Computer Science & Engg
JNTUH College of Engineering,
Hyderabad-85, T.S, India
Padmaja_jntuh@jntuh.ac.in

V Janaki
Dept. of Computer Science & Engg
Vaagdevi College of Engineering
Warangal-06, T.S, India
janakivinjamuri.cse@gmail.com

Sai Mohan Gajapaka
Dept. of Information Technology
VNR VJIET
Hyderabad-90, T.S, India
sai.mohan1607@gmail.com

Shilhora Akshay Patel
Dept. of Information Technology
VNR VJIET
Hyderabad-90, T.S, India
shilhora.akshay333@gmail.com

Lalith Bharadwaj.B
Dept. of Information Technology
VNR VJIET
Hyderabad-90, T.S, India
lalithbharadwaj313@gmail.com

***Abstract*—Coronary heart disease (CHD) is one of the leading cause of an increase in the mortality rate. Various factors are influencing to estimate the presence of CHD in an individual biologically. We made inferences with these influencing factors and their peculiarity by analyzing with a various set of algorithms which help in obtaining a precisive decision support system for the presence of CHD in an individual. We considered a set of linear models (Logistic Regression, Linear Discriminant Analysis), Naïve Bayes, Classification and Regression Trees (CART), Support Vector Machines (SVM). K-Nearest Neighbor (k-NN) for k=1 to 21. Next, we have considered a set of ensemble models. Furtherly, we computed a Multi-Layer Perceptron (MLP) and a Deep Neural Network to evaluate the performance through a deep learning approach. So, with this analysis, we found a linear model (Logistic Regression) tend to give on par results both in the case of Cleveland as well as Framingham dataset. This analysis led to design an intuitive approach for CHD classification and would give insights on how to use them in the medical research field.**

***Index Terms*—Coronary Heart Disease, Deep Learning, Disease Prediction, Feature Extraction.**

I. Introduction

Over history, heart diseases are dated back to 1230 B.C. when we had no cure for heart diseases but, over the time with the advancement of medical diagnosis heart diseases were able to detect and cure. It was during the 15th century when it was first discovered that heart pumps blood to the vital parts of the body from then we came up with diverse clinical methods. In the 20th century, Coronary Heart Disease (CHD) was considered as a potential problem and led to a rigorous study by researchers. R .D. Thomas et al., [19] has done a comprehensive study on heart diseases at Framingham [20] [22]. Another research has been done at the Cleveland clinic on CHD [15]. These studies lead computational methods, insights [14] and various analysis [36] [16] has been done by researchers to predict CHD precisely by the help of this studies we discovered that there are many other individual factors which tend to CHD such as pulse pressure [17] and hypertension [21] not only on the features from the above studies, many researches have also worked on how personal habits would affect the heart [13][18]. Now we are in 21 century and still, CHD is a threat to mortals. There are many research works which have been done on CHD but the prime focus of these works were on generic model construction over data by applying various data mining techniques. Feature selection was used in most of the research works for building a model to give more accurate results [2][5]. Feature selection might be useful in various research fields but when we are dealing with medical data every feature has its own importance. So, it is best to avoid feature selection and consider every attribute to classify a heart disease. The reason why one should avoid feature selection is that there are many uncertain possibilities which might lead to CHD. So, by avoiding feature selection it would make our model learn these kinds of uncertain possibilities that could save a person's life.

Data set description

We considered two variants which consist of CHD and the attribute description is given in the Table (I), (II).

1) Framingham dataset contains 15 features and 4,239 records. The data is updated with time this gives an overview of CHD concerning time is depicted at Table(I).
2) Cleveland dataset contains 13 features and 303 records which are used to classify CHD. The features are depicted at Table(II)

II. Previous Work

We've discussed the attributes that influence in predicting the presence of CHD and the biological significance to use them as features for the appropriate classification. Previously there was a lot of research done in the estimation of CHD using classical data mining techniques as well as designing neural networks. Sellapan et al., [11] compared Naïve Bayes,

TABLE I: *Framingham dataset [15].*

Attribute	Description	Type
Gender	Male, Female (1,0)	Categorical
Age_reported	Age reported during examination	Categorical
Education	1 = High School : 2 = GED 3 = Unspecified College or Vocational School 4 = College	Categorical
Current_Smoker	0 = no; 1 = yes	Categorical
Cigs_Per_Day	0 = Not BP medications; smoked per day (estimated average)	Numeric
BPMeds	0 = Not BP medications 1 = Taking BP medications	Categorical
prevalentStroke	1 or 0	Categorical
prevalentHyp	1 or 0	Categorical
Diabetes	1 or 0	Categorical
Total_Chol	$\frac{mg}{dL}$	Numeric
Sys_BP	mm of Hg	Decimal
Dia_BP	mm of Hg	Decimal
B.M.I	BMI reported as: Weight($\frac{KG}{m^2}$)	Decimal
Heart_Rate	$\frac{Beats}{Minutes}$ (Ventricular Beats)	Numeric
Glucose_level	$\frac{mg}{dL}$	Numeric
TenYear_CHD (target)	1 or 0	Categorical

TABLE II: *: Cleveland Data [35].*

Attribute	Description	Type
Age of person	Age(years)	Numeric
gender	Male, Female(1,0)	Categorical
Chest_pain	Chest Pain Variants(angina) : Val 1: Normal Val 2: Abnormal Val 3: non-anginal-pain Val 4: Asymptomatic	Categorical
Trest_bps	Resting B.P (in mm Hg)	Numeric
Choletrol	Serum Cholestoral (mg/dl)	Numeric
fbs	(Fasting Blood Sugar>120 mg/dl) (1 = True; 0 = False)	Categorical
Rest_ecg	Resting ECG results: Val 0: Normal condition Val 1: ST-T wave abnormality Val 2: Abnormality	Categorical
Tha_lach	Max Heart rate	Numeric
exang	exercise induced angina (1 = yes; 0 = no)	Categorical
Old_peak	ST depression induced exercise relative to rest	Decimal
Slope	slope of peak exercise ST segment: Val 1: sloping upwards Val 2: flattening Value 3: sloping downwards	Categorical
CA	No. of majority of vessels (0-3) coloured with the help of Flourosopy	Categorical
Thal	3 = Normal; 6 = static defect; 7 = inverse defect	Categorical
Target	1 or 0	Categorical

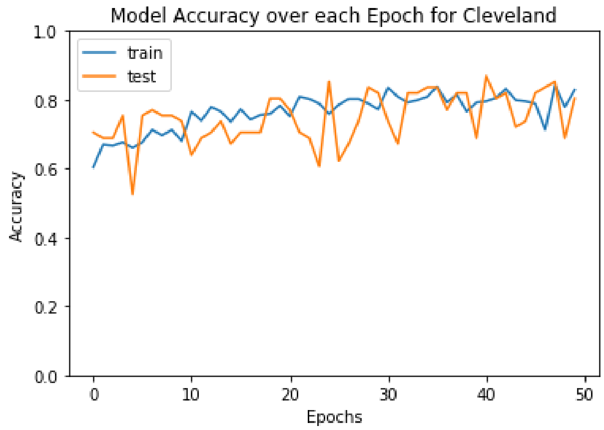
decision trees and neural networks and secured the highest accuracy of 86.12% for decision trees with equivalent train & test. Jyoti.S et al., [7] secured an accuracy score of 92.2% with decision trees by reducing dimensions of the feature set. Charitali.S.D et al., [10] performed an analysis on Cleveland and Statlog.H.D data [15] [37] by additionally adding features and evaluated the performance and demonstrated the increment before and after placing the feature set. Rajeswari.K et al ., [2] applied a brute force approach for feature selection and additionally used genetic search attribute selection mechanism and computed accuracy scores. M.A.Jabbar et al., [4] collected data manually from regional hospitals and attained the highest accuracy for the proposed algorithm which combines a genetic algorithm with K-NN.

D.S.Medhekar et al., [8] trained the complete data and then performed test methodology within the samples using Naïve Bayes and attained the highest accuracy score for their model 89.56%. S.Bashir et al., [22] compared typical data mining classifiers and proposed an ensemble technique(stacking various classifiers) which in turn attained the highest accuracy score of 81.82% and evaluated other classification metrics. S.Pouriyeh et al., [6] compared most of the existing data mining classifiers, neural networks and computed results for various metrics and obtained the highest accuracy scores 84.15% by ensemble stacking technique using 10 fold cross-validation strategy. In Further, some researchers used association mining rules and clustering algorithms to draw insightful results [34] [3] [1] for predicting the presence of CHD. We observed some were not able to generalize for newly observed data [8] [3] [10]. Some authors picked out features randomly and some specific attributes for better performance [7] [2]. But we cannot pick out features for better performance of our statistical model as they have biological significance in clinical decision making.

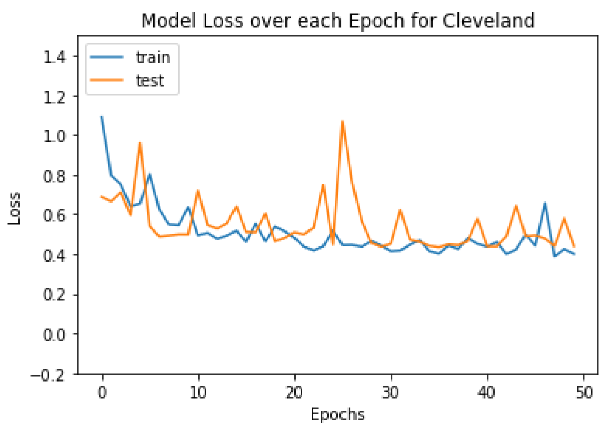
The revival of deep learning has influenced many researchers with that efficacy we designed a feed-forward neural network for predicting the risk for a individual to have CHD. Even with tremendous usage of Convolutional-Neural-Networks(CNN) in the domain of vision, text and speech it did not tend to perform precise for models consisting of attributes (a.k.a Datasets) [41]. In this paper, we provide a complete survey of statistical models (Machine Learning models) used for predicting the presence of CHD from various attributes by overcoming the stated loopholes and additionally designing a 5-layered neural-network for classification of CHD for both the datasets. We compare and contrast the proposed deep-learning-framework [42] [43] with statistical models to illustrate performance in designing a diagnostic-framework.

III. METHODOLOGY

By considering the above-proposed works we construct a new framework for an accurate diagnostic model to overcome above-mentioned loops. We considered a various set of classical data mining classifiers, KNN, decision trees, boosting and bagging of trees and ensemble models. We proposed two variants of neural network architectures for Cleveland as well as Framingham data. Firstly we clean data by preparing it for further processing. Here we do categorical encoding and perform imputation for missing values in data [12]. We performed imputation with missXGBoost by G. Madhu et al., [23] and missForest by D J. Stekhoven et al., [24]. We observed missXGBoost was computationally expensive but has

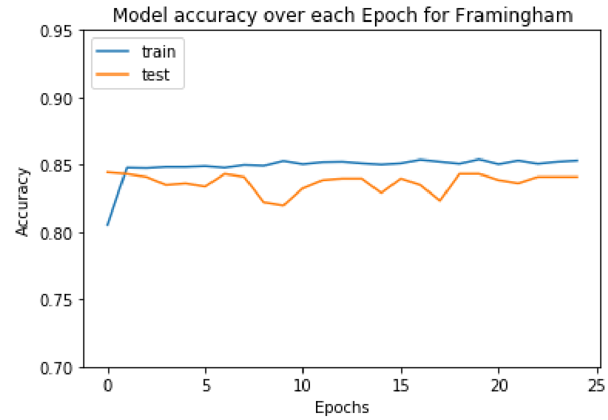


(a) *Accuracy score convergence analysis for individal epoch. Even without use of dropout[40] model tend to regularize well with good convergence.*

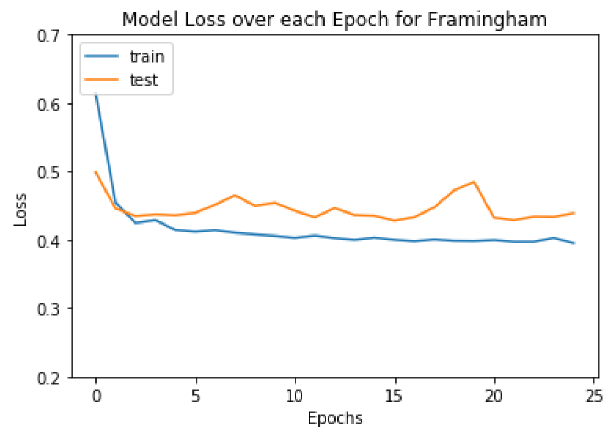


(b) *In plot, the decay for loss is consistent and a little hump between* 20^{th} *and* 30^{th} *epoch but, afterwards loss decayed consistently.*

Fig. 1: *Cleveland accuracy score and loss for proposed Neural Network at each epoch (iterating up to 50 epochs. There can be chance of overfitting and hence training is halted at* 50^{th} *epoch.)*



(a) *Accuracy score & convergence analysis for each epoch.It is observed that convergence was obtained fast and saturated (compared to Cleveland).*



(b) *In plot, the decay for loss is consistent after a little decay, i.e. between* 1^{st} *&* 5^{th} *epoch.*

Fig. 2: *Framingham accuracy score and loss for proposed Neural Network at each epoch (iterating up to 25 epochs))*

better performance. As to perform a rapid decision making, we impute with missForest. It performed 5 iterations to complete imputation on Framingham data with 15.219% missing information. Cleveland did not contain any missing values. Furtherly, we did not attempt for dimensionality reduction as it involves the reduction of feature-set, in turn, makes improper decision making in our medical diagnostic model.

We used a set of algorithms mentioned in the Table (VI). The deep neural networks proposed for each particular data. In both of case, we proposed a 5 layered neural network with variations in batch size, hidden neurons and training iterations[46] [47]. Firstly, for Cleveland data, the first layer consists of the input layer and the next 3 hidden layers consist of 32, 16, 8 neurons sequentially at each hidden layer. The final layer consists of 1 neuron with a linear sigmoid activation. As the size of Cleveland data is small (303 records) we considered a small batch size of 5 and trained for 50 epochs. In Framingham data, the first and last layers are the same as that of Cleveland but the hidden layer neurons vary as 15, 10, and 5 sequentially. The hidden layers were provided with non-linear ReLU [9] activations in both the data. As they have a good number of records (4238) we consider a batch size of 10 and trained for 25 epochs. We used optimizer as Adam [33] and fixed a learning rate of 0.01 at each level. We evaluated various metrics for each of the neural networks and plotted them Fig. 1 & Fig. 2. This led to understand and asses models performance at each level i.e. constricting not only machine learning domain but even deep learning [45]. The use of Batch-Normalization aided the performance of model by generalizing on test samples which were fed into the network [44]. This reduced the effort of using dropout for generic generalization without additional tuning[40].

TABLE III: *Results generated for Framingham data for algorithms mentioned (Table(I)).*

Models	Accuracy	F1-Score	Precision	Recall	MSE	AUC-ROC
LR	84.969 ± 1.438	5.647 ± 3.532	71.190 ± 27.473	3.012 ± 1.988	15.031 ± 1.438	64.836 ± 2.542
LDA	85.182 ± 1.708	18.064 ± 5.207	58.011 ± 16.329	10.858 ± 3.472	14.818 ± 1.708	72.420 ± 1.222
CART	75.106 ± 2.159	23.228 ± 3.703	20.909 ± 4.508	24.655 ± 3.701	25.625 ± 1.453	54.467 ± 1.823
NB	82.350 ± 2.210	26.300 ± 5.204	35.910 ± 4.905	20.920 ± 5.124	17.650 ± 2.210	71.241 ± 1.975
SVM	84.781 ± 1.488	0.548 ± 1.644	5.000 ± 15.000	0.290 ± 0.870	15.219 ± 1.488	61.734 ± 3.137
AB	85.015 ± 1.003	17.467 ± 4.792	53.561 ± 11.367	10.813 ± 3.871	14.985 ± 1.003	70.403 ± 2.396
GBC	84.956 ± 1.384	16.312 ± 5.151	50.208 ± 15.968	10.447 ± 4.212	14.985 ± 1.426	70.868 ± 2.462
RF	85.103 ± 1.167	11.801 ± 6.240	58.889 ± 20.403	6.803 ± 2.797	14.749 ± 1.286	69.703 ± 4.012
ET	84.808 ± 1.074	11.530 ± 2.733	51.376 ± 28.182	5.876 ± 2.740	15.162 ± 1.189	69.746 ± 3.830
BC	83.392 ± 1.600	15.957 ± 5.373	42.648 ± 10.418	11.218 ± 2.855	16.165 ± 1.291	65.866 ± 3.287
XGBC	85.103 ± 1.404	12.310 ± 5.086	55.789 ± 11.217	7.107 ± 3.314	14.897 ± 1.404	71.456 ± 3.146
XGBRFC	84.867 ± 1.173	1.433 ± 1.760	31.667 ± 45.000	0.751 ± 0.927	15.133 ± 1.173	70.745 ± 1.918
KNN-3	82.019 ± 1.857	17.140 ± 2.824	29.235 ± 5.666	12.274 ± 2.279	17.981 ± 1.857	58.898 ± 2.410
KNN-7	84.143 ± 1.427	13.052 ± 3.146	39.696 ± 7.310	7.929 ± 2.300	15.857 ± 1.427	63.307 ± 1.906
KNN-13	84.946 ± 1.291	8.242 ± 5.196	53.024 ± 24.843	4.559 ± 3.090	15.054 ± 1.291	64.796 ± 2.807
KNN-17	85.064 ± 1.473	7.156 ± 4.438	63.121 ± 30.385	3.866 ± 2.524	14.936 ± 1.473	65.795 ± 2.344
KNN-21	84.922 ± 1.585	5.652 ± 3.701	60.083 ± 30.112	3.021 ± 2.019	15.078 ± 1.585	65.699 ± 2.205

TABLE IV: *Results generated for Cleveland data for algorithms mentioned (Table(II)).*

Models	Accuracy	F1-Score	Precision	Recall	MSE	AUC-ROC
LR	82.839 ± 6.602	84.928 ± 5.732	81.762 ± 7.618	89.007 ± 7.686	17.161 ± 6.602	90.357 ± 4.692
LDA	82.161 ± 5.842	84.428 ± 5.222	80.500 ±6.699	89.559 ± 8.400	17.839 ± 5.842	90.671 ± 4.672
CART	78.484 ± 5.978	78.755 ± 5.963	79.305 ±6.168	76.250 ± 7.758	22.495 ± 5.858	77.548 ± 6.479
NB	80.505 ± 6.615	82.158 ± 6.371	82.091 ±7.795	82.941 ± 8.350	19.495 ± 6.615	88.974 ± 4.892
SVM	66.043 ± 8.165	72.966 ± 6.990	64.424 ±6.386	84.743 ± 10.412	33.957 ± 8.165	74.488 ± 9.697
AB	81.129 ± 7.384	83.156 ± 6.432	81.973 ±9.961	85.331 ± 7.530	18.871 ± 7.384	87.129 ± 6.785
GBC	80.140 ± 9.741	81.344 ± 10.486	80.922 ±8.637	83.971 ± 15.669	19.527 ± 9.081	89.014 ±5.211
RF	83.484 ± 5.183	84.974 ± 5.615	82.467 ±6.780	86.581 ± 8.160	15.849 ± 6.098	90.238 ±3.653
ET	83.484 ± 5.142	83.707 ± 5.036	84.690 ±7.605	84.743 ± 10.455	16.516 ± 4.950	91.841 ±4.550
BC	79.484 ± 9.419	79.880 ± 8.203	79.889 ±8.435	80.478 ± 9.218	20.849 ± 6.610	87.774 ±6.673
XGBC	80.161 ± 6.741	82.105 ± 6.661	80.041 ±6.377	85.331 ± 11.111	19.839 ± 6.741	89.803 ±4.331
XGBRFC	82.796 ± 5.433	85.179 ± 4.427	80.868 ±7.044	90.809 ± 6.945	17.204 ± 5.433	89.208 ±3.887
KNN-3	62.720 ± 7.684	66.063 ± 7.558	65.615 ±7.417	67.316 ± 10.996	37.280 ± 7.684	64.565 ±9.810
KNN-7	67.344 ± 7.399	70.616 ± 7.494	69.078 ±6.735	72.684 ± 9.703	32.656 ± 7.399	69.174 ±8.888
KNN-13	65.731 ± 8.665	69.210 ± 8.585	67.219 ±7.498	71.471 ± 10.279	34.269 ± 8.665	68.766 ±9.194
KNN-17	66.344 ± 7.788	69.536 ± 8.570	67.725 ±6.550	71.985 ± 12.203	33.656 ± 7.788	70.248 ±9.674
KNN-21	67.333 ± 7.233	70.877 ± 7.878	68.002 ±5.075	74.449 ± 11.899	32.667 ± 7.233	71.057 ±9.522

TABLE V: *Performance measure of designed deep learning models for Cleveland and Framingham*

Models	Accuracy	Loss	Precision	Recall	F1-score
Neural Network for Cleveland	74.754 ± 7.361	0.54984 ± 0.13106	68.071 ± 10.498	72.573 ± 19.981	67.471 ± 13.210
Neural Network for Framingham	84.094 ± 6.681	0.44146 ± 0.01036	71.881 ± 11.681	76.971 ± 13.961	70.751 ± 14.211

TABLE VI: *Performance measure of designed deep learning models for Cleveland and Framingham*

Short-hand	Full-Form
LR	Logistic Regression
LDA	Linear Discriminant Analysis
CART[32]	Classification and Regression Trees
NB	Naïve Bayes Classifier
SVM[27]	Support Vector Machines
AB[31]	Ada-Boost Classifier
GBC[29]	Gradient Boosting Classifier
RF[26]	Random Forest Classifier
ET[25]	Extra Trees Classifier
BC[30]	Bagging Classifier
XGBC[28]	eXtreme Gradient Boosting Classifier
XGBRFC[28][26]	XGBoost + RandomForest Classifier
k-NN	k-Nearest Neighbour

IV. Results

We computed and evaluated results with various classification metrics such as accuracy, precision, recall, f1-score, mean squared error (MSE) and AUC-ROC. As each metric has its dignity in determining model performance. We apply 10-fold cross-validation for each of the metrics considered for machine learning model construction. When performed 10-fold cross-validation, at each fold we evaluate the metric followed by aggregating them (Mean) and then finding the standard deviation accordingly. Next while training neural networks we divided the complete data into train and test portions. Where, the test portion consists of 20% of the full-length data by randomly shuffling the samples. Further, we compared the k-NN model for various k values ranging from 1 to 21 and accuracy scores for these are depicted (Fig. 3).But, to be known evaluating multiple folds for cross validation takes a lot of computational cost in case of deep learning models.

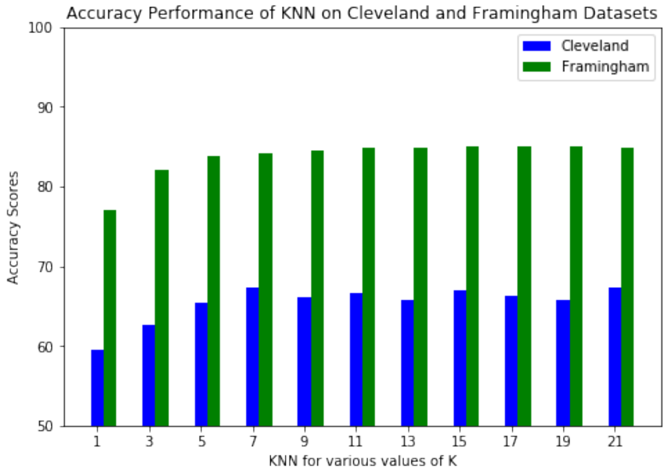


Fig. 3: *Comparing accuracy scores for k-NN model with k value ranging from 1 to 21 for both the datasets*

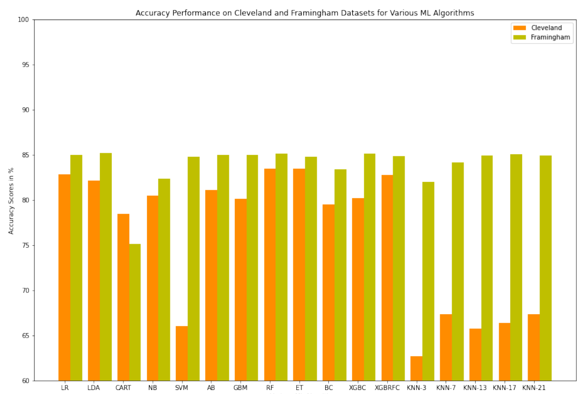


Fig. 4: *Comparing accuracy scores for various machine learning models for both the datasets*

In order to evaluate such sophisticated models we choose to split data into train and test portions. The above mentioned results are generated with Keras library where, backend using Tensor-flow [39] for computation.

V. Conclusion

We collate various machine learning classifiers and a deep neural network with insightful results. We observed that linear models (Logistic Regression (LR) and Linear Discriminant Analysis (LDA)) tend to perform well in both the data models. Trees, boosting, bagging and ensemble models competed with LDA and LR with similar accuracy scores but had a high MSE. We observed that KNN tended to perform poorly in the presence of fewer instances and acquired the least accuracy score for Cleveland data. But in the case of Framingham data, CART (Classification and Regression Trees) tend to perform poorly. As neural networks are widely used in present-day situations due their high computational ability and a desire for higher accuracy scores but, while determining performance we cannot compare our machine learning diagnostic model to that of deep learning framework because of two main reasons. First is their black box and the other is the validation procedure is different.

Even with the many state-of-the-art results in various domains deep learning can be limiting in designing diagnostic model with sensitive attributed data. So, while constructing a medical diagnostic model it is a prime factor to consider every attribute which influences clinical decision making and constructs a statistical model preferably a linear model which would lead to insightful clinical decision making and provides a precise medical diagnostic model. In the future, with the use of current technology an real-world application is designed to favour the one who is in need.

Acknowledgement:

This work was supported by a project grant from JNTUH-TEQIP-III with the proceedings no: JNTUH/TEQIP-III/CRS/2019/CSE/08.

References

[1] Nilashi M, bin Ibrahim O, Ahmadi H, Shahmoradi L. An analytical method for diseases prediction using machine learning techniques. Computers & Chemical Engineering. 2017 Nov 2;106:212-23.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] Rajeswari K, Vaithiyanathan V, Pede SV. Feature selection for classification in medical data mining. International Journal of Emerging Trends and Technology in Computer Science (IJETTCS). 2013 Apr;2(2):492-7R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[3] Nahar J, Imam T, Tickle KS, Chen YP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Systems with Applications. 2013 Mar 1;40(4):1086-93.

[4] Jabbar MA, Deekshatulu BL, Chandra P. Classification of heart disease using k-nearest neighbor and genetic algorithm. arXiv preprint arXiv:1508.02061. 2015 May 7.

[5] Nahar J, Imam T, Tickle KS, Chen YP. Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. Expert Systems with Applications. 2013 Jan 1;40(1):96-104.

[6] S. Pouriyeh, S. Vahid, G. Sannino, G. De Pietro, H. Arabnia and J. Gutierrez, "A comprehensive investigation and comparison of Machine Learning Techniques in the domain of heart disease," 2017 IEEE Symposium on Computers and Communications (ISCC), Heraklion, 2017, pp. 204-207.

[7] Soni J, Ansari U, Sharma D, Soni S. Predictive data mining for medical diagnosis: An overview of heart disease prediction. International Journal of Computer Applications. 2011 Mar 8;17(8):43-8.

[8] Medhekar DS, Bote MP, Deshmukh SD. Heart disease prediction system using naive Bayes. Int. J. Enhanced Res. Sci. Technol. Eng. 2013 Mar;2(3).

[9] Nair V, Hinton GE. Rectified linear units improve restricted boltzmann machines. InProceedings of the 27th international conference on machine learning (ICML-10) 2010 (pp. 807-814).

[10] Dangare CS, Apte SS. Improved study of heart disease prediction system using data mining classification techniques. International Journal of Computer Applications. 2012 Jun;47(10):44-8.

[11] Palaniappan S, Awang R. Intelligent heart disease prediction system using data mining techniques. In2008 IEEE/ACS international conference on computer systems and applications 2008 Mar 31 (pp. 108-115). IEEE.

[12] McEwan P, Williams JE, Griffiths JD, Bagust A, Peters JR, Hopkinson P, Currie CJ. Evaluating the performance of the Framingham risk equations in a population with diabetes. Diabetic medicine. 2004 Apr;21(4):318-23.

[13] FRIEDMAN LA, Kimball AW. Coronary heart disease mortality and alcohol consumption in Framingham. American journal of epidemiology. 1986 Sep 1;124(3):481-9.

[14] [14] D'Agostino Sr RB, Pencina MJ, Massaro JM, Coady S. Cardiovascular disease risk assessment: insights from Framingham. Global heart. 2013 Mar 1;8(1):11-23
[15] https://archive.ics.uci.edu/ml/datasets/Heart+Disease
[16] [16] D'Agostino RB, Lee ML, Belanger AJ, Cupples LA, Anderson K, Kannel WB. Relation of pooled logistic regression to time-dependent Cox regression analysis: the Framingham Heart Study. Statistics in medicine. 1990 Dec;9(12):1501-15.
[17] Franklin SS, Khan SA, Wong ND, Larson MG, Levy D. Is pulse pressure useful in predicting risk for coronary heart disease? The Framingham Heart Study. Circulation. 1999 Jul 27;100(4):354-60.
[18] Benjamin EJ, Wolf PA, D'Agostino RB, Silbershatz H, Kannel WB, Levy D. Impact of atrial fibrillation on the risk of death: the Framingham Heart Study. Circulation. 1998 Sep 8;98(10):946-52.
[19] Dawber TR, Meadors GF, Moore Jr FE. Epidemiological approaches to heart disease: the Framingham Study. American Journal of Public Health and the Nations Health. 1951 Mar;41(3):279-86.
[20] Detrano R, Janosi A, Steinbrunn W, Pfisterer M, Schmid JJ, Sandhu S, Guppy KH, Lee S, Froelicher V. International application of a new probability algorithm for the diagnosis of coronary artery disease. The American journal of cardiology. 1989 Aug 1;64(5):304-10.
[21] Franklin SS, Wong ND. Hypertension and cardiovascular disease: contributions of the Framingham Heart Study. Global heart. 2013 Mar 1;8(1):49-57.
[22] Bashir S, Qamar U, Javed MY. An ensemble-based decision support framework for intelligent heart disease diagnosis. International Conference on Information Society (i-Society 2014) 2014 Nov 10 (pp. 259-264). IEEE.
[23] G. Madhu, B. L. Bharadwaj, G. Nagachandrika and K. S. Vardhan, "A Novel Algorithm for Missing Data Imputation on Machine Learning," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 173-177.
[24] Daniel J. Stekhoven, Peter Bühlmann, MissForest—non-parametric missing value imputation for mixed-type data, Bioinformatics, Volume 28, Issue 1, 1 January 2012, Pages 112–118, https://doi.org/10.1093/bioinformatics/btr597.
[25] [25] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine learning. 2006 Apr 1;63(1):3-42.
[26] Ho TK. Random decision forests. In Proceedings of 3rd international conference on document analysis and recognition 1995 Aug 14 (Vol. 1, pp. 278-282). IEEE.
[27] Cortes C, Vapnik V. Support-vector networks. Machine learning. 1995 Sep 1;20(3):273-97.
[28] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug 13 (pp. 785-794).
[29] Friedman JH. Greedy function approximation: a gradient boosting machine. Annals of statistics. 2001 Oct 1:1189-232.
[30] Breiman L. Bagging predictors. Machine learning. 1996 Aug 1;24(2):123-40.
[31] Quinlan JR. Bagging, boosting, and C4. 5. In AAAI/IAAI, Vol. 1 1996 Aug 4 (pp. 725-730).
[32] De'ath G, Fabricius KE. Classification and regression trees: a powerful yet simple technique for ecological data analysis. Ecology. 2000 Nov;81(11):3178-92.
[33] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014 Dec 22.
[34] Jabbar MA, Chandra P, Deekshatulu BL. Cluster based association rule mining for heart attack prediction. Journal of Theoretical and Applied Information Technology. 2011 Oct 31;32(2):196-201.
[35] https://www.kaggle.com/amanajmera1/framingham-heart-study-dataset.
[36] Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. Journal of chronic diseases. 1967 Jul 1;20(7):511-24.
[37] https://archive.ics.uci.edu/ml/datasets/statlog+(heart).
[38] Pedregosa F, Varoquaux, Ga"el, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12(Oct):2825–30.
[39] Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, Devin M, Ghemawat S, Irving G, Isard M, Kudlur M. Tensorflow: A system for large-scale machine learning. In12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) 2016 (pp.265-283).
[40] Srivastava, Nitish, et al. "Dropout: a simple way to prevent neural networks from overfitting." The journal of machine learning research 15.1 (2014): 1929-1958.
[41] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012.
[42] Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. Deep learning. MIT press, 2016.
[43] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." nature 521.7553 (2015): 436-444.
[44] Ioffe, Sergey, and Christian Szegedy. "Batch normalization: Accelerating deep network training by reducing internal covariate shift." arXiv preprint arXiv:1502.03167 (2015).
[45] Bengio, Yoshua, Aaron Courville, and Pascal Vincent. "Representation learning: A review and new perspectives." IEEE transactions on pattern analysis and machine intelligence 35.8 (2013): 1798-1828.
[46] Arora, Sanjeev, et al. "Fine-grained analysis of optimization and generalization for overparameterized two-layer neural networks." arXiv preprint arXiv:1901.08584 (2019).
[47] Allen-Zhu, Zeyuan, Yuanzhi Li, and Yingyu Liang. "Learning and generalization in overparameterized neural networks, going beyond two layers." Advances in neural information processing systems. 2019.