

Comparative Analysis of Meta Learning Algorithms for Liver Disease Detection

Maruf Pasha^{1*}, Meherwar Fatima²

¹ Department of Information Technology, Bahauddin Zakariya University, Pakistan.

² Department of Computer Science and Information Technology, Women University, Pakistan.

* Corresponding author. Email: maruf.pasha@bzu.edu.pk

Manuscript submitted January 10, 2017; accepted March 8, 2017.

doi: 10.17706/jsw.12.12.923-933

Abstract: Various kinds of pressure and unbalanced eating behaviors, along with alcohol inhalation and on-going toxic gases, absorption of tainted nutrients, unnecessary intake of cured food and ingestion of drug, enables patients to increase year by year from liver disease. For this purpose, the type of data mining algorithms can help medical doctors to diagnose patients in hospital. This paper analyzes meta learning algorithms to classify the Indian liver patient dataset. The Data set is attained from UCI repository that contains 583 instances. Adaboost, logitboost, Bagging and Grading meta learning algorithms are applied to this data set. These algorithms are compared on the basis of Correct Classification, Incorrect Classification and Time to build model. Grading is the best algorithm among these meta learning algorithms as it provides highest Correct Classification rate and minimum rate of incorrect classification. Execution time for Grading is less than Adaboost, Logitboost and Bagging. Key role is played by Grading algorithm in shaping enhanced classification accuracy (Correct Classification Rate) of a data set.

Keywords: Disease diagnostic, data mining, machine learning, meta learning.

1. Introduction

Now a days Liver disease is rising rapidly in the world and has become a major health problem in the world. Liver is the biggest glandular organ in living organisms/animals and it works on many complex tasks to retain the body free from toxins and harming materials. An approximate weight of liver is three pounds. It is composed of four sections. 1.5 quarts of blood is taken every minute through hepatic artery and portal vein. Liver is thought about taken as Gland- an organ that takes out chemicals as it creates bile, in which fats are digested. To make it able to absorb in small intestine, bile's salts convert fats into very fine pieces. Other than creating bile, the liver performs following functions. 1) To make the blood pure from damaging drugs. 2) It keeps iron, vitamins, Sugar and glucose. 3) If the glucose level in the body goes down from normal/ average then it changes the stored sugar into functional sugar. 4) Hemoglobin, insulin and some hormones are destroyed/broken. 5) Ammonia is transformed into urea that is important in metabolism of body. 6) Abolishment of RBC's which makes waste materials which is of brown color, discoloration of stool/ fecal matter, make the urine dark, can be the cause of liver infection e.g. Viral hepatitis.

Other symptoms of liver infection are jaundice in which skin & eyes turn yellow. Some other liver infections comprise hepatitis, fatty liver disease, causes & affliction from alcohol, acetaminophen that relieve pain & other cancer causing medicines.

1.1. Liver Dialysis

In this process a machine detoxifies the liver, is a modern medical cure, in this a patient no longer survives than few years. Dialysis is processed during kidney failure and liver transplantation.

1.2. Anatomy of Liver

Liver is found in the top right-handed part of abdominal cavity just below the diaphragm and above the stomach, intestines and right kidney. Two major sources of blood supply towards liver are (1) Hepatic artery functions to stream oxygenated blood. (2) Nutrient-rich blood streams through hepatic portal vein.

Liver is composed of two major lobes, each of which is made up of eight compartments. Hepatic duct is formed by the connection of lobules to little ducts that join with the larger ducts. This common hepatic duct transfers bile that is made by liver cells to duodenum and gallbladder.

1.3. Types of Liver Infection

Figure 1 presents a variety of liver Infections. Viruses such as hepatitis A, B and C are the famous causes of hepatitis. Hepatitis can also have non-toxic reasons, which includes, drinking, obesity, drugs and allergic reactions etc. Scarring of liver caused by continuous damage, called cirrhosis. Due to this, liver cannot perform functioning perfectly. It takes place when oxygen turns stiff. Chronic liver disease is the main reason of this disease due to abusive use of alcohol and hepatitis C inflammation. Liver Cancer is fifth most known cancer in the world, after the presence of cirrhosis hepatocellular carcinoma remains in it permanently. Liver failure has many reasons like inflammation, genetic infection and heavy intake of alcohol. Ascites is a collection of flood in the cavity of abdomen, which is the result of cirrhosis. Gallstones are hard deposit in gall bladder that can eventually block existing bile. Hemochromatosis functions as the deposition of iron in liver, which damages it.

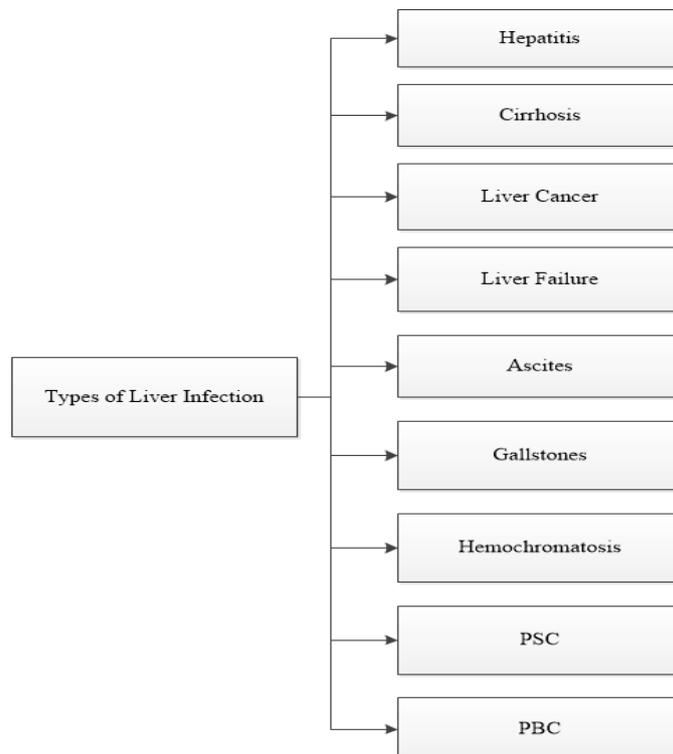


Fig. 1. Types of liver infection.

Deposition of iron causes many medical complications. Primary sclerosing cholangitis is a chronic liver disease. It results in damage of bile ducts in the liver. Primary biliary cirrhosis is an unknown mechanism gradually damages the bile ducts in liver. Liver cirrhosis is developed permanently.

Hepatitis, Cirrhosis, Liver Cancer, Liver Failure, Ascites, Gallstones, Hemochromatosis, Primary sclerosing cholangitis and Primary biliary cirrhosis are the different forms of Liver infection. These Infections damage the liver functioning badly.

Liver Tests

Following blood tests are performed for diagnosis of liver disease.

Liver function panels: This panel consists of many different blood tests and checks that how properly liver is functioning.

ALT (Alanine Aminotransferase): The higher level of ALT assists in the recognition of disease of liver such as hepatitis.

AST (Aspartate Aminotransferase): In addition with increased ALT, Aspartate Aminotransferase AST also analyze the causes of liver destruction.

Alkaline phosphates: In bones as well as bile secreting cells in liver, alkaline phosphates are found. high level often means bile flow out of the liver is blocked.

Bilirubin: a difficulty with liver is occurred by elevated bilirubin quantity.

Albumin: Albumin analyses how perfect the liver is functioning, as it is a component of protein.

Ammonia: when liver is not working accurately or perfectly, the level of ammonia is increased.

Hepatitis A test: The doctors will test working of liver in addition with antibodies to identify the hepatitis A virus, if the hepatitis A is diagnosed.

Hepatitis B test: If hepatitis B is suspected, doctors will determine the level of antibodies to analyse whether the person is suffering from hepatitis B virus or not.

Hepatitis C test: Its blood test analyse if a person has been suffering with hepatitis C and also examine the working of liver.

Prothrombin Time (PT): This test is accomplished to check if someone is taking adequate dosage of medicine of blood thinning warfarin. It can check problems of blood clotting.

Partial Thromboplastin Time (PTT): This test is done to determine blood-clotting trouble.

In numerous automatic medical diagnoses, classification techniques are generally used. At early stage problems with liver patients are not simply revealed even if the liver is partially damaged, it will perform all its functions normally [1].

Knowledge is the most important strength of any organization in Information Technology (IT) driven society. People are expecting better and reasonable healthcare due to the on-going development in IT applications in the area of healthcare. Valuable medical information's are now easily accessible due to computerized hospital information systems (HIS). Use of sophisticated equipment in the practice of modern medicine produces huge amount of data. After storing this data in digital form, to extract knowledge by automated methods of data analysis, significant amount of effort is being made for this purpose. Knowledge can be used for well speedy clinical decision-making. The area that consists of data mining and knowledge discovery tools are very helpful in achieving these goals. Including data mining, knowledge discovery is a distinct process that consists of different steps. Usually the healthcare environment is very informative but lacks in knowledge. Though, for better knowledge in healthcare environment data mining techniques can be applied for this purpose. As the large amount of medical data being produced, it is need of the hour to improve methods of data analysis and knowledge discovery using appropriate data mining techniques. A convenient resource for data mining and knowledge discovery provides a proper medical database that was created with intention mining. Knowledge discovery in databases (KDD) and data mining are the two terms, which are interchangeable. Process of finding useful information and extracting knowledge is called KDD and data mining respectively [2]. Learning from the learned knowledge Is roughly called Meta-learning. It is a modern technology used to compute high-level models. In principled fashion, Meta classifiers are

integrated. Separate learned classifier cleans the information to enhance the predictive performance. In the process of Meta learning, number of learning program is applied on the number of data subsets, parallel. Then in the form of classifier combined outcome is gathered [3]. For detecting liver diseases, it is a thought to recommend an computerized method that is based on previous data.

2. Materials and Method

This section presents the existing literature, data set description, WEKA: Data mining Tool and Meta Learning Algorithms.

2.1. Existing Literature

Prediction of liver disease by using SVM and Naïve bayes algorithms are done by the Vijayarani and Dhayanand. For this experiment data-set is attained from UCI machine Learning repository. It contains 10 attributes and 560 rows. Analysis is done on the basis of Accuracy and time to build model by using MATLAB. In 1670.00 ms, Naive bayes attained 61.28% accuracy. 79.66% of correctness is obtained by SVM in 3210.00 ms. In case of accuracy, SVM shows highest result but Naïve bayes takes less time when compared with SVM [4].

For the analysis of liver disease, Rajeswari and Reena, used Naïve bayes, K star and FT tree data mining algorithms. Data set contains 345 instances and 7 attributes. It is obtained from UCI. In 0 sec, Naive bayes offers 96.52% accuracy. In 0.2 sec, 97.10% of correctness is attained after utilizing FT tree. K star algorithm provides 83.47% of accuracy in 0 sec. These results are obtained after applying 10 cross fold validation test in the tool of WEKA. FT offers highest accuracy [5].

Performance of classification algorithms is analyzed by Jin[6]. They consider the characteristics of algorithms to the data of liver patient. Algorithms that were previously used are, Naïve bayes, Decision Tree, Multi-Layer Perceptron and K-NN. They proposed Random forest and Logistic algorithms to develop data set. Evaluation criteria for these algorithms are precision, recall, sensitivity, specificity, and so on. Naïve bayes algorithm provides better precision as compared to other algorithms. On the basis of Recall and sensitivity Logistic and Random forest take preference.

For Liver Patient Classification, Gulia and colleagues [7] shows computational intelligence techniques. J-48, MLP, SVM, Random Forest and Bayesian Network are selected for the classification. Data set is collected from UCI. They implement hybrid model construction. To enhance accuracy they make comparative analysis in three phases. In first phase, algorithms are implemented before feature selection. In second phase, Classification algorithms are applied after feature selection. Before feature selection SVM provides highest accuracy as compared to other algorithms. But, after feature selection, Random Forest algorithm shows highest accuracy. In third phase, comparative analysis (after and before feature selection) is made for improving accuracy. Obtained results indicate that the Random Forest provides the highest prediction accuracy of 71.8696% after feature selection.

Authors, [8] performs a critical study for diagnosing of liver disease. In this paper NB, C4.5, Back propagation Neural Network and SVM are thought to be as classification algorithms. Evaluation criteria for these classification algorithms are Accuracy, Precision, Sensitivity and Specificity.

Karthik and colleagues, [9] performs a work to diagnose liver disease. In the first step, for classification of liver disease, ANN is applied. In second step, to produce classification rules, rough set rule induction using Learning by Example algorithm is applied. In third step, to recognize the types of the liver disease, fuzzy rules are useful.

2.2. Data Set Description

Indian Liver patient data set is obtained from UCI [10] machine learning repository. This data set was

gathered from Andhra Pradesh, India. Data set consists of 583 instances in which disease is present in 416 patients. Disease is excluded in 167 instances. Data set contains 11 attributes that are age, Gender, TB (total Bilirubin), DB (direct Bilirubin), Alkphos (Alkaline Phosphotase), SGPT Alamine Aminotransferase, SGOT Aspartate Aminotransferase, TP (total proteins), ALB (albumin), A/G ratio (Albumin and Globulin Ratio) and Selector field. Selector is a class that divides the whole data set into two parts, "Liver patients or non-Liver Patient".

2.3. WEKA: Tool of Data Mining

WEKA [11], is an open source data-mining tool with various algorithms implemented in it. It has four interfaces that are Explorer, Experimenter, Knowledge Flow and Simple CLI. 10 fold cross validation test with four meta learning algorithms are applied on the original data set for liver disease diagnostics.

2.4. Meta Learning Algorithms

Various meta learning techniques are discussed in this section.

2.4.1. Ada-boost

Yeave Freud and Robert Scapire, developed Adaptive boosting meta learning algorithm known as Ada-Boost. With various algorithms Ada-boost meta learning algorithm may be used in conjunction to enhance their performance. As linear combination, Algorithm of Ada Boost is used to construct a "powerful" classifier. Ada Boost is adaptive just because "succeeding classifiers constructed are weak in approving of those instances that are incorrectly classified by the preceding classifiers". For noisy data and outliers Adaptive boosting is very sensitive. In some cases, it may be less vulnerable to the problem of over fitting as compared to many learning algorithms. Classifier that is used by this algorithm may be weak i.e. "shows a considerable error rate" but their process of performing task is not unexpected/unusual that results 0.5 error rate for binary classification, final model will be improved [12][3]. Even classifiers having an error rate greater than would be estimated from an unexpected classifier will be applicable, since in the concluding linear combination of classifiers, they will have negative coefficients and therefore act like their opposites [12].

2.4.2. Logit boost

Jerome Friedmome, Trevor Hastie, and Robert Tibshirani constructed a boosting algorithm known as LogitBoost. An application of logistic regression to AdaBoost method is presented by LogitBoost [12] [3].

2.4.3. Bagging

Bagging that is Recognized re-sampling ensemble method, are used to produce and combine different classifiers by using the same learning algorithm for the base-classifiers. Bagging, a name derived from "bootstrap aggregating" is to improve unstable assessment or classification algorithm. For a given base method, bagging as a variance reduction technique, was encouraged by Breiman such as DT or methods that can be used to ensure the selection of variables and appropriates in a linear type[13]. Bagging technique is the process of merging the decisions of different techniques. For Classification process Bagging was actually planned and is generally applicable to models of decision trees, however it can also be employed with any classification or regression model. It can also be defined as combining the several results into one calculation. Basic idea for this is to take average. In bagging algorithm, equal weights are received by Models. Randomly different training sets having the equal size are selected. For each dataset, assume that specific ML techniques are used to build a decision tree. It can be believed that these trees are almost same and making the identical prediction for ever new test instance [12]. When using unstable Nonlinear models, bagging is effective.

2.4.4. Grading

Grading is another Meta learning algorithm that are examined. Main concept behind it is to forecast whether the prediction for every actual learning algorithm for specific example is right or not. On training set, one classifier for every actual learning algorithms is being trained which includes the actual examples with class labels that encrypt whether this learner’s forecast was correct on specific example [12]. The sense of classification to be marked as correct or incorrect can be done by the technique named as “graded”. By this method, the classification made by the k different classifiers into k training sets is converted by using the occurrences k times and assigning them a new binary class in each instance. Comparing the real class of occurrence, the class directs whether the k-th classifier generated is correct or incorrect classification.

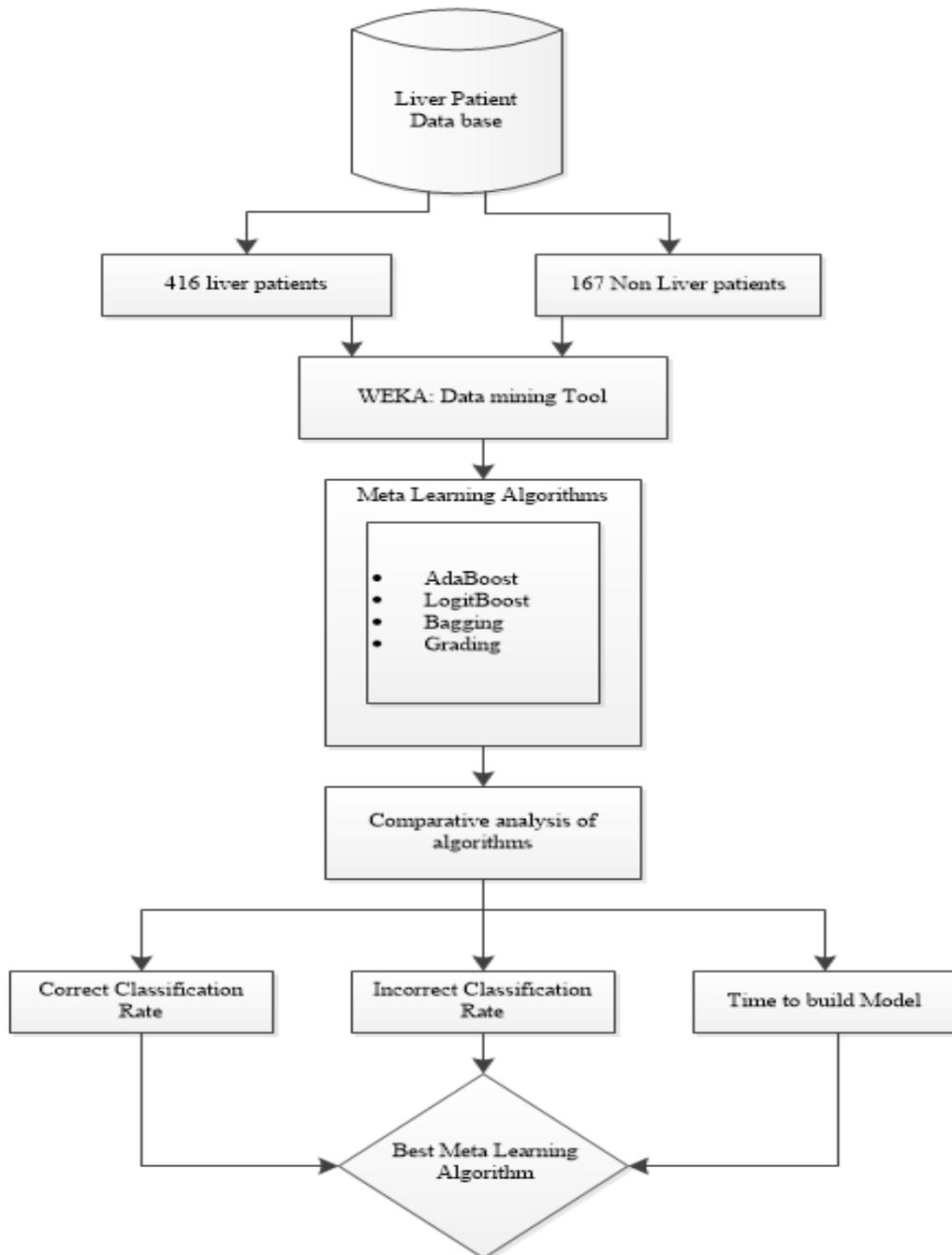


Fig. 2. Comparative analysis of algorithms for liver disease detection.

For the indication of each classifier when it inclines to misclassify a given occurrence, one meta classifier is being trained for every classifier for this purpose. At the time of classification, each classifier attempts to classify non-labeled occurrence. By gathering the outputs of the classifiers that are accepted as correct by meta-classification schemes, we get the final classification. For the partitioning of the training set into k-sets, grading is the process for the generalization of the cross validation selections. It also helps to form k-1 classifiers by reducing one set at a time and then uses it for finding the incorrect classification rate. At the end, the process basically indicates the classifier consistent to the subset with the smallest incorrect classification. For all occurrences grading attempts to make this decision individually, by using only those classifiers that are expected to classify that occurrence accurately [14].

3. Comparative Analysis of Algorithms

These four Meta learning Algorithms are analyzed on the basis of Correct Classification, Incorrect Classification and Time to build model. Figurative approach for performing these steps is present in Fig. 2.

Liver disease data sets are evaluated by using WEKA: Data mining Tool. Four Meta learning algorithms Adaboost, Logitboost, Bagging and Grading are compared to find best Meta learning algorithm among these four Algorithms. Correct Classification, Incorrect Classification and Time to build Model are the parameters to analyse these Algorithms.

4. Results and Discussion

4.1. Analysis on the Basis of Correct Classification Rate

Meta learning Algorithms are analysed on the basis of Correct Classification. Table 1 shows Correct Classification rates of Algorithms. Adaboost Algorithm provide lowest Correct Classification rate of 70.3259%. Logitboost and Bagging both offers 70.4974% of Correct Classification Rate. Grading meta-learning Algorithm shows the highest Correct Classification rate of 71.3551%. Figurative approach of Correct Classification rate of algorithms is shown in Fig. 3.

Table 1. Comparative Analysis of Algorithms on the basis of Correct Classification

Meta Learning Algorithms	Correct Classification rate
Adaboost	70.3259%
Logitboost	70.4974%
Bagging	70.4974%
Grading	71.3551%

Four meta learning algorithms (Adaboost, Logitboost, Bagging and Grading) are compared to analyse the Correct classification rate. Grading Algorithm Shows the highest Correct Classification rate of 71.3551%. Logitboost and Bagging offers 70.4974% of correct Classification rate. Minimum rate of correct Classification is showed by Adaboost algorithm that is of 70.3259%. Comparison of algorithms shows that, Grading Algorithm provides highest rate of correct classification.

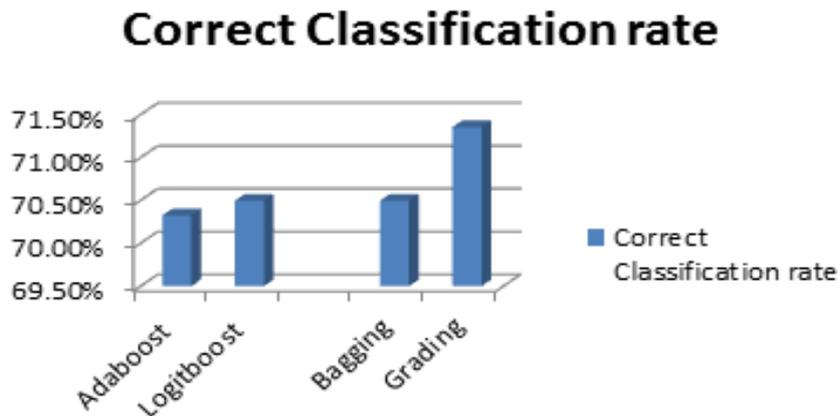


Fig. 3. Correct classification rate of meta learning algorithms.

4.2. Analysis on the Basis of In Correct Classification Rate

After the analysis of Meta learning algorithms on the basis of In correct Classification Rate Table 2, maximum In-correct classification rate is shown by Adaboost that is of 29.6741%. Logitboost and Bagging shows 29.5026% of In correct Classification rate. Minimum incorrect Classification rate of 28.6449% offered by Grading. Fig 4 shows the graphical representation of Incorrect Classification rate of Algorithms.

Table 2. Comparative Analysis of Algorithms on the Basis of In Correct Classification

Meta Learning Algorithms	In correct Classification Rate
Adaboost	29.6741%
Logitboost	29.5026%
Bagging	29.5026%
Grading	28.6449%

Analysis takes place on the basis of incorrect classification rate. As shown in figure 4 Grading algorithm Shows the lowest incorrect Classification rate of 28.6449%. Logitboost and Bagging offers 29.5026% of Incorrect Classification rate. Adaboost Algorithm that is of 29.6741% shows maximum rate of Incorrect Classification.

Grading shows lowest In correct Classification rate as compared to other Algorithms.

4.3. Analysis on the Basis of Time to Build the Model

Execution time for grading is minimum as compared to other algorithms as shown in Table 3. Time to build model for Grading is 0 sec. While both Adaboost and Logitboost takes 0.1 sec to build model. Execution time for Bagging is 0.2 sec. Figurative approach of the time taken by the Algorithms to build model is shown in Fig 5. Adaboost and Logitboost take 0.1 sec to build model. While bagging takes 0.2 sec.

However Grading takes minimum time of 0 sec to build model. Sense of Classification for grading technique is to be marked as Correct or incorrect. Analysis shows that Grading is the best Algorithm as it provides highest Correct Classification rate, Lowest Incorrect Classification rate and minimum execution time (Time to build model).

In correct Classification Rate

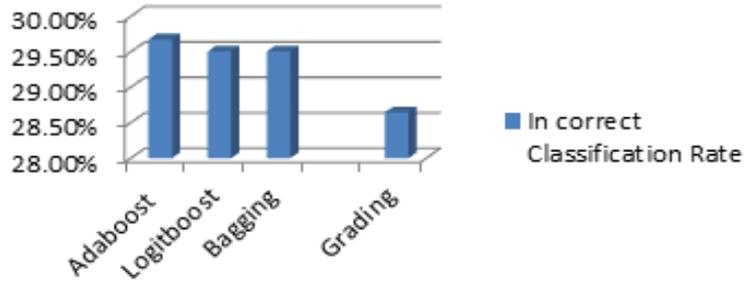


Fig. 4. Incorrect classification rate of meta learning algorithms.

Table 3. Comparative Analysis of Algorithms on the Basis of Execution Time (Time to Build Model)

Algorithms	Execution Time in sec
Adaboost	0.1
Logitboost	0.1
Bagging	0.2
Grading	0

Execution Time in sec

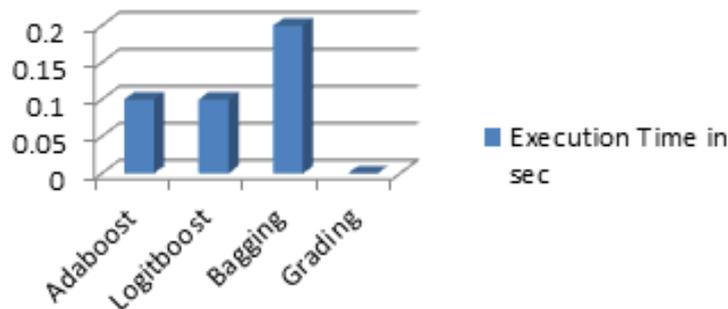


Fig. 5. Time to build model of meta learning algorithms.

Grading takes 0 sec to build model. Grading is the best algorithm as it takes minimum time as compared to other algorithms as depicted in Fig. 5.

5. Conclusions

For clinical evaluation, different data mining techniques are present. In this study, four (Adaboost,

Logitboost, Bagging and Grading) meta learning algorithms are used to classify the data. Criteria for their evaluation are Rate of Correct Classification, Incorrect Classification and Time to build model. Among these algorithms, Grading Technique is one of the effective data mining algorithms that are used in Liver disease diagnosis. Grading algorithm shows highest correct Classification rate of 71.3551% and lowest In-correct Classification rate of 28.6449% as compared to other Adaboost, LogitBoost and Bagging meta learning Algorithms. The total time taken to build the model is at 0 seconds in the diagnosis of liver disease patients. Hence, for this dataset, Grading meta learning model can be recommended for diagnosis of liver disease.

Acknowledgments

The authors would like to thanks UCI machine learning repository for ILPD (Indian liver Patient dataset) Dataset and developers of Waikato Environment for Knowledge Analysis (WEKA).

References

- [1] Pahareeya, J., Vohra, R., Makhijani, J., & Patsariya, S. (2014). Liver patient classification using intelligence techniques. *Int J Adv Res Comput Sci Softw Eng*.
- [2] Wasan, S. K., Bhatnagar, V., & Kaur, H. (2006). The impact of data mining techniques on medical diagnostics. *Data Sci J*.
- [3] Sen, S. K., & Dash, S. (2014). Application of meta learning algorithms for the prediction of diabetes disease. *Int J Adv Res Comput Sci Manag Stud*.
- [4] Vijayarani, S., & Dhayanand, S. (2014). Liver disease prediction using SVM and naïve bayes algorithms. *Int J Sci Eng Technol Res*, 816-820.
- [5] Rajeswari, P., & Reena, G. S. (2010). Analysis of liver disorder using data mining algorithm. *Glob J Comput Sci Technol*.
- [6] Jin, H., Kim, S., & Kim, J. (2014). Decision factors on effective liver patient data prediction. *Int J Biosci Biotechnol*.
- [7] Gulia, A., Vohra, R., & Rani, P. (2014). Liver patient classification using intelligent technique. *Int J Comput Sci Inform Technol*.
- [8] Ramana, B. V., Prasad, B. M. S., & Venkateswarlu, N. B. (2011). A critical study of selected classification algorithms for liver disease diagnosis. *Int J Database Manag Syst*.
- [9] Karthik, S., Priyadarishini, A., Anuradha, J., & Tripathy, B. K. (2011). Classification and rule extraction using rough set for diagnosis of liver disease and its types. *Pelagia Research Library, Adv Appl Sci Res*.
- [10] ILPD(Indian Liver Patient Dataset)Data set. UCI machine learning repository. Retrieved from: <http://archive.ics.uci.edu/ml/datasets/ILPD+%28Indian+Liver+Patient+Dataset%29/3/5/2016>
- [11] WEKA: Data mining software. WEKA The university of Waikato. Retrieved from: <http://www.cs.waikato.ac.nz/ml/weka/4/5/2016>.
- [12] Sen, S. K., & Dash, S. (2013). Meta learning algorithms for credit card fraud detection. *Int J Eng Res Dev*.
- [13] Nithya, R., Manikandan, P., & Ramyachitra, D. (2015). Performance analysis of meta classifiers algorithms using yeast dataset. *Int J Innov Res Comput Commun Eng*.
- [14] Rani, M. U., & Prasanna, K. G. T. (2012). A study of meta-learning in ensemble based classifier.



Maruf Pasha received his PhD from University of South Brittany, France and he has obtained his MS (IT) from NUST, Pakistan and is currently serving as head, Department of Information Technology at Bahauddin Zakariya University, Multan. His research encompasses semantic web, big data and IoT systems. He has number of publications in international journals and conferences.



Meherwar Fatima received her MS Degree in information technology in 2015 from Bahauddin Zakariya University Multan Pakistan. She obtained her BS in information technology in 2013 from Islamia University Bahawalpur Pakistan. She is currently working as a lecturer, in Institute of Computer Science and Information Technology, The Women University Multan, Pakistan. Her research interests include data mining; machine learning, artificial intelligence and internet of things.