Volume 3, Issue 1, 2025

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

EARLY DIAGNOSIS OF HEPATIC DISORDERS USING MACHINE LEARNING

Ahmed Khoso[*1], Hadia Rehman[2], Sasuee Rajper[3], Sarfaraz Shahani[4], Danish Rathor[5]

[*1,2, 3,4,5]*Institute of Biomedical Engineering & Technology, Liaquat University of Medical & Health Sciences, Jamshoro.*

[1]ahmedbaloch181@gmail.com

DOI: https://doi.org/10.5281/zenodo.15141055

Keywords
Machine Learning, classification model, Healthcare, liver disease

Article History
Received on 24 February 2025
Accepted on 24 March 2025
Published on 31 March 2025

Corresponding Author: *

Abstract

In several nations, liver illnesses are now the deadliest illness. The number of people suffering from liver disease has been rising due to drug and alcohol abuse, inhaling toxic gasses, consuming contaminated food, and drinking alcohol. The goal of studying liver patient datasets was to create classification models that can accurately predict liver disease. By applying prediction and classification algorithms such as KNN and SVM to the publicly available patient liver dataset, we significantly reduced the effort required by the healthcare practitioners. By utilizing a dataset that included ten variables—such as age, gender, and other biochemical parameters we sought to precisely categories people as either liver patients or non-liver patients. The dataset included 7904 records for liver patients and 3173 records for non-liver patients generated from the global cohort of liver patients. We implemented two machine learning algorithms to test which gives more accurate results. Our investigation found that the KNN algorithm outperformed the SVM technique in predicting liver disease. Utilizing a proximity-based methodology, KNN exhibited superior ability in identifying underlying patterns within the dataset, leading to more accurate predictions. To sum up, our study demonstrated that KNN achieved higher accuracy of 0.89, recall of 0.95, and F1-score of 5508. These findings highlight KNN's superior reliability and precision in liver disease prediction using the given dataset and features. This means quicker findings, hence leading to better ill care and treatment plans. Through utilizing it doctors can make more informed judgements and enhance results for their patients.

INTRODUCTION

In the human body, liver is considered as the main organ, which plays a main a central role in several bodily functions. The production of glucose processing waste products, producing protein removing worn-out cell, blood clotting to cholesterol, and iron metabolism are the core functions of the liver. According to World Health Organization (WHO) and World Gastroenterology Organization (WGO), each year 35 million deaths occur due to liver failure, these deaths can be prevented by early prompt actions in the right direction. Its necessity of time because liver is responsible for various essential functions like it breaks down bilirubin by a process called glucuronidation, which further helps its defecation into bile (Regev & Schiff, 2001). Furthermore, it has extremely important role in regulation toxic and allergic substances. We call it drug is divided into two enormous sections: the left estimate and the privileged portion. The gallbladder is situated close to the pancreas, beneath the liver. These organs, together with the liver, aid in the consumption and provision of nutrients. Its function

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

is to facilitate the passage of blood from the stomach into the remaining portion of the body by assisting the flow of wound materials. Liver diseases are brought on by damage to the liver or impairment of its functionality. An irregularity in the liver's function that leads to sickness is called liver disease. It performs a multitude of essential bodily functions, and when it is injured or infected, the body's ability to function normally may suffer. Liver disease is also referred to as hepatic disorder. This general term includes a variety of potential issues that hinder the liver's ability to carry out its designated functions. The effectiveness of the liver will be significantly decreased even if only 25% of it remains healthy and the remainder is impaired. (Pan et al., 2024). The largest hard structure in the human body, the liver is a well-designed gland because, among its many functions, it produces and secretes bile. The rib cage protects the liver, which is in the upright portion of the abdomen. Its two central lobes are filled with tiny lobules. There are two different blood source bases for the liver cells. (Yilmaz et al., 2024). The vein which delivers nutrients from the intestines is portal, while the hepatic artery transports oxygen-rich blood driven by the heart. The vein's primary function is to carry blood from all other organs to the heart; however, the portal vein allows nutrients from the digestive tract to enter the liver, where they are treated and purified, before entering the bloodstream. (Bhutto et al., 2024). The liver cells need certain chemicals, which the portal vein efficiently carries to produce the proteins, cholesterol, and glycogen required for regular bodily functions. Liver disease happening is a complex process that is influenced by several properties that determine an individual's vulnerability to the disease. Bodily mass index (BMI), concomitant disorders such as diabetes, gender, heritage, genetics, and environmental exposures (viruses, alcohol, nutrition, and toxins) are a few of them. Liver issues relate to a high mortality rate and are considered life- threatening diseases. The initial step in diagnosing liver problems is the standard urine and blood testing. (Burnt and E.M 2004). There are numerous activities that prompt liver maladies. The classifications are:

Infection: Viruses and parasites can infect the liver, compromising liver function and causing edema or inflammation. The virus that usually causes liver damage is mainly contracted by contaminated food, impure water, or being together with an infected person. It is transmitted through blood or sperm. People can contract liver infections such as hepatitis A, C, and B.

Immune system abnormality: Certain illnesses give the body's immune system the opportunity to target other body parts. There's also liver damage. One of these illnesses may be autoimmune hepatitis. Furthermore, primary sclerosing cholangitis and primary biliary cholangitis could be the cause.

Inheritance: A gene that you received genetically from both of your parents may lead to a build-up of different compounds in the liver that can damage the liver.

Cancer: Liver adenoma, bile duct cancer, and liver cancer are cancers that can cause liver diseases.

Others: Uncontrolled alcohol abuse, non-alcoholic fatty liver disease (NAFLD), specific medications or over-the-counter remedies, and specific unani (herbal) mixtures are the common causes.

Risk factors: Excessive alcohol consumption, being overweight, having type 2 diabetes, having tattoos or body piercings, injecting drugs with used needles, receiving blood transfusions, encountering foreign blood, unprotected sexual contact, being exposed to chemicals, and inheritance are all factors that may increase the risk of liver diseases. (Khan et al., 2022). A liver function test, or ALFT, is advised for the patient based on the symptoms. Millions of individuals worldwide are impacted by liver disease, which is a serious health problem. Improved patient outcomes and a lighter load on the healthcare system can result from early identification and precise classification of liver illnesses. (Jayaraman et al., 2024). Non-alcoholic fatty liver disease is a developing health problem that affects one-third of adults and a rising percentage of children in wealthy countries. The initial indication of the illness is an abnormal accumulation of triglycerides in the liver, which can progress to cirrhosis and liver cancer in certain individuals. Although overweight, insulin resistance, and non-alcoholic fatty liver disease (NAFLD) are

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

significantly corelated pathophysiology of NAFLD is still not well understood, and there are few treatment options available. (Mburu et al., 2023). Based on patient data, machine learning techniques have nevertheless shown promising results in the prediction and classification of liver diseases. These methods predict results and find patterns in large datasets by using complex algorithms to analyze and learn from them. The applicability of the machine learning methods about the hepatopathy diagnosis and prognosis would be increasing gradually to the further increase of validate and discounted medical costs.

Literature Review

Thus, today the human diseases are equally being experienced nowadays as were experienced in past decades. Liver diseases can be also compared to other serious diseases and it is possible to state that the cases of people with the liver diseases are increasing constantly. (Kumar et al., 2023). Machine learning utilizes several algorithms whose primary goal is to explore the information stored in data in form of parameters to enable it determine the nex relationship between the several parameters it's fed with, and this makes it very effective in the predicting pattern and results of liver diseases, where a multitude of factors including but not limited to age, gender, lifestyle, and inherited tendencies come into play. As (Andrade et al., 2012) pointed out, deciding trees, SVM and ANNs have shown a strong ability in predicting the status of liver diseases using large datasets that include patient's medical history, lab tests and demography. Another strength of using Machine Learning techniques in the prediction of liver diseases is that the training data in such cases are big and unstructured, which are not easy to analyze using conventional statistical methods. This capability helps in sorting useful information from large databases including the electronic health records EHRs as identified by (Shobhika et al., 2023). These improvements made through integration of ML into liver disease models include systems able to not only predict the probability of liver disease existence but also the most likely type of disease and stage of the disease's progression, helping clinicians make better decisions. Besides, more recent developments in a subfield of ML called deep learning have improved the precision of liver disease models. Specialised forms of deep learning known as CNNs have demonstrated potential for the analysis of imaging information including, for example, liver biopsies and ultrasound. According to Behera et al., (2023), these types of models enhance features from imaging data needed for beyond distiantial liver fibrosis, cirrhosis, and hepatocellular carcinoma; better than manually diagnostic methods.

Yet, some obstacles are still present with the application of the ML in liver disease prediction. The first one is the accuracy and accessibility of information. Several datasets are required to build high-quality and diverse datasets on which to train reliable ML applications, but these are often missing in the context of the liver because of privacy constraints and due to the variability of their manifestations across populations. Secondly, there is a problem of explainability of the results of ML solutions since clinicians have to know how the predictions were made to use them.Some authors such as Choudhary et al., (2021) agree with the call for the creation of models that are explainable that is, models that offer clear reasons behind their decisions in order to enhance their applicability in clinical practice.023). Machine learning makes use of data-driven models to identify patterns and correlations within complex datasets, making it particularly suited for the prediction of liver diseases, where multiple factors such as age, gender, lifestyle, and genetic predispositions interact. As stated by (Andrade et al., 2012), techniques like decision tree, support vector machines (SVM), artificial neural networks (ANN) have shown considerable performance metrics for the utopia of liver disease. Smooth performance of extensive medical history records of patient, biochemical tests, and demographic data. Another advantage of using ML in the prediction of liver disease is that it can learn from a big amount of data, which can be disorganized and difficult in the case of statistical approaches. This capability facilitates the process of mining EHR to derive useful insights from them, about which more from (Shobhika et al., 2023). Integrating ML into the liver disease early prediction models created more complex systems that may not only predict the probability of developing liver disease, but also the type and stage of the disease where this should be helpful to clinicians.

Frontier in Medical & Health Research

ISSN: (e) 3007-1607 (p) 3007-1593

Furthermore, the recent development in layering of deep learning, a category of ML, has improved the forecasting precision of the liver disease models even more. Convolutional neural networks have been tested in some imaging data, for instance, liver biopsies and ultrasound images, as explained by (Behera et al., 2023) such models are capable of learning features from imaging data that are important in diagnosis of liver fibrosis, cirrhosis and hepatocellular carcinoma better than other diagnostic methods.

Nonetheless, there are a few challenges which cropped up despite the improvements made in the application of ML for liver disease prediction. There is always a question about quality as well as the availability of the data used in a particular study. Large, variegated footsteps of fine-quality data are prerequisite for the formation of most excellent ML models, however, the data of this type are generally limited for numerous reasons including data privacy and the variability of patterns of liver diseases within and across different populations. Further, the interpretability of developed ML models is also a challenge because clinicians require comprehensible explanations as to how those predictions are made to be able to trust and act accordingly. From such considerations, several scholars such as (Choudhary et al., 2021) argue that there is need to build post-hoc explanation methods that explain the ML models' decisions hence enhancing their applicability in clinical practice. Vast datasets containing patient medical history, biochemical tests, and demographic information. One of the key advantages of ML in liver disease prediction is its ability to handle large and unstructured datasets, which are often challenging for traditional statistical methods. This capability enables the mining of EHRs in order to derive pattern-of-care information as noted by (Shobhika et al., 2023). Integrating ML into liver disease prediction models means that systems with capabilities to predict the likelihood of the liver disease have been developed alongside the type and stage of the disease that a clinician has to treat. Additionally, the newest development in a specialized branch of ML, known as deep learning, has only improved the effectiveness of liver disease predictions. Deep learning algorithms, specifically the convolutional neural networks CNNs have found application in evaluation of imaging data including liver biopsies and ultrasound scans. According to (Behera et al., 2023) these models can learn from imaging data the important features for diagnosis of liver fibrosis cirrhosis and hepatocellular carcinoma better than conventional methods.

However, there are current gaps limiting the practical application of ML for liver disease prediction. One of the primaries has to do with the quality and accessibility of the information that is downloaded. Well-labeled diverse datasets are a critical ingredient of powerful ML algorithms, but such data are hard to come by because of patients' privacy and liver disease heterogeneity across various populations. Moreover, there is a problem of the interpretability of the built ML models as clinicians should be aware of how the prediction is done to be able to make decisions based on it. Researchers like (Choudhary et al., 2021) emphasize the importance of developing interpretable ML models that can provide clear explanations for their predictions, thereby increasing their utility in clinical setting.

Methodology

This paper suggested the adherence to the SVM and KNN algorithm, and as part of the technique that involves the prediction of liver diseases from the chosen dataset, the preparation of the dataset had to be done meticulously.

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

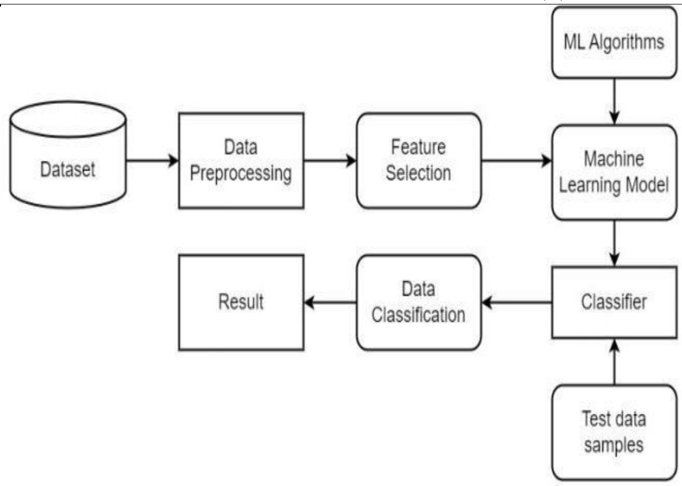


Figure 1: Processing of Supervised Machine Learning Algorithms for Early Diagnosis of Hepatic Disorders

In this study we initially loaded the dataset which then went for preprocessing, in this phase we made sure our dataset was relevant and clean. We managed outliers, standardized numerical features, and missing values. To make categorical variables compatible with machine learning techniques, we also encoded them. The dataset included 7904 records for liver patients and 3173 records for non-liver patients. It was generated from a global cohort of liver patients. Interestingly, anyone older than 89 were classified as "90" years old. Furthermore, we know in the field of machine learning, feature selection is crucial. This work applies feature selection to hundreds of samples of clinical data related to liver illness. There are three techniques for selecting features: filter, wrapper, and embedding techniques. The filter method is a frequently employed technique during the pre-processing phase. Another technique that cores the feature subset is the wrapper method. Lastly, the filter and wrapper methods are combined in the embedded method. Moreover, in order to understand the distribution and linkages within our dataset, we performed exploratory data analysis, or EDA. Trends, relationships, and anomalies were visualized. This process helps in comprehending the properties of the data and helps with decision-making on feature selection. Additionally, classification model was utilized to make the system learn from data and make predictions without direct programming. In this project we used KNN and SVM algorithms to predict liver diseases. K-Nearest Neighbors (KNN) and Support Vector Machines (SVM) are two popular machine learning algorithms that can be applied to liver disease analysis

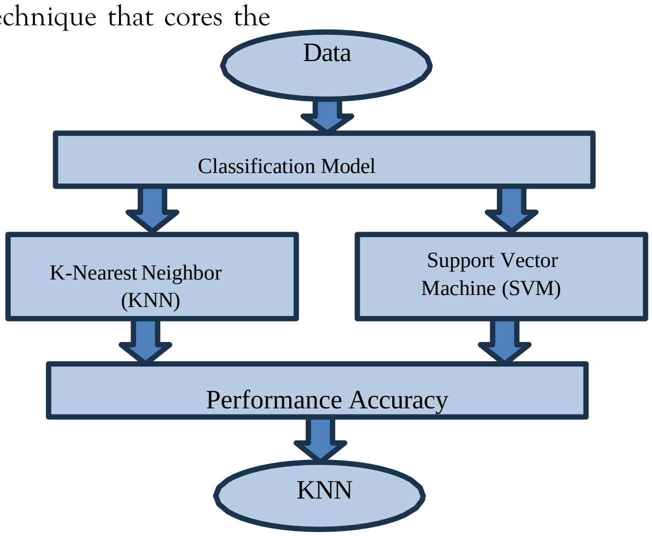


Figure 2: Illustrating comparison between SVM and KNN and showing the better Classification Algorithm as KNN

ISSN: (e) 3007-1607 (p) 3007-1593

KNN algorithm

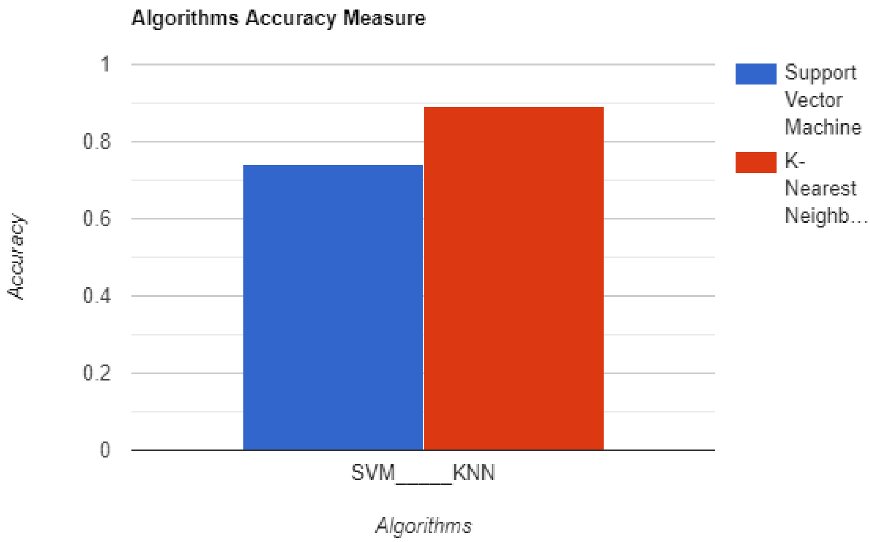
KNN is a pattern recognition supervised classification algorithm. Based on characteristics taken from their medical records, KNN can be used to categorize patients with liver disease into distinct groups (such as healthy and liver disease). The way it operates is by calculating the difference between the patient's data point and the dataset's known cases. The number of nearest neighbors to consider is represented by the "k" in KNN. (Loomba et al., 2009). A certain object is classified using the class label of majority of the k-nearest neighbors. Since KNN can be applied for evaluating the outcomes of the early liver diseases diagnosis by choosing patients' medical records similar to the records of the positively diagnosed patients, KNN can be useful when evaluating patients' testing and diagnosis.

SVM algorithm

SVM is a binary linear classification model that also includes regression. (Ou et al., 2024). In the case of liver disease, SVM is used to select a hyperplane that will partition patient records into sets. The main function of SVM is to try and establish the best hyperplane that shall enable maximum separation of different data points in two different classes. Since SVM is capable of mapping the data into a higher dimensional space through kernel functions so that separation is possible, then this algorithm is useful in such challenging cases where data cannot be separated through linear techniques. SVM is applicable for early predictions of the development of the disease or for sorting out the stages of liver disease. During analyzing the data pertaining to liver disease, KNN and SVM are the tools and which may be followed by the medical records and patient's data to make decision on diagnosis, prognosis and treatment Depending on the type of data and objectives chosen for the research work it is better to use which algorithm.

Results

Figure 3: Illustrating the result of both algorithms in bar chart form and KNN has a clear edge over SVM



Out of the five classifiers employed in the construction of the liver prediction model, K- Nearest Neighbors together with the Support Vector Machine classifiers, were selected for a deeper analysis. The results analyzed suggested that the accurate result of KNN was better as compared to the result of SVM. Standard Scaler was then applied to the features of the data set used in both training and testing of the model.

The result of classifiers showed that KNN had higher accuracy than that of the SVM. The classification report also revealed that, based on NAFLD data, KNN model had higher accuracy, higher recall, and higher F1-score for liver diseases predictions. This shows that the KNN algorithm produced a more accurate and dependable prediction of liver health within the specified dataset and feature set. According to figure 1, KNN has a greater accuracy in predicting liver

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

diseases than SVM algorithm. KNN has an accuracy of 0.89% while SVM has an accuracy of 0.74%.

Table 1 Showing the performance matrix for SVM algorithm

Metrics	Precision	Recall	Percentage	Support
0.0	0.92	0.54	0.68	2788
1.0	0.67	0.95	0.78	2720
Accuracy	-	-	0.74	5508
Macro Average	0.79	0.74	0.73	5508
Weighted Average	0.79	0.74	0.73	5508

Table 2 Showing the performance matrix for KNN algorithm

Metrics	Precision	Recall	Percentage	Support
0.0	0.94	0.84	0.89	2788
1.0	0.85	0.95	0.90	2720
Accuracy	~	~	0.89	5508
Macro Average	0.90	0.89	0.89	5508
Weighted Average	0.90	0.89	0.89	5508

Table 3 Illustrating comparison between the accuracy of both algorithms SVM and KNN

ML Models	Accuracy
Support Vector Machine	0.74
Nearest Neighborhood	0.89

Conclusion

In summary, we tested two methods, K nearest neighbors and support vector machines, to predict liver disease. Each method analyzed data about nearly ten variables to determine whether a person had liver disease. Our findings indicate that KNN was better at predicting the disease than SVM. This technique worked by comparing the dataset values internally, creating complex predictions that drew upon the patterns present in the dataset as it was form. We believe that a more practical use for K nearest neighbor would be in a hospital setting where rapid identification of liver disease would enable rapid medical intervention.

Acknowledgements

We want to give special thanks to our families and friends who have supported us throughout this project.

Conflict of interest

This is our own work; we have no conflicts of interest related to this study.

REFERENCES

Andrade, A., Silva, J. S., Santos, J., & Belo-Soares, P. (2012). Classifier Approaches for Liver Steatosis using Ultrasound Images. *Procedia Technology*, 5, 763–770. Retrieved 6 August 2024 from https://doi.org/10.1016/J.PROTCY.2012.09.084

Behera, M. P., Sarangi, A., Mishra, D., & Sarangi, S. K. (2023). A Hybrid Machine Learning algorithm for Heart and Liver Disease Prediction Using Modified Particle Swarm Optimization with Support Vector Machine. *Procedia Computer Science*, 218, 818–827. Retrieved 6 August 2024 from https://doi.org/10.1016/J.PROCS.2023.01.062

Frontier in Medical & Health Research
ISSN: (e) 3007-1607 (p) 3007-1593

Bhutto, J. A., He, Z., Najeeb, J., Naeem, S., Mahmoud, E. A., & Elansary, H. O. (2024). Data driven analysis of aromatase inhibitors through machine learning, database mining and library generation. *Chemical Physics*, 577, 112143. Retrieved 21 May 2024 from https://doi.org/10.1016/J.CHEMPHYS.2023.112143

Brunt, E. M. (2004). Nonalcoholic Steatohepatitis. *Seminars in Liver Disease*, 24(1), 3–20. Retrieved 25 July 2024 from https://doi.org/10.1055/S-2004-823098

Choudhary, R., Gopalakrishnan, T., Ruby, D., Gayathri, A., Murthy, V. S., & Shekhar, R. (2021). An Efficient Model for Predicting Liver Disease Using Machine Learning. *Data Analytics in Bioinformatics: A Machine Learning Perspective*, 443–457. Retrieved 6 August 2024 from https://doi.org/10.1002/9781119785620.CH18

Jayaraman, A., Jangalapuram, P. K., Ambayiram, C., & Alphonse, M. J. K. (2024). An optimal Bi-gated recurrent unit and ResNET using IoT for healthcare monitoring. *Expert Systems with Applications*, 247. Retrieved 25 July 2024 from https://doi.org/10.1016/J.ESWA.2023.122818

Khan, R. A., Luo, Y., & Wu, F. X. (2022). Machine learning based liver disease diagnosis: A systematic review. *Neurocomputing*, 468, 492–509. Retrieved 25 July 2024 from https://doi.org/10.1016/J.NEUCOM.2021.08.138

Kumar, A., Dev Mahato, K., Azad, C., & Kumar, U. (2023). Liver Disease Prediction Using Different Machine Learning Algorithms. *Proceedings - 2023 International Conference on Advanced and Global Engineering Challenges*, AGEC 2023, 118–123. Retrieved 25 July 2024 from https://doi.org/10.1109/AGEC57922.2023.00034

Loomba, R., Sirlin, C. B., Schwimmer, J. B., & Lavine, J. E. (2009). Advances in pediatric nonalcoholic fatty liver disease. *Hepatology (Baltimore, Md.)*, 50(4), 1282–1293. Retrieved 25 July 2024 from https://doi.org/10.1002/hep.23119

Mburu, A. N., Laving, A., Macharia, W. M., & Sande, J. (2023). Prevalence of non-alcoholic fatty liver disease in overweight and obese children seeking ambulatory healthcare in Nairobi, Kenya. *BMJ Open Gastroenterology*, 10(1), e001044. Retrieved 25 July 2024 from https://doi.org/10.1136/BMJGAST-2022-001044

Ou, J., Jiang, L., Bai, T., Zhan, P., Liu, R., & Xiao, H. (2024). ResTransUnet: An effective network combined with Transformer and U-Net for liver segmentation in CT scans. *Computers in Biology and Medicine*, 177. Retrieved 25 July 2024 from https://doi.org/10.1016/J.COMPBIOMED.2024.108625

Pan, A., Mukhopadhyay, S., & Samanta, S. (1AD). Liver Disease Detection: Evaluation of Machine Learning Algorithms Performances With Optimal Thresholds. *Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/IJHISI.299956*, 17(2), 1–19. Retrieved 25 July 2024 from https://doi.org/10.4018/IJHISI.299956

Qi, C., Yilmaz, E., & Chen, Q. (2024). Background of machine learning. *Machine Learning Applications in Industrial Solid Ash*, 93–130. Retrieved 21 May 2024 from https://doi.org/10.1016/B978-0-443-15524-6.00015-7

Shobhika, Kumar, P., & Chandra, S. (2023). Prediction and comparison of psychological health during COVID-19 among Indian population and Rajyoga meditators using machine learning algorithms. *Procedia Computer Science*, 218, 697–705. Retrieved 6 August 2024 from https://doi.org/10.1016/J.PROCS.2023.01.050

Regev, A., & Schiff, E. R. (2001). LIVER DISEASE IN THE ELDERLY. Gastroenterology Clinics of North America, 30(2), 547–563. Retrieved 14 August 2024 from https://doi.org/10.1016/S0889-8553(05)70195-3