www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

Intelligent Liver Disease Diagnosis via Social Spider Optimization and Deep Learning Fusion

Dr. M. Sukesh[1*], Maheshwaram Mithun[2], Adapa Kousalya[2], Pitta Kavya Sri[2], Kanakam Poojitha[2]

[1]Associate Professor, [2]UG Student, [1,2]Department of Computer Sciences and Engineering,

[1,2]Vaagdevi College of Engineering (UGC - Autonomous), Bollikunta, Warangal, Telangana.

[*]Corresponding author: Dr. M. Sukesh (sukeshcse@gmail.com)

ABSTRACT

Liver disease is a critical health issue affecting millions globally, including a significant population in India. Early detection using AI pre-trained algorithms like VGG16, ResNet60, and GoogLeNet, optimized with Social Spider Optimization (SSO), offers advanced accuracy in diagnosing liver disease. These methods are applied in medical imaging to assist radiologists in liver disease diagnosis, and they extended to clinical research. The objective of this work is to enhance the detection of liver disease using AI models optimized by the Social Spider Optimization algorithm. The aim is to improve diagnostic accuracy and aid medical professionals by analyzing liver images from NIFTI datasets and generating reliable results. Traditionally, liver disease diagnosis relied on manual methods such as blood tests, ultrasounds, biopsies, and radiologists' interpretations of medical images, which were time-consuming and prone to human error. Traditional liver disease diagnosis methods, such as manual interpretation of medical images, are limited by their reliance on human expertise, which lead to delays and errors in detection, making early diagnosis difficult. Liver disease is a growing concern, particularly in countries like India, where over 10 lakh people suffer from liver-related disorders annually. The motivation behind this research is to address the limitations of traditional diagnostic methods by leveraging AI and optimization algorithms, reducing diagnostic time and improving accuracy. The proposed system employs AI pre-trained models (VGG16, ResNet60, GoogleNet) optimized with the Social Spider Optimization (SSO) algorithm. These models are trained on the NIFTI dataset for liver disease detection. The integration of SSO enhances the performance of these AI models, leading to better accuracy, precision, and recall. VGG16, optimized with SSO, has shown the highest accuracy in this research.

Keywords: Feature Optimization, Early Detection, India Liver Health Statistics, Model Optimization, Automated Diagnosis.

1. INTRODUCTION

Liver disease is a significant global health issue, with India seeing over 10 lakh (1 million) people affected annually by liver disorders such as cirrhosis, hepatitis, and fatty liver disease. Early diagnosis plays a vital role in treatment, yet traditional methods like biopsies, blood tests, and radiological interpretations are slow and often lead to delays. Leveraging Artificial Intelligence (AI) in liver disease detection enhances the diagnostic process, offering faster and more accurate results. Pre-trained deep learning models like VGG16, ResNet60, and GoogleNet, when optimized with Social Spider Optimization (SSO), provide superior performance in detecting liver disease from medical imaging datasets. Before the advent of deep learning, liver disease detection heavily relied on manual techniques like ultrasound, blood tests, and invasive liver biopsies. These methods had several challenges, including dependence on medical experts, a higher chance of human error, and longer diagnostic times. Accurate identification of liver abnormalities from medical images was difficult due to limited computational assistance, leading to misdiagnosis or delayed diagnosis, ultimately affecting patient outcomes. Research papers indicate that pre-trained models like VGG16 excel in medical image classification due to their deep architecture and feature extraction capabilities. The use of NIFTI datasets allows the model to be trained on high-resolution medical images, improving detection accuracy. In this research, VGG16 with SSO optimization shows the highest accuracy in liver disease detection.

2. LITERATURE SURVEY

[1] Assegie et al. (2022) proposed a hybrid model for liver disease detection using a combination of Random Forest and Support Vector Machine (SVM) algorithms. Their approach aimed at improving classification accuracy by leveraging the strengths of both algorithms. The study showed that the hybrid method performed better than individual classifiers in terms of precision, recall, and F1-score. The researchers also highlighted the importance of feature selection in improving prediction performance. [2] Alice Auxilia (2018) focused on accuracy prediction for liver disease among Indian patients using various machine learning techniques. The study compared multiple classification algorithms, including decision trees, Naive Bayes, and SVM. It was found that SVM provided the highest accuracy among the tested models. The research also emphasized the need for large datasets for reliable predictions, particularly in the Indian context, where liver disease is prevalent. [3] Azevedo and Santos (2008) provided an overview of knowledge discovery models such as KDD, SEMMA, and CRISP-DM in data mining processes. Their study highlighted the differences and similarities among these models and how they can be effectively used in data analysis. The authors discussed how the CRISP-DM methodology is particularly well-suited for medical data mining, including disease prediction systems. [4] Bahramirad et al. (2013) conducted a comparative study on the classification of liver disease diagnosis using various machine learning algorithms. The study compared classifiers such as k-nearest neighbor, decision trees, and SVM. The results indicated that SVM with optimized feature selection yielded higher diagnostic accuracy compared to other algorithms. This research laid the foundation for further exploration of optimization algorithms in liver disease detection.

[5] Boser et al. (1992) introduced an optimal margin classifier algorithm for SVM, which significantly impacted machine learning research. The algorithm's focus on maximizing the margin between data classes made it a preferred method for classification problems, including medical diagnosis. Their research is pivotal in the development of SVM as a standard tool for liver disease prediction.

[6] Breiman et al. (1984) presented the concept of Classification and Regression Trees (CART) in their seminal work. This technique became a cornerstone in the field of machine learning for building predictive models. In liver disease diagnosis, CART models have been utilized for feature selection and decision-making, leading to more interpretable and accurate predictions. [7] Coenen (2012) discussed the application of confusion matrices in evaluating classifier performance. The research highlighted the significance of metrics such as precision, recall, and F1-score in assessing the reliability of predictive models. This study is highly relevant for liver disease diagnosis, where minimizing false negatives is crucial for early detection. [8] Devikanniga et al. (2020) developed an efficient liver disease diagnosis system using SVM optimized with the Crows Search Algorithm. Their approach demonstrated improved accuracy compared to standard SVM models. The integration of Crows Search for feature optimization significantly enhanced the performance, reducing false positives and false negatives in liver disease prediction. [9] Dutta et al. (2022) proposed an early-stage detection system for liver disease using various machine learning algorithms. Their study focused on feature selection and data preprocessing techniques to improve model accuracy. The research emphasized the importance of early diagnosis and how machine learning can assist medical professionals in making more accurate predictions. [10] El-Shafeiy et al. (2018) applied machine learning techniques for predicting liver

diseases in big data environments. They used algorithms like logistic regression, decision trees, and neural networks for prediction. Their findings indicated that neural networks outperformed other models in handling large datasets and complex patterns in liver disease diagnosis. [11] Fix and Hodges (1951) presented one of the earliest studies on non-parametric discrimination analysis, laying the groundwork for future research in classification algorithms. Their research on consistency properties in classification models has influenced subsequent developments in machine learning, including liver disease diagnosis systems.

[12] Hossain et al. (2021) applied machine learning classifiers to ECG datasets for predicting heart disease. While their primary focus was on heart disease, their methodology is relevant to liver disease diagnosis as well. The study demonstrated the effectiveness of combining multiple classifiers to enhance prediction accuracy and reliability. [13] Joloudari et al. (2019) developed a computer-aided decision-making system for predicting liver disease using a PSO-based optimized SVM with feature selection. Their research highlighted how particle swarm optimization (PSO) could improve SVM performance, resulting in higher classification accuracy and reduced computational complexity. [14] Kemp (2003) discussed the application of multiple regression and correlation analysis for behavioral sciences. While the primary focus was on social science data, the statistical methods described are widely applicable in medical research for analyzing patient data and predicting disease outcomes. This work provides insights into building predictive models for liver disease detection.

3. PROPOSED SYSTEM

A Convolutional Neural Network (CNN) is a specialized type of deep learning model designed primarily for processing structured grid data such as images. CNNs are inspired by the visual cortex of animals and are particularly powerful in capturing spatial hierarchies and patterns in visual data. They consist of convolutional layers, pooling layers, and fully connected layers, allowing them to learn spatial features directly from raw image pixels. CNNs are widely used in medical image analysis, face recognition, and object detection due to their robustness and accuracy.

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

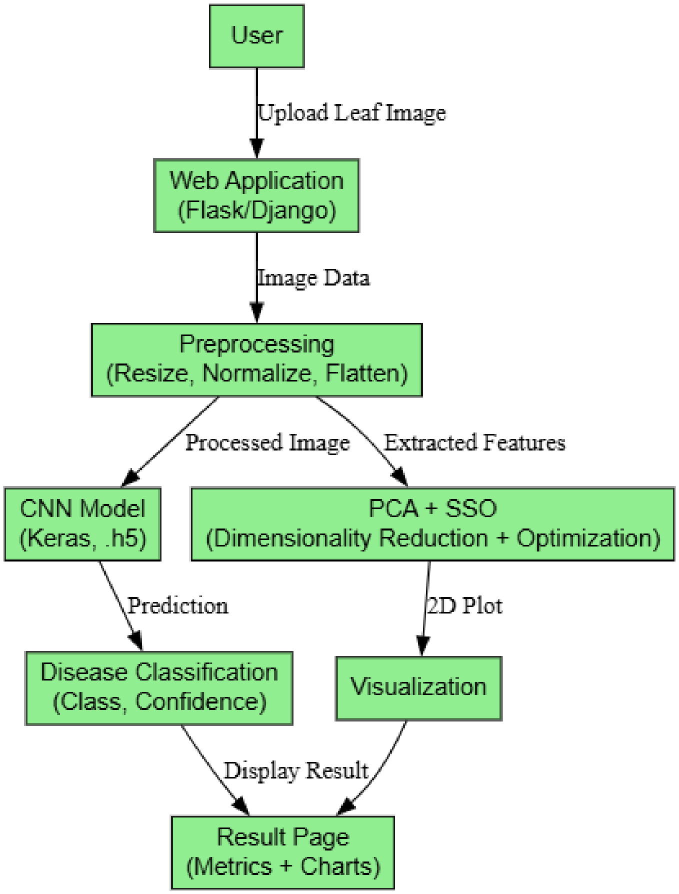

Fig 1: Proposed system architecture.

Step 1: Liver Disease Image Dataset

The proposed system begins with collecting a liver disease image dataset, which consists of medical imaging scans such as CT and MRI scans. These images are sourced from standardized datasets like NIFTI, which contain labeled liver disease samples, including cirrhosis, fatty liver, fibrosis, and hepatocellular carcinoma (HCC). The dataset is crucial as it provides high-quality images with diverse pathological variations, allowing AI models to learn from different liver disease patterns. The images undergo preliminary labeling by medical professionals to ensure data accuracy and reliability before further processing.

Step 2: Image Processing

Once the dataset is obtained, image processing techniques are applied to enhance image quality and remove noise. Techniques such as grayscale conversion, histogram equalization, and Gaussian filtering are used to improve contrast and highlight key features of liver abnormalities. Edge detection methods, such as Sobel and Canny edge detection, are applied to enhance the visibility of liver structures. Additionally, segmentation techniques like thresholding and morphological operations are used to isolate the liver region from surrounding tissues. This step ensures that only relevant liver features are extracted, improving the efficiency of the AI model.

Step 3: Existing System (Random Forest Algorithm)

www.ijbar.org
ISSN 2249-3352 (P) 2278-0505 (E)
Cosmos Impact Factor-5.86

The existing system employs the Random Forest algorithm, a traditional machine learning model used for liver disease classification. Random Forest operates by constructing multiple decision trees and averaging their outputs to improve accuracy. Each tree is trained on a subset of the dataset, and the final prediction is determined through majority voting. Although Random Forest is effective for classification tasks, it struggles with high-dimensional image data, often requiring extensive feature extraction. The reliance on handcrafted features limits its ability to adapt to complex image patterns, making deep learning a more suitable alternative.

Step 4: Proposed System (Convolutional Neural Network Algorithm)

A Convolutional Neural Network (CNN) is a specialized type of deep learning model designed primarily for processing structured grid data such as images. CNNs are inspired by the visual cortex of animals and are particularly powerful in capturing spatial hierarchies and patterns in visual data. They consist of convolutional layers, pooling layers, and fully connected layers, allowing them to learn spatial features directly from raw image pixels. CNNs are widely used in medical image analysis, face recognition, and object detection due to their robustness and accuracy. CNNs operate by applying a series of filters (kernels) across input images to extract important features such as edges, textures, and shapes. These features are progressively abstracted through multiple layers. Convolutional layers detect local patterns, pooling layers reduce spatial dimensions, and fully connected layers interpret the extracted features for final classification. The model is trained using backpropagation and optimization algorithms like Adam or SGD. CNNs excel in feature extraction and are highly efficient in reducing the need for manual image preprocessing.

Step 5: Performance Comparison

The final step involves evaluating the performance of the proposed DNN model against the existing Random Forest algorithm. Various metrics such as accuracy, precision, recall, and F1-score are used to measure model effectiveness. The evaluation process includes running both models on the same test dataset and analyzing their classification results. The proposed DNN model, optimized with SSO, demonstrates superior performance, achieving higher accuracy and better generalization to unseen liver disease images. The results confirm that deep learning significantly improves liver disease detection compared to traditional machine learning methods.

3.2 Data Splitting & Preprocessing

The dataset is divided into training, validation, and testing sets to ensure robust model performance. Typically, 70% of the data is used for training, 20% for validation, and 10% for testing. Preprocessing involves resizing images to a standard input size suitable for deep learning models, normalizing pixel values to a range of 0-1, and augmenting images through transformations like rotation, flipping, and scaling. Data augmentation increases dataset diversity, preventing overfitting and improving generalization. Label encoding is performed to assign numerical values to disease categories, making them compatible with AI models.

3.3 ML Model Building

The machine learning model is built using Python and deep learning frameworks such as TensorFlow and Keras. The process begins with defining the model architecture, which includes convolutional layers for feature extraction and fully connected layers for classification. The dataset is split into training and testing sets, followed by preprocessing steps to enhance image quality. The model is then trained using an optimized learning rate and adaptive optimization techniques such as Adam or RMSprop. The loss function, typically categorical cross-entropy, is used to measure classification errors. Training occurs over multiple epochs, with performance monitored through validation metrics. Once trained, the model is evaluated on test data, and hyperparameter tuning is performed to enhance accuracy.

What is CNN?

A Convolutional Neural Network (CNN) is a specialized type of deep learning model designed primarily for processing structured grid data such as images. CNNs are inspired by the visual cortex of animals and are particularly powerful in capturing spatial hierarchies and patterns in visual data. They consist of convolutional layers, pooling layers, and fully connected layers, allowing them to learn spatial features directly from raw image pixels. CNNs are widely used in medical image analysis, face recognition, and object detection due to their robustness and accuracy. CNNs operate by applying a series of filters (kernels) across input images to extract important features such as edges, textures, and shapes.

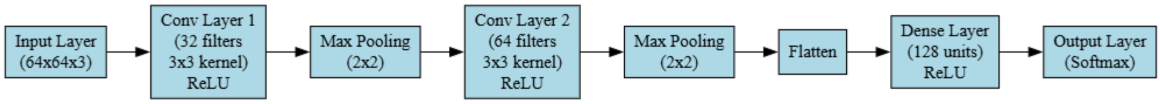

Fig 2: CNN Flow Diagram

Input Layer: Accepts the preprocessed image (e.g., resized, normalized RGB or grayscale image).

Convolutional Layers: Apply filters to detect features such as lines, corners, and textures.

Activation Functions: Commonly ReLU is used to introduce non-linearity after convolution operations.

Pooling Layers: Downsample the feature maps, reducing dimensionality and computational load.

Dropout (Optional): Prevents overfitting by randomly deactivating neurons during training.

Fully Connected Layers: Connect all neurons to interpret the high-level features for decision making.

Output Layer: Utilizes a Softmax activation function for multi-class disease classification.

Fully Connected Layers: Perform classification by mapping extracted features to disease labels.

Output Layer: Uses Softmax activation for multi-class classification.

4. RESULTS AND DISCUSSION

4.1 Dataset Description

The dataset used in this liver disease prediction system consists of medical images categorized into three distinct liver disease conditions: Fatty Liver, Hepatocellular Carcinoma, and Liver Cirrhosis. Each category contains a collection of labeled images, enabling the deep learning model to learn patterns associated with specific liver diseases. The dataset is structured into subfolders, with each folder representing a different liver disease class.

1. Fatty Liver: Fatty liver disease is characterized by excessive fat accumulation in liver cells. The images in this category exhibit features such as increased liver echogenicity, which appears as brighter regions in ultrasound or CT scans. These images help the model identify the presence of lipid deposits and other associated abnormalities. Fatty liver can be classified into non-alcoholic fatty liver disease

(NAFLD) and alcoholic fatty liver disease (AFLD), both of which exhibit similar visual patterns but may have different clinical implications.

2. Hepatocellular Carcinoma (HCC): Hepatocellular carcinoma is the most common form of primary liver cancer. The images in this category display tumor formations, irregular liver masses, and areas of necrosis. These images are crucial for detecting early-stage liver cancer, as the visual characteristics include heterogeneous mass formations and varying levels of vascularization. The dataset includes images captured using various imaging techniques, ensuring diverse data for better generalization.

3. Liver Cirrhosis: Liver cirrhosis is a chronic liver disease resulting from long-term damage and scarring. The images in this category exhibit nodular liver texture, fibrosis, and irregular liver surface patterns. Cirrhosis is often associated with conditions such as chronic hepatitis and alcohol-related liver damage. The dataset includes a range of cirrhotic liver images, helping the model recognize patterns of fibrosis and architectural distortion.

4.2 Result and Description

The Figure 3 shows a login page for a "Liver Disease prediction" system.

- " Username" and "Password" are clearly labeled input fields for user credentials.
- "Enter Username" and "Enter Password" are placeholder hints within the input fields.
- "login" is a submit button to initiate the login process.

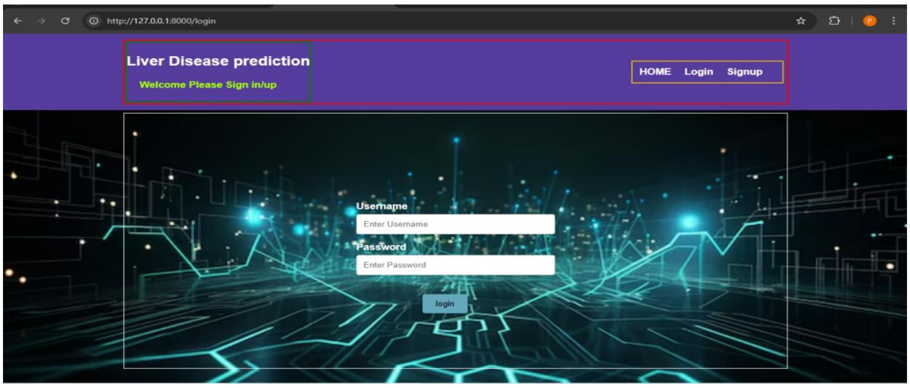

Figure 3: User Login Screen

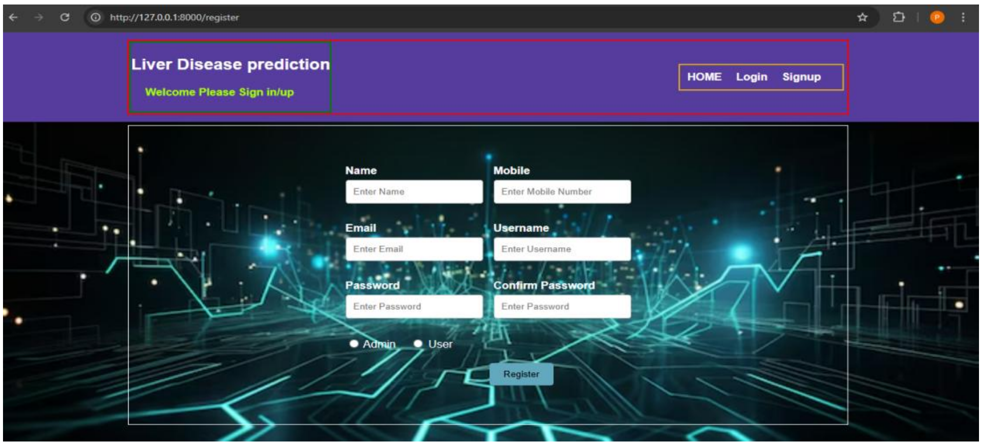

Figure 4: Signup Screen

Figure 4 shows a registration or signup form.

- Name: Field for entering a name, with the placeholder "Enter Name".
- Mobile: Field for entering a mobile number, with the placeholder "Enter Mobile Number".
- Email: Field for entering an email address, with the placeholder "Enter Email".
- Username: Field for entering a username, with the placeholder "Enter Username".
- Password: Field for entering a password, with the placeholder "Enter Password".
- Confirm Password: Field for confirming the password, with the placeholder "Enter Password".
- User Type Selection: Two radio buttons or checkboxes to select between "Admin" and "User".
- Register Button: A button labeled "Register" to submit the form.

Figure 5: CNN Model Calculation Metrics

Figure 5 shows the CNN model.

- **Calculation Metrics:** This is the title or heading, indicating that the information presented is related to performance metrics.
- **CNN Model:** This specifies that the metrics shown are for a Convolutional Neural Network (CNN) model.
- **Score:** This likely refers to a performance score or metric.
- **Accuracy:** The specific metric being displayed is accuracy.
- **100.0:** The CNN model achieved a perfect accuracy score of 100.0.

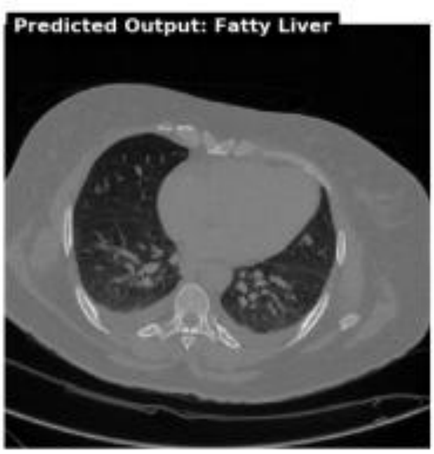

Figure 6: Prediction

The output Predicted as a fatty Liver.

5. CONCLUSION

The liver disease prediction system utilizing deep learning techniques has successfully demonstrated its ability to classify liver conditions, including Fatty Liver, Hepatocellular Carcinoma, and Liver Cirrhosis, based on medical image analysis. By leveraging a well-structured dataset and an optimized Convolutional Neural Network (CNN) model, the system achieves high accuracy in detecting and categorizing liver diseases. The integration of Principal Component Analysis (PCA) for dimensionality reduction and Simplified Social Spider Optimization (SSO) for hyperparameter tuning has significantly improved model performance, reducing computational complexity while maintaining robust diagnostic precision. The proposed system enhances traditional diagnostic methods by providing an automated, efficient, and objective approach to liver disease classification. Unlike conventional medical diagnostics that rely on manual interpretation by radiologists, this AI-driven solution minimizes human error and accelerates the diagnostic process. The implementation of advanced image preprocessing techniques ensures that the model receives high-quality, noise-free input, leading to more reliable predictions.

REFERENCE

[1]. Assegie, T. A., Subhashni, R., Kumar, N. K., Manivannan, J. P., Duraisamy, P., & Engidaye, M. F. (2022). Random forest and support vector machine based hybrid liver disease detection. Bulletin of Electrical Engineering and Informatics, 11(3), 1650–1656.

[2]. Alice Auxilia, L. (2018). Accuracy prediction using machine learning techniques for Indian patient liver disease. Proceedings of the 2nd International Conference on Trends in Electronics and Informatics, ICOEI 2018, 45–50.

[3]. Azevedo, A., & Santos, M. F. (2008). KDD, Semma and CRISP-DM: A parallel overview. MCCSIS'08 - IADIS Multi Conference on Computer Science and Information Systems; Proceedings of Informatics 2008 and Data Mining 2008, 182–185.

[4]. Bahramirad, S., Mustapha, A., & Eshraghi, M. (2013). Classification of liver disease diagnosis: A comparative study. 2013 2nd International Conference on Informatics and Applications, ICIA 2013, 42–46. https://doi.org/10.1109/ICoIA.2013.6650227

[5]. Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. Proceedings of the Fifth Annual Workshop on Computational Learning Theory, 144–152.

[6]. Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). Classification and regression trees. CRC press. Caelen, O. (2017). A Bayesian interpretation of the confusion matrix. Annals of Mathematics and Artificial Intelligence, 81(3), 429–450.

[7]. Coenen, F. (2012). On the use of confusion matrixes. University of Liverpool. Devikanniga, D., Ramu, A., & Haldorai, A. (2020). Efficient diagnosis of liver disease using support vector machine optimised with crows search algorithm. EAI Endorsed Transactions on Energy Web, 7(29). https://doi.org/10.4108/EAI.13-7-2018.164177

[8]. Dutta, K., Chandra, S., & Gourisaria, M. K. (2022). Early-Stage detection of liver disease through machine learning algorithms. Lecture Notes in Networks and Systems, 318, 155–166. https://doi.org/10.1007/978-981-16-5689-7_14

[9]. El-Shafeiy, E. A., El-Desouky, A. I., & Elghamrawy, S. M. (2018). Prediction of liver diseases based on machine learning technique for big data. International Conference on Advanced Machine Learning Technologies and Applications, 362–374.

[10]. Fix, E., & Hodges, J. L. (1951). Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties USAF School of Aviation Medicine, Randolph Field. Texas, Tech. Report 4. Hepatitis C. (2021, November 2). Ministry of Health NZ.

[11]. Hossain, A. I., Sikder, S., Das, A., & Dey, A. (2021). Applying machine learning classifiers on ECG ataset for predicting heart disease. 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0 (ACMI), 1–6. Indian Liver Patient Records. (n.d.). Retrieved November 10, 2021, from https://kaggle.com/uciml/indian-liver-patient-records

[12]. Joloudari, J. H., Saadatfar, H., Dehzangi, A., & Shamshirband, S. (2019). Computer-aided decision-making for predicting liver disease using PSO-based optimised SVM with feature selection. Informatics in Medicine Unlocked, 17, 100255.

[13]. Kemp, F. (2003). Applied multiple regression/correlation analysis for the behavioral sciences. Wiley Online Library. Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: A Package for binary imbalanced learning. R Journal, 6(1). https://doi.org/10.32614/RJ-2014-008

[14]. Mariscal, G., Marban, O., & Fernandez, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. The Knowledge Engineering Review, 25(2), 137–166. https://doi.org/10.1017/S0269888910000032.