Performance Analysis of Machine Learning Algorithms for Prediction of Liver Disease

Vinayak Singh
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar, India
vinayaksooryavanshi@gmail.com

Mahendra Kumar Gourisaria
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar, India
mkgourisaria2010@gmail.com

Himansu Das
School of Computer Engineering
KIIT Deemed to be University
Bhubaneswar, India
das.himansu2007@gmail.com

Abstract— Liver diseases like fatty liver disease, chronic active hepatitis, and cirrhosis are the major cause of mortality in India. Alcohol consumption, inhalation of harmful toxic gases, improper consumption of contaminated pickles, drugs, and foods are the major cause of diseases in the liver. Diagnosis of liver disease needs high accuracy and precise results for predicting whether a person is suffering from liver disease or not. Major disastrous repercussions can be the result of minor errors in the diagnosis of liver diseases. The major goal of this paper is for the detection of liver disease at right time and helping the doctors and combating the increasing number of cases. In this paper, we implemented various machine learning techniques like logistic regression, KNN, XG-Boost, SVM, Gaussian NB, Random forest, Decision tree, Gradient Boosting, CatBoost, AdaBoost, and LightGBM on selected features from the dataset for predicting liver disease and it was found that Random Forest performed best among all the technique and gained high accuracy and performed outstandingly in all metric evaluations.

Keywords— machine learning, liver disease, data mining,

I. INTRODUCTION

This Liver disease is the most dangerous disease and the major reason behind mortality among human beings. The liver is among the largest and most essential organs of our body. Being one of the exocrine glands liver secrets bile into the intestine. Observing the last two decades, liver disease is among most top 12 dangerous causes of death and is among the top reason for death cause in adults who are in the age group from 45-54 [1]. Fatty liver disease is caused due to extra deposit of fat in the liver cells. Steatosis is the process in which there is the deposition of fat in our liver cells and steatosis can be caused by consumption of alcohol, contaminated food, pickles, etc., and also by metabolic syndrome [2]. Noncholestatic cirrhosis and hepatocellular carcinoma are majorly caused by fatty liver diseases [3]. Biopsy has been used for distinguishing liver patients from other normal persons and this method is costly and can give sampling errors for diagnosing fatty liver disease (FLD).

Nowadays ultrasonography helps to diagnose the FLD with high accuracy but while finding the accuracy is highly dependent on the operator [4]. Restricted Boltzmann Machines i.e. Unsupervised neural networks were used for creating a time-aware recommendation system to suggest the movies to the users [5]. Machine learning plays an important role in every field nowadays from detection of skin cancer [6] to detection of sentiments of tweets and classifying them into positive and negative via NLP [7]. Machine learning [8] and generative modeling [9] are contributing major hands in various fields like medical fields and recommendation systems. As deep learning [10] is the backbone of machine learning so both of them are playing a crucial role in every field like COVID19 detection by using CNN ranking approach [11], text recognition [12]. Various machine learning algorithms like Random Forest, Decision trees, etc. were used for detecting credit card frauds [13]. In this paper, we have implemented various machine learning algorithms like Gaussian NB, Random Forest, Logistic Regression, KNN, Support Vector Machine, Decision tree, Gradient Boosting, CatBoost, AdaBoost, LightGBM, and XG Boost on selected features from the correlation matrix obtained from the liver dataset and we have compared all of them in terms of Sensitivity, Specificity, Precision, Matthews Correlation Coefficient, F1 score, Accuracy, Confusion matrix, and FNR, FDR, NPV and selected the best suitable machine learning algorithm for predicting the liver disease. The rest of the paper is organized as follows:-2. Related Work, 3. Data Preparation 4. Techniques 5. Evaluation and Results, 6.Conclusion and Future Work. By finding the best algorithm and implementing it we can save computational time and gain high accuracy for diagnosing liver disease in patients.

II. LITERATURE REVIEW

Diseases like cancer, tuberculosis, pneumonia [26], and other dangerous disease are getting a major contribution from machine learning deep learning fields. Andrade et al. [14] used extracted features of ultrasound images for diagnosing liver steatosis by comparing the performance of three different classifiers. Automatic ROI selection and hierarchical method were used by Ojwimehr et al. [15] for classifying normal persons from persons who are suffering from fatty liver diseases like steatosis, fibrosis, and cirrhosis. Support vector machine was used by Li et al. [16] for diagnosing and analyzing B-mode ultrasonic images of FLD with different textures. The Bayesian framework was used by Ribeiro et al. [17] for extracting features and for classifying normal liver ultrasonic images from fatty liver ultrasonic images where the anatomic and echogenic information of ultrasonic images were used.

The SVM classifier was used by Subramanya et al. [18] in 2014 for classifying liver images which were from the US liver image dataset and they achieved an accuracy of 84.9%. An accuracy of 81.2% was achieved by Ma et al. [19] in 2015 by developing a signal processing approach that was based on kurtosis scanning and grading the fatty liver diseases and this was implemented on the US liver image dataset.

Backpropagation Neural Network (BPNN) was used by Saba et al. [20] in 2016 where BPNN had 10 hidden layers and 128 features were used for extraction on US liver images by using different feature extraction algorithm total 6 different types of algorithms was used and BPNN gained accuracy of 97.6%. Vijayrani et al. [21] used the ILPD dataset which was collected from UCI and they implemented Support vector machine and Naïve Bayes classification algorithms for classification where they used MATLAB for implementation and according to the result SVM performed good gained high accuracy than Naïve Bayes which was 79.66% but Naïve Bayes took less time for training when compared to SVM. In 2014 Dhamodharan et al. [22] used the Naïve Bayes and F1 tree algorithm for the prediction of liver diseases where three main liver diseases were classified as Liver Cancer, Cirrhosis, and Hepatitis where accuracy was used for selecting the best suitable algorithm where Naïve Bayes performed better than other algorithms. In medical field machine contributes an important hand like ECG classification [23], distinguishing between COVID 19 and Pneumonia [24], Parkinson disease detection [25] and malaria detection [30], heart disease prediction[29]. There are many contributions in every field as nowadays machine learning is serving as the backbone in all the niche areas. All recent work was satisfactory but they did not have the full comparison between all the algorithms. Few of the works lacked in the dataset. Liver diseases are a major reason behind mortality among humans. FLD needs high accuracy for detection as ultrasonic classification is costly too.

III. Data Preparation

This Data Preparation will give a proper description regarding the dataset we have used and about the various tuning before implementing the various algorithms. Dataset had various missing values, unnecessary information that is not essential for training and testing. So for better performance, we have performed various tunings and made the dataset up to the point for building a good connection with the algorithms for giving good and accurate results.

A. *Dataset Used*

Dataset was collected from UCI ML Repository where it was deployed by Lichman, M. in 2013 and the dataset had a record of 516 persons that were collected from the North-East part of Andhra Pradesh, India [31]. The data set consisted of 583 records where 441 were male and 142 were female. Fig. 1 shows the distribution between the liver patient and the normal person.

Dataset had 11 column attributes which were Age', 'Gender', 'Total Bilirubin', 'Direct Bilirubin', 'Alkaline Phosphotase', 'Alamine Aminotransferase', 'Aspartate Aminotransferase', 'Total Protein', 'Albumin', 'Albumin and Globulin Ratio', 'Dataset'.

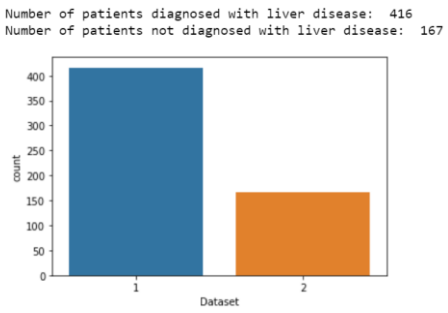

Fig. 1. The distribution between the liver patient and normal person

B. *Data Exploration*

This is the most essential part of building a good machine learning model. For finding the errors, missing values, analyzing the dataset and extracting the features, and finding the most correlated attributes the data exploration part is necessary so that our algorithms can build up a good connection with the dataset. Dataset needs to be cleaned up before applying the machine learning algorithm for good results.

1) *Missing Values, Correlation Matrix for Model*

When data exploration was done it was found that 'Albumin and Globulin Ratio' had 4 missing values. Record numbers 209, 241, 253, and 312 had missing values. So for balancing up the dataset we replaced the missing values with the mean of 'Albumin and Globulin Ratio' which was found to be 0.9470. No outliers and Skewness were present in the dataset. For further implementation of algorithms, the correlational matrix was used for extracting the features shown in Fig. 2 and after watching over the matrix it was noticed that two columns that were 'Gender' and 'Dataset' were dropped from the dataset before implementation of the machine learning algorithm.

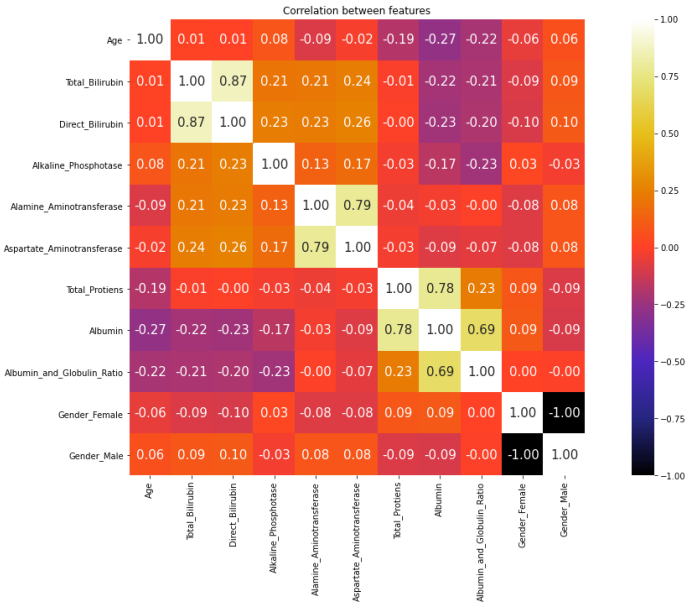

Fig. 2. Correlation Matrix

C. *Dataset Splitting and Software and Hardware Used*

Dataset was distributed in two parts where 70% was selected for the training set and the rest 30% was for testing set and for the initialization of the internal random number generator we had a value of 101 for the random state.

All of the machine learning implementations were done and analyzed on Python 3 with the Keras and TensorFlow were used with a jupyter notebook. The hardware workstation is with Intel i5 8th generation and 16 GB RAM.

IV. TECHNOLOGY USED

Most common algorithms like Random Forest, Boosting technique like XGBoost, Logistic Regression, Gaussian Naïve Bayes, KNN, Support Vector Machine, Gradient Boosting, CatBoost, AdaBoost, LightGBM, and Decision Tree were used for training and testing of the dataset and they have a wide range of application on various recent research works like breast cancer prediction [27], real estate cost prediction [28].

Logistic Regression is a regression classification method that uses a logistic function and which is used for classifying categorical variables. They are statistical models and the logistic function we implement of sigmoid function on linear regression turns it into logistic regression. (1) shows the logistic function.

$$a_0 + a_1 x = \ln(\frac{y}{1-y}) \tag{1}$$

XGBoost uses gradient boosting. It is a boosting technique that is based on a decision tree ensemble ML algorithm. It is like a boosted tree algorithm. In this, every tree learns from their predecessors and the update the residual errors. So the next tree will learn from the new updated version of the residuals. XGBoost has advantages like the implementation of regularization which helps in overfitting and can handle sparse data. Let X be a set of base learners like $X = \{x_1 + x_2 \dots \dots + x_n\}$

Then Final Prediction will be,

$$\hat{A}_i = \sum_{b=1}^{n} x_b (z_i) \tag{2}$$

Random Forest is a bagging method and non-parametric algorithm. It is an ensemble learning algorithm used for both classification and regression problems. Generally, Random forest reduces over-fitting and efficiently improves accuracy. For diagnosing and classification of liver patients Random forest model performed very well and achieved good accuracy as it can work for both categorical and continuous values.

Support Vector Machine (SVM) is used for regression and classification techniques. It is one of the robust algorithms in supervised learning models. Majorly SVM is for a two-group classification problem and it can categorize the new text or record very efficiently after giving a labeled training set they are also helpful for outlier detections too. SVM is not a good performer on a large dataset. SVM finds a suitable separable hyper-plane for separating the liver patients from normal persons. It is beneficial for the noisy dataset.

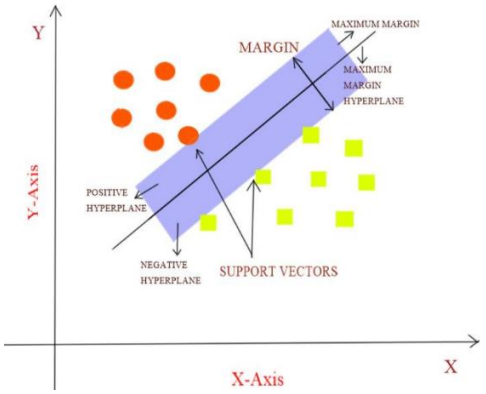

Fig. 3. Support Vector Machine

Gaussian Naïve Bayes or Probabilistic Classification is a supervised machine learning algorithm used for classification problems. It is based on Bayes' theorem i.e. finding the probability of occurring event when we have the probability of other occurred events. It also follows the Gaussian Normal Distribution which has high functionality. This is a linear classifier and after implementation. Mathematically we can see (3) for a better understanding of Naïve Bayes.

$$A = argmax_A \, P(X) \prod_{i=1}^{y} P(\frac{Y_i}{X}) \tag{3}$$

KNN stands for K-Nearest Neighbours which is a supervised ML algorithm that is a non-parametric method and it can efficiently solve classification and regression problems. KNN consumes much time on a large dataset and becomes slow when the dataset gets enlarged. It stores all the data points available and can classify new points based on the previous classification stored data. Another name of the KNN algorithm is the Lazy Learner Algorithm. Euclidean distance plays an important in classification and finding the ideal value of K. Fig. 4 shows the graph for K-values for our dataset for predicting liver disease.

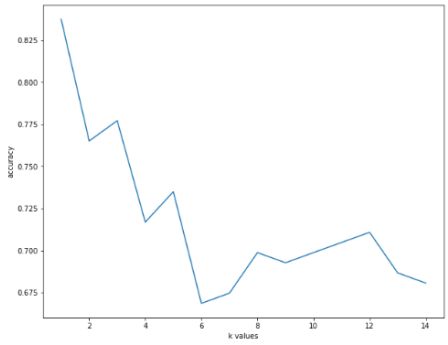

Fig. 4. Graph of K-Values for the liver disease prediction

$$P(x = a|Z = z) = \frac{1}{K} \sum_{i \in Y} I(x^{(i)} = a) \tag{4}$$

The decision tree is a supervised machine learning algorithm that creates a training architecture that can be used for predicting the classes of the targeted variable by learning from simple decision rules which were inferred from the earlier dataset. It is

a multilevel approach that breaks the complex situation into simpler nodes. In this algorithm Parent node divides into two parts then the child node works as parent nodes for further nodes and or works as a decision node for further decisions which creates a tree this whole process is known as a decision tree algorithm. Fig. 5 representation of decision tree.

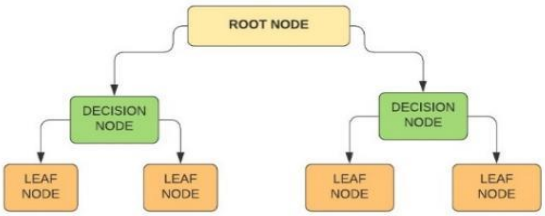

Fig. 5. Representation of Decision Tree

Gradient Boosting reduces the overall prediction error and the main focus is that the target outcomes for the upcoming model will get minimum error from the previous model it gives the best possible outcome for the successor model by reducing the errors. Light Gradient Boosting Machine known as LightGBM produces good accuracy and has fast training speed and this boosting technique is based on tree-based learning and models are fit by any arbitrary differentiable loss function and gradient descent optimization algorithm.

CatBoost is boosting algorithm which can be implemented on various types of data categories like audio, images, and text. As it is diversified so it takes more time for implementation. CatBoosting gives good results on a small dataset.

Adaptive Boosting is known as Adaboost which uses an ensemble method for implementation and it mainly reduces the bias and variance. It is a statistical classification meta-algorithm that is beneficial when other algorithms are overfitting then Adaboost can give better results without getting overfit.

V. Implementation And Results

After data exploration and implementation of various machine learning and boosting algorithms, it is essential to evaluate your results in terms of evaluation metrics for finding the best suitable classifier for the dataset. Abbreviations used in Table I and equations are given in Table I. The performance of all the classifiers is shown in Table II. In Table II we have compared all the classifiers in terms of their Sensitivity, Specificity, Precision, MCC, F1 score, Accuracy, Confusion matrix, and FNR, FDR, NPV.

TABLE I. Abbreviations used in Table 2

Abbreviations	Meaning
MCC	Matthews Correlation Coefficient
NPV	Negative Predicted Value
FDR	False Discovery Rate
FNR	False Negative Rate

$$\text{Sensitivity} = {}^{\text{TP}}/_{(\text{TP}+\text{FN})} \quad (5)$$

$$\text{Specificity} = \text{TN}/(\text{FP} + \text{TN}) \quad (6)$$

$$Precision = \text{TP}/(\text{TP} + \text{FP}) \quad (7)$$

$$F1\ Score = 2\text{TP}/(2\text{TP} + \text{FP} + \text{FN}) \quad (8)$$

$$\text{MCC} = \frac{\text{TP x TN} - \text{FP x FN}}{\sqrt{((\text{TP}+\text{FP})*(\text{TP}+\text{FN})*(\text{TN}+\text{FP})*(\text{TN}+\text{FN}))}} \quad (9)$$

$$\text{NPV} = \text{TN}/(\text{TN} + \text{FN}) \quad (10)$$

$$\text{FDR} = \text{FP}/(\text{FP} + \text{TP}) \quad (11)$$

$$\text{FNR} = \text{FN}/(\text{FN} + \text{TP}) \quad (12)$$

Where TP TN FP FN is True Positive, True Negative, False Positive, and False Negative respectively. The confusion matrices of all the used classifiers have been shown in Fig. 6, Fig. 7, Fig. 8, Fig. 9, Fig. 10, and Fig. 11.

TABLE II. Metric Evaluation Of All Models

S.no.	Model	Accuracy	Sensitivity	Specificity	Precision	MCC	F1- Score	FNR	NPV	FDR
1	XG Boost	0.8675	0.9155	0.8316	0.8025	0.7394	0.8553	0.0845	0.9294	0.1975
2	Support Vector Machine	0.6627	0.7907	0.6179	0.4198	0.3581	0.5484	0.2093	0.8941	0.5802
3	Decision Tree	0.8313	0.9077	0.7822	0.7284	0.6736	0.8082	0.0923	0.9294	0.2716
4	KNN	0.7349	0.8033	0.6952	0.6049	0.4808	0.6901	0.1967	0.8588	0.3951
5	Logistic Regression	0.7289	0.8103	0.6852	0.5802	0.4727	0.6763	0.1897	0.8706	0.4198
6	**Random Forest**	**0.8855**	0.9189	**0.8587**	**0.8395**	**0.7733**	**0.8774**	0.0811	0.9294	**0.1605**
7	Gaussian Naïve Bayes	0.7169	**0.9250**	0.6508	0.4568	0.4926	0.6116	**0.0750**	**0.9647**	0.5432
8	Gradient Boosting	0.8012	0.8871	0.7500	0.6790	0.6165	0.7692	0.1129	0.9176	0.3210
9	CatBoost	0.8193	0.8806	0.7778	0.7284	0.6462	0.7973	0.1194	0.9059	0.2716
10	AdaBoost	0.7289	0.7195	0.7381	0.7284	0.4577	0.7239	0.2805	0.7294	0.2716
11	LightGBM	0.8434	0.8873	0.8105	0.7778	0.6907	0.8289	0.1127	0.9059	0.2222

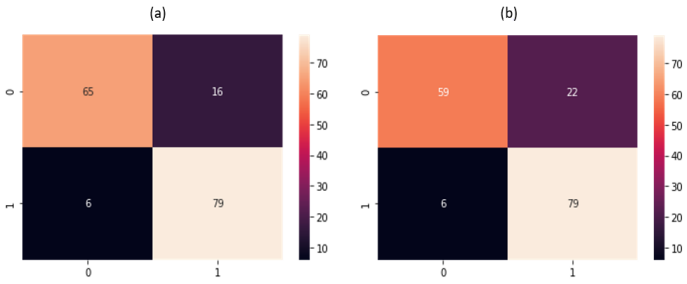

Fig. 6. Confusion Matrix of (a) XG Boost (b) Decision Tree

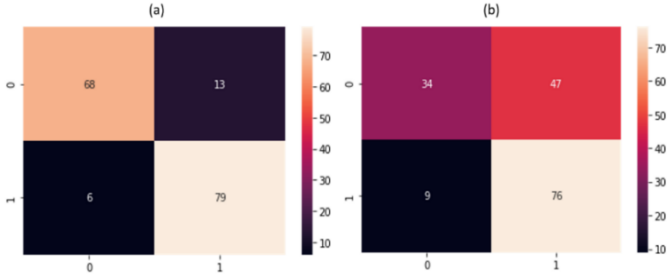

Fig. 7. Confusion Matrix of (a)Random Forest (b) Support Vector Machine

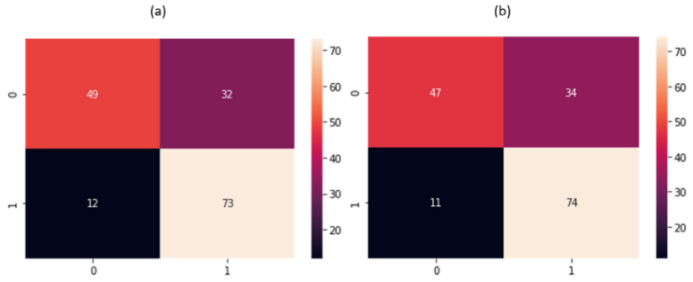

Fig. 8. Confusion Matrix of (a) KNN (b) Logistic Regression

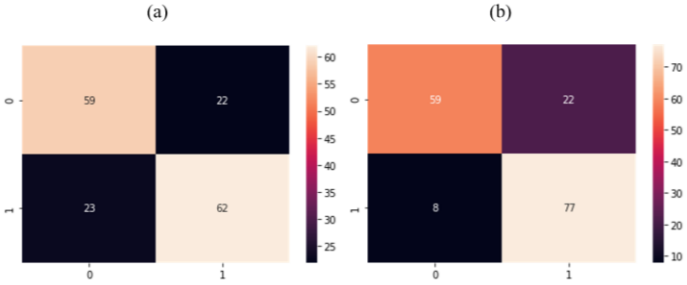

Fig. 9. Confusion Matrix of (a) AdaBoost (b) CatBoost

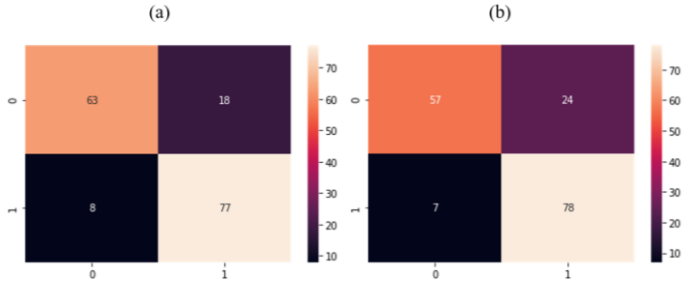

Fig. 10. Confusion Matrix of (a) LightGBM (b) Gradient Boosting

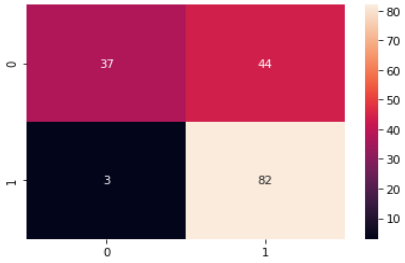

Fig. 11. Confusion Matrix of Gaussian Naïve Bayes

From confusion matrices, we have calculated all the evaluation metrics like Sensitivity, Specificity, and many more. From Table II we can observe that Random Forest performed very well and maintained a good and stable relationship with the dataset and gained an accuracy of 0.8855 followed by XGBoost with 0.8675 and XGBoost was followed by LightGBM with an accuracy of 0.8434. In every field, Random Forest performed outstandingly when compared with other models. In Sensitivity and NPV Gaussian Naïve Bayes performed well than Random Forest with a score of 0.9250 and 0.9647 respectively and also scored well in FNR metric section with a score of 0.0750. Support Vector Machine was the least scorer in the accuracy section with a value of only 0.6627 and other areas too and lacked in metric evaluation section and had very low MCC values of 0.3581. SVM lacked in building a good connection with the dataset might be due to the large dataset or SVM was not able to fine-tune the parameters for finding good results. A full comparison between all the models is mentioned in Table II. Logistic regression was a stable performer in all the fields gained an accuracy of 0.7289 and had an intercept value of 1.47300. In the Decision Tree for criterion ‘entropy’ was used and others were default parameters. From Fig. 4 we can observe the K-value graph for our dataset and KNN was a good performer in overall performance and gave stable and accurate results for detecting liver patients. SVM was a bad performer in all fields and the Confusion matrix was also not good when compared with XGBoost or Random Forest. XGBoost was the close competitor of Random forest they both were close and gave results but Random forest was able to build a better connection than XGBoost. After data exploration and feature selection performance were very good for the detection of liver diseases

VI. CONCLUSION AND FUTURE WORK

Liver diseases are dangerous diseases and need high accuracy for prediction. In this paper, we used different machine learning algorithms like Decision tree, Random Forest, XG Boost, Gaussian Naïve Bayes, Gradient Boosting, AdaBoost, CatBoost, KNN, Logistic Regression, Support Vector Machine, and LightGBM for predicting liver disease by using real-time deployment of algorithms on the Indian liver patient dataset and we compared the different models in all the fields and we founded that Random forest performed very well and gained good accuracy and was stable and precise. SVM was not a good performer when compared to other algorithms. The main goal was for finding the best ML algorithm for liver disease prediction for predicting liver disease by using the real-time deployment of algorithms on the Indian liver patient dataset and we compared the different models in all the fields and we

founded that Random Forest performed very well and gained good accuracy and was stable and precise. For future work, we can implement ANN architecture and compare them with different Machine Learning Algorithm and we can do more feature engineering and selection of more related columns. Further, more advanced machine learning techniques can be implemented for liver disease prediction.

ACKNOWLEDGMENT

The author would like to acknowledge and proffer my warmest thanks to all the co-authors for their efforts and co-operation.

REFERENCES

[1] S. H. Saverymuttu, A. E. Joseph and J. D. Maxwell, "Ultrasound scanning in the detection of hepatic fibrosis and steatosis," Br Med J (Clin Res Ed), vol. 292(6512), pp. 13-15, 1986

[2] J. M. Clark, "The epidemiology of the nonalcoholic fatty liver disease in adults," Journal of clinical gastroenterology, vol.40, pp. S5-S10, 2006

[3] M. Lazo and J. M. Clark, "The epidemiology of nonalcoholic fatty liver disease: a global perspective," In Seminars in liver disease © Thieme Medical Publishers, Vol. 28(4), pp. 339-350, November, 2008.

[4] Q. M. Anstee, G. Targher, and C. P. Day, "Progression of NAFLD to diabetes mellitus, cardiovascular disease, or cirrhosis," Nature Reviews Gastroenterology & Hepatology, vol. 10(6), pp. 330, 2013

[5] H. GM, M. K. Gourisaria, S. S. Rautaray, and M. Pandey, "UBMTR: Unsupervised Boltzmann machine-based time-aware recommendation system," Journal of King Saud University-Computer and Information Sciences, 2021

[6] M. A. Taufiq, N. Hameed, A. Anjum, and F. Hameed, "m-Skin Doctor: A mobile-enabled system for early melanoma skin cancer detection using a support vector machine," eHealth 360° Springer, Cham, pp. 468-475, 2019

[7] S. Chandra, M. K. Gourisaria, H. GM, S.S. Rautaray, M. Pandey and S. N. Mohanty, "Semantic Analysis of Sentiments through Web-Mined Twitter Corpus," In Proceedings of CEUR Workshop Proceedings, vol. 2786, pp. 122-135, 2021

[8] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," Artificial Intelligence in medicine, vol. 23(1), pp. 89-109, 2001

[9] H. GM, M. K. Gourisaria, M. Pandey and S. S. Rautaray, "A comprehensive survey and analysis of generative models in machine learning," Computer Science Review, vol. 38, pp.100285,2020

[10] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," Annual review of biomedical engineering, vol. 19, pp. 221-248, 2017

[11] G. Jee, H. GM, and M. K. Gourisaria, "Juxtaposing inference capabilities of deep neural models over posteroanterior chest radiographs facilitating COVID-19 detection," Journal of Interdisciplinary Mathematics, pp.1-27, 2021.

[12] J. H. AlKhateeb, J. Ren, J. Jiang, and H. Al-Muhtaseb, "Offline handwritten Arabic cursive text recognition using Hidden Markov Models and re-ranking," Pattern Recognition Letters, vol. 32(8), pp. 1081-1088, 2011

[13] A. Sahu, H. GM, and M. K. Gourisaria, "A Dual Approach for Credit Card Fraud Detection using Neural Network and Data Mining Techniques," 17th India Council International Conference (INDICON) IEEE, New Delhi, India, pp. 1-7, 2020

[14] A. Andrade, J. S. Silva, J. Santos, and P. Belo-Soares, "Classifier approaches for liver steatosis using ultrasound images," Procedia Technology, vol. 5, pp. 763-770, 2012

[15] M. Owjimehr, H. Danyali, M.S. Helfroush, and A. Shakibafard, "Staging of fatty liver diseases based on hierarchical classification and feature fusion for back–scan–converted ultrasound images," Ultrasonic Imaging, vol. 39(2), pp. 79-95, 2017

[16] G. Li, Y. Luo, W. Deng, X. Xu, A. Liu and E. Song, "Computer-aided diagnosis of fatty liver ultrasonic images based on support vector machine: engineering in medicine and biology society," 30th Annual International Conference of the IEEE EMBS 2008, pp. 4768–4771, 2008

[17] R. Ribeiro, and J. Sanches, "Fatty liver characterization and classification by ultrasound," Iberian conference on pattern recognition and image analysis. Springer, Berlin, Heidelberg, pp. 354-361, June, 2009

[18] M. B. Subramanya, V. Kumar, S. Mukherjee, and M. Saini, "A CAD system for B-mode fatty liver ultrasound images using texture features," Journal of medical engineering & technology, vol 39(2), pp. 123-130, 2015

[19] H. Y. Ma, Z. Zhou, S. Wu, Y. L. Wan, and P. H. Tsui, "A computer-aided diagnosis scheme for detection of fatty liver in vivo based on ultrasound kurtosis imaging," Journal of medical systems, vol. 40(1), pp. 1-9, 2016

[20] L. Saba, N. Dey, A. S. Ashour, S. Samanta, S.S. Nath, et al., "Automated stratification of liver disease in ultrasound: an online accurate feature classification paradigm," Computer methods and programs in biomedicine, vol. 130, pp. 118-134, 2016

[21] S. Vijayarani, and S. Dhayanand, "Liver disease prediction using SVM and Naïve Bayes algorithms," International Journal of Science, Engineering and Technology Research (IJSETR), vol. 4(4), pp. 816-820, 2015

[22] S. Dhamodharan, "Liver disease prediction using Bayesian classification," COMPUSOFT: An International Journal of Advanced Computer Technology, 2014

[23] R. Sharma, M. K. Gourisaria, S. S. Rautray, M. Pandey and S. S. Patra, "ECG Classification using Deep Convolutional Neural Networks and Data Analysis," International Journal of Advanced Trends in Computer Science and Engineering, vol. 9(4), pp. 5788-5795, 2020

[24] I. Saha, M. K. Gourisaria, and H. GM, "Distinguishing Pneumonia and COVID-19: Utilizing Computer Vision to Mimic Clinician Efficacy," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS) IEEE, pp. 834-841, March, 2021

[25] S. Bind, A. K. Tiwari, A. K. Sahani, P. M. Koulibaly, F. Nobili, M. Pagani, et al., "A survey of machine learning-based approaches for Parkinson disease prediction," International Journal of Computer Science and Information Technologies, vol. 6(2), pp. 1648-1655, 2015

[26] H. GM, M.K. Gourisaria, S.S. Rautaray and M. Pandey, "Pneumonia detection using CNN through chest X-ray," Journal of Engineering Science and Technology (JESTEC), vol. 16(1), pp. 861-876, 2021

[27] H. Asri, H. Mousannif, H. Al Moatassime, and T. Noel, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," Procedia Computer Science, vol. 83, pp.1064-1069, 2016

[28] S. Khare, M. K. Gourisaria, H. GM, S. Joardar, and V. Singh, "Real Estate Cost Estimation Through Data Mining Techniques," IOP Conference Series: Materials Science and Engineering IOP Publishing, vol. 1099(1), pp. 012053, March, 2021

[29] J. Patel, D. Tejal Upadhyay, and S. Patel, "Heart disease prediction using machine learning and data mining technique," Heart Disease, vol. 7(1), pp. 129-137, 2015

[30] M. K. Gourisaria, S. Das, R. Sharma, S. S. Rautaray, and M. Pandey, "A Deep Learning Model for Malaria Disease Detection and Analysis using Deep Convolutional Neural Networks," International Journal of Emerging Technologies, vol. 11(2), pp. 699-704, 2020.

[31] M. Lichman, UCI Machine Learning Repository, Irvine, CA: University of California, School of Information and Computer Science, 2013 http://archive.ics.uci.edu/ml