Improved Classification of Liver Diseases Using Machine Learning Techniques

Kundan Kumar[1], Chinmay Bhatt[2]
CSE, SRK University, Bhopal, India[1, 2]
Kundan42me@gmail.com[1], chinmay20june@gmail.com[2]

Abstract: *Both industrialized and developing nations exhibit elevated prevalence rates of chronic ailments, including diabetes and Chronic Liver Disease (CLD). The prevalence of liver disease has been consistently increasing due to the escalating rates of excessive alcohol intake, exposure to harmful gases, and ingestion of tainted food, medicines, and pharmaceuticals. Due to the liver's involvement in numerous essential functions, liver illness poses significant challenges for the provision of medical treatment. Classification algorithms are commonly employed in many automatic medical diagnosis systems. The early identification of cancer has significantly enhanced the patients' prospects of survival in numerous instances. Support vector machines, or SVMs, are supervised learning models that analyze data using specific learning techniques to detect patterns. In this study, we proposed a novel methodology utilizing support vector machines, neural networks, and optimization for classification purposes. This approach, which draws inspiration from biology, aims to improve the results of classifiers in terms of performance criteria such as accuracy, precision, recall, etc. It is built upon the classification techniques of support vector machines, neural networks, and optimization. We assess and enhance all the performance metrics related to the dataset concerning liver patients. The required dataset for the study can be accessed from the UCI machine learning repository, where the dataset including information about all the patients was sourced. Our experimental results demonstrate that our method has a higher classification recognition rate for performance attributes compared to other currently employed strategies.*

Keywords: Liver Disease, Chronic Disease Detection, SVM, Precision Medicine, Bioinformatics.

Introduction

One vital organ that serves as a crucial barrier between the gastrointestinal blood, which is heavily contaminated with toxins and antigens, and the rest of the body is the liver. The liver is a major producer of hormones and enzymes and carries out a number of vital biological processes. By transforming toxins into absorbable materials, it is also the organ in charge of eliminating toxins from the blood. The term "liver disease" encompasses a wide range of illnesses and disorders that can lead to decreased liver function. Although intrinsic malfunction may occur, disorders of other organ systems frequently have a secondary effect on the liver due to its involvement in numerous metabolic and detoxifying processes. Significant disruption in glucose homeostasis is linked to the occurrence of chronic liver disease (CLD). About 30 to 60% of CLD patients have frank diabetes, whereas up to 80% of patients have glucose intolerance [1]. CLD significantly affects the metabolism of glucose in the liver, depending on its etiology.

The care of diabetes in patients with liver disease is conceptually complicated by liver-related abnormalities in drug metabolism, potential therapeutic interactions, and the prevalence of hepatotoxicity. Liver disease is a significant cause of death for specific types of diabetic patients.

A complex group of diseases with many different causes is called diabetes. High blood glucose, commonly known as hyperglycemia or high blood sugar, is a condition that affects people with diabetes. When the body is unable to use insulin efficiently or produces insufficient amounts of it, diabetes develops. Consequently, glucose does not get absorbed by body cells but instead accumulates in the blood. Then, even with elevated blood glucose levels, the body's cells become energy-starved. Amputations, heart attacks, strokes, renal diseases, blindness, dental disorders, and other issues can result from long-term damage to blood vessels and neurons caused by high blood glucose. Depression, difficulties during pregnancy, aging-related loss of mobility, and an increased risk of developing other diseases are possible side effects of diabetes. Numerous liver disorders, such as fatty liver disease, cirrhosis, hepatocellular carcinoma, and acute liver failure, are linked to certain forms of diabetes. Global worry over hepatic fibrosis and associated end stage cirrhosis is growing. The permanent outcome of fibrous scarring is cirrhosis, in which the normal architecture of the liver is replaced by bands of fibrous tissue that link to one another. The three most frequent etiological variables that lead to cirrhosis are alcohol abuse, hepatitis B, and hepatitis C [2]. The virus that causes hepatitis C (HCV) is a single-stranded RNA avivirus that is transmitted by blood and codes for two envelope proteins, a capsid protein, and a few nonstructural proteins [3]. Compared to patients receiving transplants for other liver diseases, HCV patients had a higher risk of developing diabetes [4]. When considered collectively, these findings point to a potential pathogenic function for HCV in certain forms of diabetes.

A chronic HCV infection typically progresses slowly and without symptoms for several years after infection. While some individuals experience a prolonged illness with little discernible liver damage, others soon develop liver cirrhosis and may even develop hepatocellular carcinoma [5]. Hepatocellular carcinoma (HCC) and cirrhosis are mostly caused by chronic HCV infection. There may be an increase in alpha fetoprotein levels in this disease. The prevalence of the incidence of hepatocellular carcinoma is rising, and this trend is anticipated to last for some time [6]. Figure 1 illustrates that, among other cancer forms, liver cancer is the leading cause of mortality in Egypt.
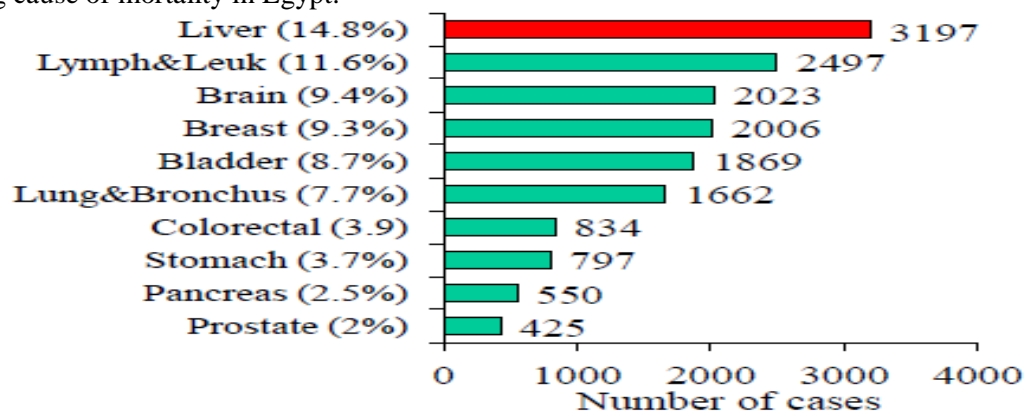

Fig. 1: Egypt Mortality Statistics.

The majority of HCC patients, according to current research, developed the disease as a result of a build-up of genetic anomalies, most likely brought on by external etiological factors, including HBV and HCV infections [7]. Risk factors such as aflatoxin can cause p53 mutations and DNA damage caused by the HBV genome incursion [8] can also cause mutations and damage to DNA sequences. Physician workload might be lessened by automatic classification tools. In the first phase of the two-phase process of data classification, the classifier algorithm builds a classifier using the training set of datasets. In the second phase, the model is used for classification, and its performance is examined using the testing set of datasets [9].

Filter methods and wrapper methods are the two main categories into which existing feature selection techniques can be divided. Independent of any classification techniques, filter approaches choose features based on certain discriminating criteria that depend on the properties of the data [10].The goodness of a subset of characteristics is assessed using the predicted accuracy of predefined classification algorithms in wrapper approaches [11, 12].

Taking into consideration the combinatorial impacts of features, the majority of wrapper techniques use complex multivariate machine learning algorithms like Support Vector Machines (SVMs). These have been demonstrated in numerous trials to be more effective than filter approaches in terms of classification accuracy [13]. Back up For many classification issues, vector machines have shown to be successful. SVM creates the best separating hyper plane with the maximum margin between the positive and negative classes for binary-class classification. With inequality constraints, it can be stated as a quadratic programming problem [14, 15].

SVMs are among the most promising machine learning algorithms, and there are numerous applications where they are effectively applied, such as bioinformatics, facial recognition, and text classification. SVMs deliver very good results on these data sets, typically outperforming other conventional methods [16].Statistics, learning theory, and engineering have all seen a huge increase in the use of SVMs [17, 18], as evidenced by the numerous references in these fields. Most support vector learning methods have been developed for binary issues, with a few notable exceptions. SVM generalization to multiclass issues has been attempted a few times [19].

Related Work

Numerous investigations have been carried out utilizing diverse machine learning techniques to forecast liver illnesses [20]. Using a clinical liver disease dataset, a framework for liver disease prediction is created. Regression analysis and hybrid feature selection are used to apply the described system. The model's diagnosis accuracy for liver disease was 89.21%.

A decision tree model was created by Hashem and Mabrouk [21] to forecast a patient's typical early stages of cirrhosis. Furthermore, the study contrasts the random forest (RF) model with the decision tree model. The simulation result demonstrates that the RF model outperforms the choice tree model in terms of accuracy, achieving a higher 70.67%. A model for liver disease diagnostics based on the support vector machine (SVM) method is presented in [22]. A clinical liver disease dataset gathered from the University of

Irvine UCI machine learning repository is used to train the SVM. The experiment's outcome demonstrates that the built SVM model can predict liver disease with promising results. Using the UCI data repository, the SVM model achieves a prediction accuracy of 73.2%.

Furthermore, a comparative study between SVM and an adaptive boosting algorithm for liver disease prediction was carried out by Afrin et al. [23]. The University of California, Irvine (UCI) data dataset's 583 liver disease samples are used to train the Support Vector Machine (SVM) and an adoptive boosting method. With a prediction accuracy of 74.65%, the simulation result demonstrates that adaptive boosting performs better than the SVM model. Similar to this, a comparative analysis of the UCI liver disease dataset is used in [24], [25] to examine the effectiveness of K-nearest neighbor (KNN), random forest, decision tree, and adoptive boosting algorithm. The outcome demonstrates that, when it comes to predicting liver disease, the decision tree model outperforms KNN, random forest, and adaptive boosting algorithms.

A comparative analysis of four machine learning algorithms—random forest, logistic regression, artificial neural network (ANN), and Naïve Bayes (NB)—was carried out by Geethaet and Arunachalam [26]. Accuracy is used as a performance criterion in the experimentation for the comparative examination of the four algorithms' performances. The outcome demonstrates that random forest beats logistic regression, ANN, and NB with the maximum accuracy of 84.29%. In terms of predicting liver disease, a comparison between SVM and logistic regression reveals that SVM performs better than logistic regression. With SVM, 75.04% prediction accuracy for liver disease is attained. Used NB and KNN to create a liver disease prediction model in [27] and [28]. The KNN model demonstrated a prediction accuracy of 72.5%, whilst the NB model demonstrated a prediction accuracy of 63.19%. Therefore, for the prediction of liver disease, the KNN model outperforms the NB model.

Additional research [29]–[31] demonstrates the significance of feature selection in enhancing the machine learning model's performance in the diagnosis of liver disease. The overlapping symptoms of a disease that are used to train the machine learning process can be interleaved with feature selection. Therefore, the researchers choose to use recursive feature removal to identify the important features for enhanced performance based on the literature. Additionally, SVM is used to train models since several researches [32]–[36] demonstrate the model's efficacy in multi-class classification tasks, such the prediction of liver disease.

It is evident from the literature review (section 2) that the impact of class imbalance on the anticipated accuracy was not taken into account in the earlier research. Even while earlier research has produced greater accuracy, this accuracy falls short of the class-wise performance of an ML model for the diagnosis of liver disease. Furthermore, the literature does not give the features that are more important to the various Machine Learning algorithms' learning process. In order to fill this gap, this work looks into liver disease diagnosis features that are important for support vector machine learning and uses the synthetic minority oversampling technique (SMOTE) to balance the dataset.

Proposed Work

In this section we proposed the enhanced method for the health care system to compute some performance parameters such as accuracy using classification and optimization techniques. Here we improve the efficiency rate in the terms of accuracy for the proposed system compare than existing system which is provide better results in the medical science domain.

Particle Swarm Optimization (PSO) method is types of swarm intelligence family methods which is particularly used for the optimization of results or improve the rate of any classification dataset which is used for various numbers of applications. Particle swarm optimization mostly used with four improves the results for any area mostly used with data mining methods and some other classification methods.

Here we describe proposed algorithm for the particular proposed heart based and other patients' dataset and they are following:-

1. Generate random population of n solution (particles)
2. For each individual i€N: calculate fitness (i)
3. Initialize the value of the weight factor, w
4. For i=1 to n particles
   Set pBest as the best position of particle i
   If fitness (i) is better than pBest then
   Set pBest (i) =fitness (i)
   End for i
5. Set gBest as the best fitness of all particles
6. For i=1 to n particles
7. Update the value factor of the weight w
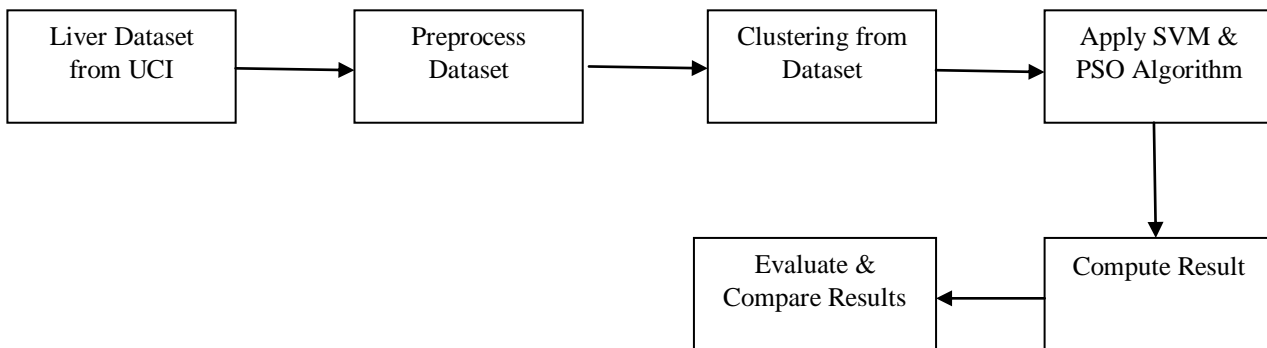8. Check if termination is true.
9. End

PROPOSED MODEL

Fig. 2: Proposed model.

Result Analysis

In this section, experimental process of we show that the comparative result analysis study for the Health care sector with disease diagnosis of Liver dataset is performed. This process of disease diagnosis of Liver dataset is done by using two methods that are support vector machine and PSO, the base paper method and proposed method. For the evaluation of performance parameter we used MATLAB software for the authenticity and effectiveness of results.

In this section we discuss about the dataset which we used for the diseases detection in the field of health care. Liver dataset is fetched from the UCI machine learning repository for the research purpose.

Result Parameters:
- **Precision-** Precision is defined as the proportion of projected positives/negatives that are really positives/negatives. Precision is measured in percentages.
- **Recall -** It is the ratio of the total positives and negatives that are projected to be positive and negative
- **Accuracy-**It is defined as the function of number of predictions that were right, or the percentage of cases that were properly categorized in a certain time period.

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

$$Accuracy = \frac{TP+TN}{TP+TN+FN+FP}$$

Comparative Result Analysis:

Dataset name	Method	Elapsed Time (Sec)	Accuracy in (%)	Precision in (%)	Recall in (%)
Liver	SVM	9.22	78.56	77.56	75.62
	OLD METHOD	10.24	80.78	78.24	74.38
	PROPOSED	11.36	94.56	88.56	89.62

Table 1: Comparative result analysis for the Liver dataset.

Comparative Result Graph
Figure 3: Show that the comparative result analysis for the Liver disease diagnosis using SVM and PSO techniques, our empirical result study shows that better accuracy than existing methods.
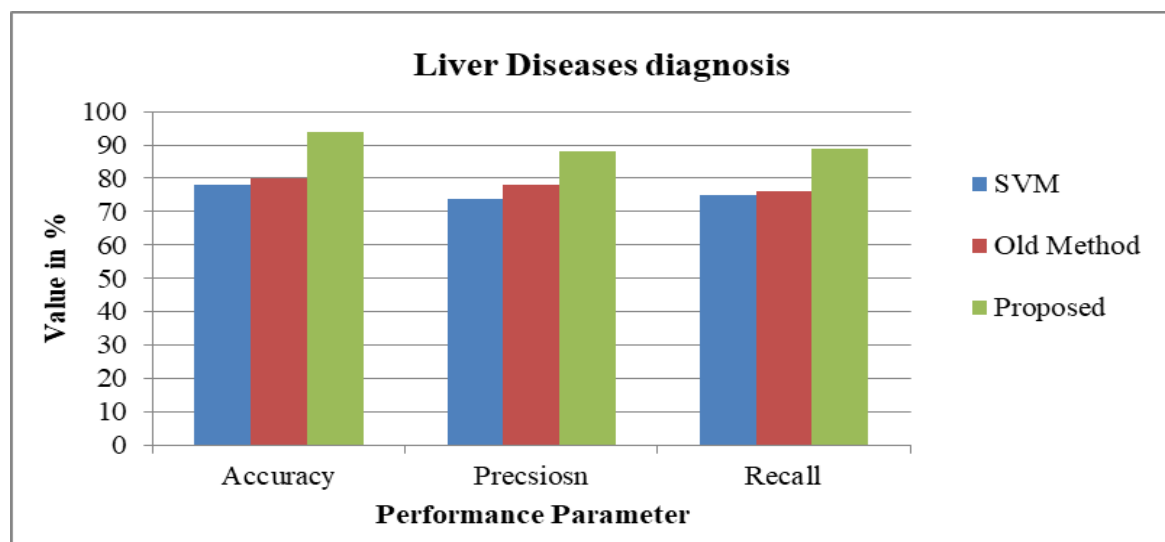
Fig. 3: Comparative result analysis for the Liver disease diagnosis.

Conclusion

In this research paper, we have proposed an enhanced method for the healthcare system using Particle Swarm Optimization (PSO). The proposed method is able to improve the efficiency rate in terms of accuracy compared to the existing system. The proposed method has been evaluated on a Liver dataset and the results show that it outperforms the existing methods in terms of accuracy, precision, and recall. The proposed method has the potential to be used in a variety of healthcare applications, such as disease diagnosis, treatment planning, and patient monitoring. It can help to improve the quality of healthcare services and reduce the costs of healthcare delivery. The proposed method has several potential implications for healthcare practice. First, it can help to improve the accuracy of disease diagnosis. This is important because accurate diagnosis is essential for effective treatment. Second, the proposed method can help to improve the efficiency of treatment planning. This is important because it can help to reduce the time and cost of healthcare delivery. Third, the proposed method can help to improve the quality of patient monitoring. This is important because it can help to identify and address potential complications early on. One limitation of the proposed method is that it has only been evaluated on a single dataset. In the future, it is important to evaluate the proposed method on other healthcare datasets to assess its performance on a wider range of diseases and conditions. Additionally, it is important to develop a hybrid PSO algorithm that combines PSO with other optimization algorithms to further improve the performance of the proposed method. Finally, it is important to develop a software tool that implements the proposed method and makes it easy to use for healthcare professionals.

References

[1] G.Compean D, J.Quintana , "M.Garza .Hepatogenous diabetes: current views of an ancient problem." Ann Hepatol ,pp 8:13,vol.20, 2009.

[2] K. Golla, J B. Epstein,and J. Robert. "Liver disease: Current perspectives on medical and dental management". Medical management update, vol. 98 , No. 5 ,November 2004.

[3] J.W. Little, D.A. Falace, C.S. Miller, N.L. Rhodus. "Dental management of the medically compromised patient." 5th ed. St. Louis: Mosby, 1997.

[4] H. Knobler, A .S.Green, S .Wallenstein, M .Schwartz, SH. Roman. "Higher incidence of diabetes in liver transplant recipients with hepatitis C." J Clin Gastroenterol pp.26:30 ,vol.33, 1998.

[5] T.J. Liang, B .Rehermann, L.B. Seeff, J.H. Hoofnagle. "Pathogenesis, natural history, treatment, and prevention of hepatitis C.". Ann Intern Med, pp132:296,vol.305 ,2000.

[6] P.J. Johnson. "Hepatocellular carcinomaa: is current therapy really altering outcome". Gut, pp51:459, vol.62, 2002.

[7] P .Pineau, A. Marchio, C. Battiston, E .Cordina, A .Russo, B. Terris and LX .Qin et al. ''Chromosome instability in human hepatocellular carcinoma depends on p53 status and aflatoxin exposure. '' Mutat Res, vol.653 (1-2), pp. 6–13, 2008.

[8] K. Saigo, K. Yoshida, R. Ikeda, Y .Sakamoto, Y. Murakami, T. Urashima and T. Asano et al. ''Integration of hepatitis B virus DNA into the myeloid/lymphoid or mixed- lineage leukemia (MLL4) gene and rearrangements of MLL4 in human hepatocellular carcinoma. '' Hum Mutat, vol.29 (5), pp. 703–708, 2008.

[9] Mitchell TM. Machine learning. Boston, MA: McGraw-Hill, 1997.

[10] C.Ding and H. Peng. ''Minimum redundancy feature selection from microarray gene expression data''. In CSB '03: Proceedings of the IEEE Computer Society Conference on Bioinformatics, pp 523, 2003.

[11] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik. "Gene selection for cancer classification using support vector machines." Machine Learning,vol.46(1-3),pp 389– 422, 2002.

[12] K. B. Duan, J. C. Rajapakse, H. Wang, and F. Azuaje. "Multiple svm for gene selection in cancer classification with expression data." IEEE TransNanobioscience, vol.4(3),pp 228–234, September 2005.

[13] H. Chai and C. Domeniconi. "An evaluation of gene selection methods for multi- class microarray data classification." In Proceedings of the Second European Workshop on Data Mining and Text Mining in Bioinformatics, pp3:10, 2004.

[14] C. J. Burges. "A tutorial on support vector machines for pattern recognition." Data Miningand Knowledge Discovery,vol 2(2),pp121:167, 1998.

[15] N. Cristianini and J.S. Taylor. "Support Vector Machines and other Kernel-based Learning Methods". Cambridge University Press, 2000.

[16] K. P. Bennett and C. Campbell. "Support vector machines: hype or hallelujah" SIGKDD Explor.Newsl ,vol.2,pp 1:13, 2000.

[17] V. N. Vapnik. ,Statistical Learning Theory. Wiley, 1998 .

[18] B. Sch¨olkopf, C. Burges, and A. Smola .", Advances in Kernel Methods,Support Vector Learning," 1998

[19] J.Weston and C. Watkins. "Support vector machines for multi-class pattern recognition." In Proceedings of the Seventh European Symposium on Artificial Neural Networks, pp219:224, April 1999.

[20] N. Nahar and F. Ara, "Liver Disease Prediction by Using Different Decision Tree Techniques," International Journal of Data Mining & Knowledge Management Process, vol. 8, no. 2, pp. 1–9, 2018, doi: 10.5121/ijdkp.2018.8201.

[21] E. M. Hashem and M. S. Mabrouk, "A Study of Support Vector Machine Algorithm for Liver Disease Diagnosis," American Journal of Intelligent Systems, vol. 4, no.1, pp. 9–14, 2014, doi: 10.5923/j.ajis.20140401.02.

[22] D. Devikanniga, A. Ramu, and A. Haldorai, "Efficient Diagnosis of Liver Disease using Support Vector Machine Optimized with Crows Search Algorithm," EAI Endorsed Transactions on Energy Web, vol. 20, no. 29, 2020, doi: 10.4108/eai.13-7-2018.164177.

[23] S. Afrin et al., "Supervised machine learning based liver disease prediction approach with LASSO feature selection," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 6, pp. 3369–3376, 2021, doi: 10.11591/eei.v10i6.3242.

[24] R. A. Khan, Y. Luo, and F.-X. Wu, "Machine learning based liver disease diagnosis: A systematic review," Neurocomputing, vol. 468, pp. 492–509, 2021, doi: 10.1016/j.neucom.2021.08.138.

[25] C.-C. Wu et al., "Prediction of fatty liver disease using machine learning algorithms," Computer Methods and Programs in Biomedicine, vol. 170, pp. 23–29, 2019, doi:.1016/j.cmpb.2018.12.032.

[26] C. Geethaet and A. R. Arunachalam, "Evaluation based Approaches for Liver Disease Prediction using Machine Learning Algorithms," 2021 International Conference on Computer Communication and Informatics (ICCCI), 2021, doi: 10.1109/ICCCI50826.2021.9402463.

[27] H. Hartatik, M. B. Tamam, and A. Setyanto, "Prediction for Diagnosing Liver Disease in Patients using KNN and Naïve Bayes Algorithms," 2020 2nd International Conference on Cybernetics and Intelligent System (ICORIS), 2020, doi: 10.1109/ICORIS50180.2020.9320797.

[28] F. Mostafa, E. Hasan, M. Williamson, and H. Khan, "Statistical Machine Learning Approaches to Liver Disease Prediction," Livers, vol. 1, no. 4, 294–312, 2021, doi: 10.3390/livers1040023.

[29] S. J. Sushma, T. A. Assegie, D. C. Vinutha, and S. Padmashree, "An improved feature selection approach for chronic heart disease detection," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 6, December 2021, pp. 3501–3506, doi: 10.11591/eei.v10i6.3001.

[30] M. H. Arif, A.-R. Hedar, T. H. A. Hamid, and Y. B. Mahdy, "SS-SVM (3SVM): A New Classification Method for Hepatitis Disease Diagnosis," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 4, no. 2, pp. 53–58, 2013, doi: 10.14569/IJACSA.2013.040208.

[31] N. Razali, A. Mustapha, M. H. Abd Wahab, S. A. Mostafa, and S. K. Rostam, "A Data Mining Approach to Prediction of Liver Diseases," Journal of Physics: onference Series, vol. 1529, pp. 1–7, 2020, doi:10.1088/1742-6596/1529/3/032002.

[32] S. Perveen, M. Shahbaz, K. Keshavjee, and A. Guergachi,"A Systematic Machine Learning-Based Approach for the Diagnosis of Non-Alcoholic Fatty Liver Disease Risk and Progression," Scientific Reports, vol. 8, no. 1, 2018, doi: 0.1038/s41598-018-20166-x

[32] C.-L. Liu, R.-S. Soong, W.-C. Lee, G.-W. Jiang, and Y.-C. Lin, "Predicting Short-term Survival after Liver Transplantation using Machine Learning," Scientific Reports, vol. 10, pp. 1–10, 2020, doi: 10.1038/s41598-020-62387-z.

[33] S. Ambesange et al., "Optimizing Liver disease prediction with Random Forest by various Data balancing Techniques," 2020 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM), 2020, doi: 10.1109/CCEM50674.2020.00030.

[34] K. Thirunavukkarasu, A. S. Singh, Md Irfan, and A. Chowdhury, "Prediction of Liver Disease using Classification Algorithms, 2018 4th Int. Conference on Computing Communication and Automation (ICCCA), 2019, doi: 10.1109/CCAA.2018.8777655.

[35] T. A. Assegie, "Support Vector Machine and K-nearest Neighbor Based Liver Disease Classification Model," Indonesian J. of Electronics Electromedical Engineering and Medical Informatics, vol. 3, no. 1, pp. 9–14, Nov. 2020, doi: 10.1234/jeeemi.v1i1.9xx.

[36] J. H. Joloudari, H. Saadatfar, A. Dehzangi, and S. Shamshirband, "Computer-aided decision-making for predicting liver disease using PSO-based optimized SVM with feature selection," Informatics in Medicine Unlocked, vol. 17, 2019, doi:10.1016/j.imu.2019.100255.