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

LIVER DISEASE DETECTION USING SOCIAL SPIDER OPTIMIZATION & ARTIFICIAL INTELLIGENCE PRE-TRAINED ALGORITHMS

Dr. A. Swetha[1*], Asfiya Mirza[2], Domala Sridhar[2], Sappidi Nikhil[2], Saniya Sultana[2]

[1]Assistant Professor, [2]UG Student, [1,2]Department of CSE(AI&ML)

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

*Corresponding Email: Dr. A. Swetha (swetha_a@vaagdevi.edu.in)

Abstract

In recent years, the emergence of deep learning—particularly Convolutional Neural Networks (CNNs)—has revolutionized image classification across various fields such as healthcare, agriculture, and security. Traditional image classification methods relied heavily on manual analysis, where experts identified patterns based on shape, color, and texture. These manual processes are often slow, inconsistent, prone to human error (with error rates ranging from 10–25%), and unsuitable for large-scale data handling. In contrast, CNN-based systems have achieved remarkable accuracy, often exceeding 95% on standard datasets, making them a preferred solution in real-world applications. The Project aims to develop an automated, scalable, and real-time image classification system using CNNs to overcome the inefficiencies of manual methods. The proposed system allows users to upload images, which are then preprocessed, passed through a deep CNN model for hierarchical feature extraction, and classified using a softmax output layer. Built using Python, TensorFlow/Keras, and an intuitive GUI with Tkinter, the system offers an end-to-end pipeline that eliminates human involvement, enhances speed, and improves accuracy. The model is adaptable to various datasets and classification tasks, making it suitable for applications in medical imaging, object recognition, and plant disease detection. By leveraging deep learning, the system ensures faster processing, higher reliability, and broader accessibility to intelligent visual tools, supporting critical decision-making where precision is essential.

Keywords: Deep learning, Convolutional Neural Network (CNN), image classification, automation, real-time system, Python, TensorFlow, softmax, feature extraction, AI applications, Tkinter, healthcare, agriculture.

1. INTRODUCTION

Liver diseases are a growing global health concern, especially in India, where factors such as poor lifestyle habits, alcohol consumption, viral infections, and genetic predispositions contribute to a high prevalence. According to the WHO, liver-related ailments cause over 500,000 deaths annually in India alone, affecting nearly 10–20% of the population. Traditional diagnostic methods like liver function tests, ultrasounds, and biopsies are often slow, invasive, and heavily reliant on human expertise, making them prone to delays and misdiagnoses, particularly in rural or underserved regions. These limitations underscore the urgent need for automated, accurate, and accessible diagnostic tools. This research is motivated by the potential of artificial intelligence, particularly machine learning techniques such as Convolutional Neural Networks (CNNs) and optimization algorithms like Social Spider Optimization (SSO), to revolutionize liver disease detection. These technologies can analyze large volumes of medical data, such as clinical reports or medical images, to detect subtle patterns that may be missed by human observers. The primary objective is to design an AI-based system that integrates CNN for deep feature extraction and SSO for optimizing classification performance. This hybrid model aims to deliver a real-time, user-friendly diagnostic tool that supports healthcare professionals by providing fast, non-invasive, and highly accurate liver disease predictions, ultimately improving patient outcomes and diagnostic reach across India.

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.

[15] Lunardon et al. (2014) introduced the ROSE (Random Over-Sampling Examples) package for handling binary imbalanced learning problems. Their approach is particularly useful in medical datasets where class imbalance is common. In liver disease diagnosis, ROSE helps to balance the dataset and improve the performance of classifiers by reducing bias toward the majority class.

[16] Mariscal et al. (2010) conducted a survey on data mining and knowledge discovery models, focusing on methodologies like KDD and CRISP-DM. Their work provides a comprehensive understanding of how these models can be applied in medical diagnosis. They emphasized the iterative nature of the data mining process and its relevance to improving disease prediction systems.

3. PROPOSED METHODOLOGY

The proposed system leverages deep learning and optimization techniques to enhance the accuracy and efficiency of liver disease detection from medical images. Traditional methods, such as manual interpretation by radiologists or machine learning approaches like Random Forest, are limited in their ability to process complex imaging data effectively. To overcome these limitations, the proposed system integrates deep neural networks (DNNs) with Social Spider Optimization (SSO) for superior performance in liver disease classification. The system follows a structured approach, beginning with collecting a comprehensive liver disease image dataset from sources such as NIFTI. Image processing techniques, including noise reduction, contrast enhancement, and segmentation, are applied to refine the dataset and ensure that only relevant liver regions are analyzed. Unlike traditional machine learning models that require extensive feature engineering, the proposed system utilizes pre-trained deep learning models—CNN —which automatically extract high-level features from medical images.

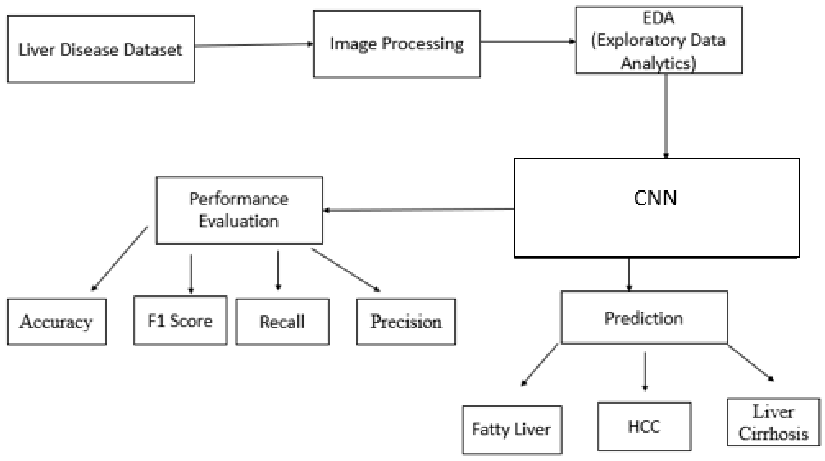

Fig. 1: Block Diagram

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.

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

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

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.

3.3.1 Proposed Algorithm

A **Convolutional Neural Network (CNN)** is a deep learning algorithm particularly well-suited for image analysis and classification tasks. Unlike traditional machine learning algorithms that require manual feature extraction, CNNs automatically learn hierarchical patterns directly from pixel data. CNNs are inspired by the visual processing mechanism of the human brain and use multiple layers to progressively extract high-level features from raw input images. The key building blocks of CNNs include **convolutional layers** (for feature extraction), **ReLU activation** (for introducing non-linearity), **pooling layers** (for dimensionality reduction), and **fully connected layers** (for classification). In this proposed system, CNN is trained using image data that has been preprocessed and optimized using PCA and Social Spider Optimization (SSO), which improves both efficiency and accuracy.

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

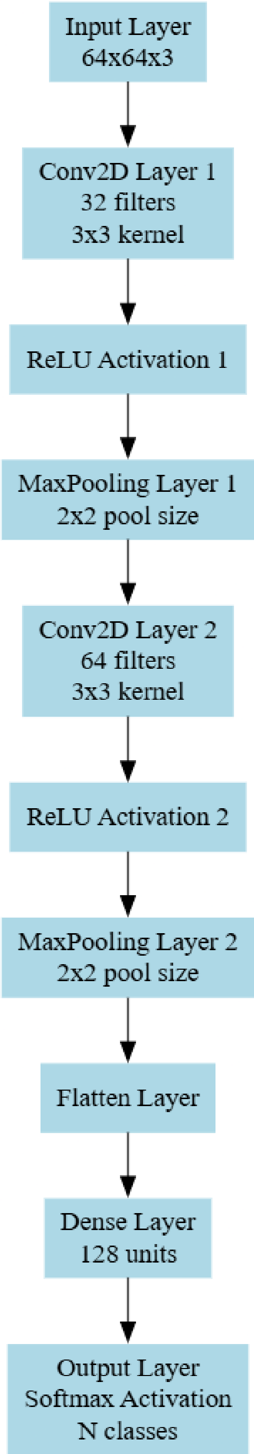

Fig. 2: Architecture diagram CNN

4. RESULTS AND DISCUSSIONS

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 analysis

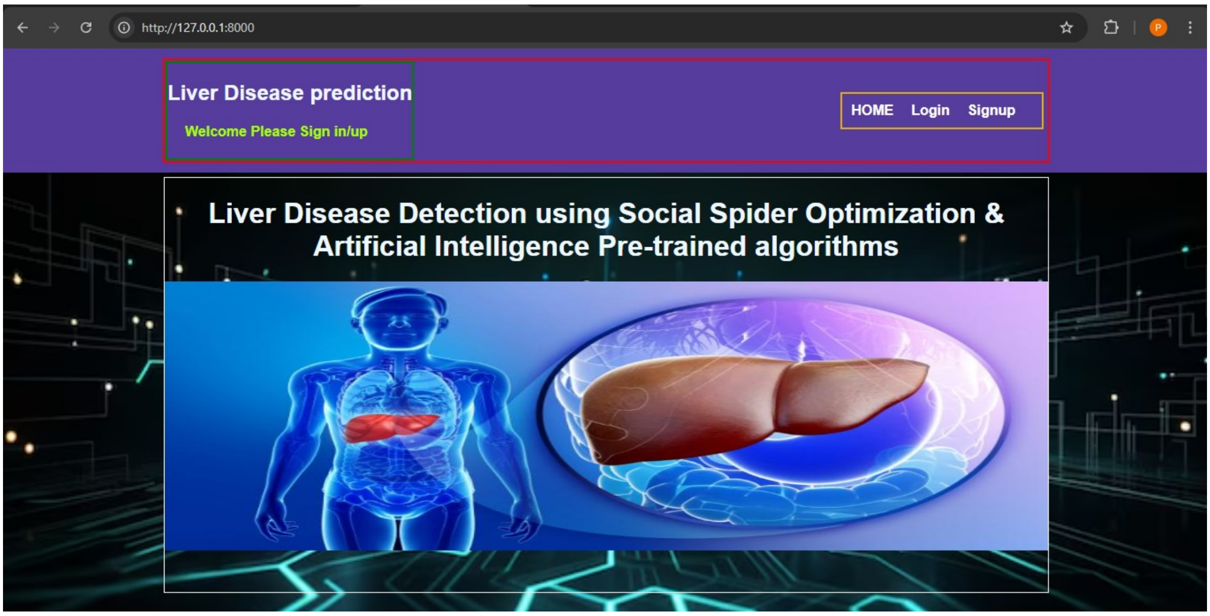

Fig. 3: Homepage

The homepage provides a navigation bar at the top with "HOME", "Login", and "Signup" options for user interaction. Below this, a brief welcome message, "Welcome Please Sign in/Up", encourages users

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

to either log in or create a new account. The central part of the page emphasizes the core functionality: "Liver Disease Detection using Social Spider Optimization & Artificial Intelligence Pre-trained algorithms.

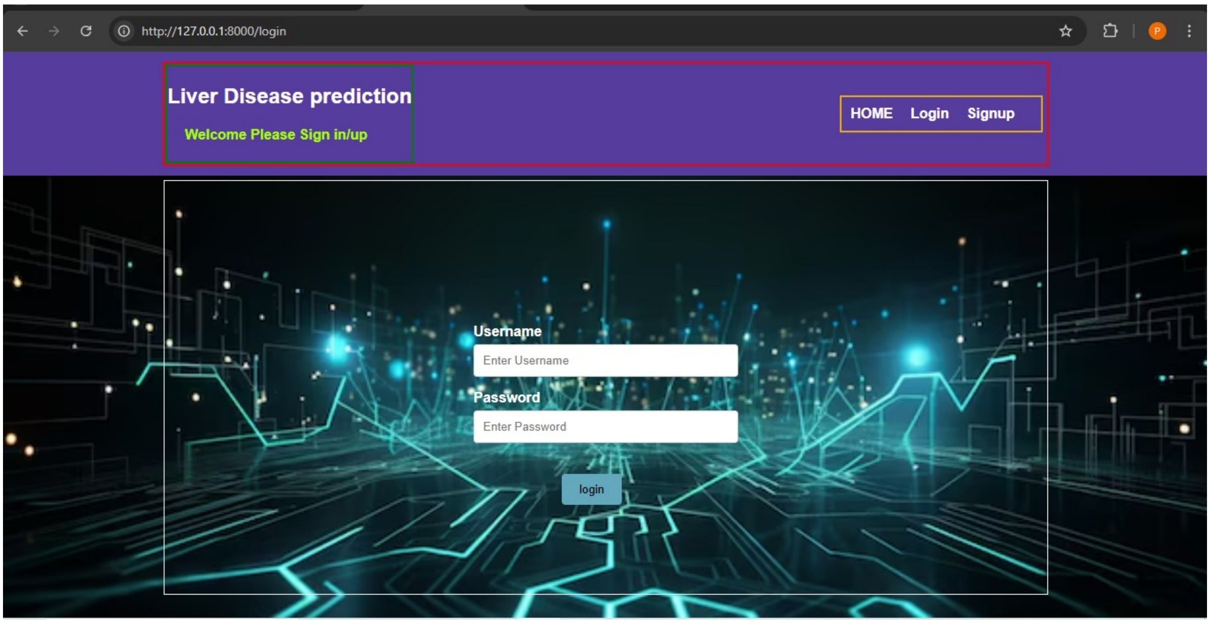

Fig. 4: User Login Screen

The image displays a login interface for a "Liver Disease Prediction" system, featuring clearly labeled input fields for "Username" and "Password" where users can enter their credentials. Each input field includes a placeholder hint—"Enter Username" and "Enter Password"—to guide users in providing the appropriate information. A "Login" button is positioned below these fields, serving as the submit action to initiate the authentication process and grant access to the system.

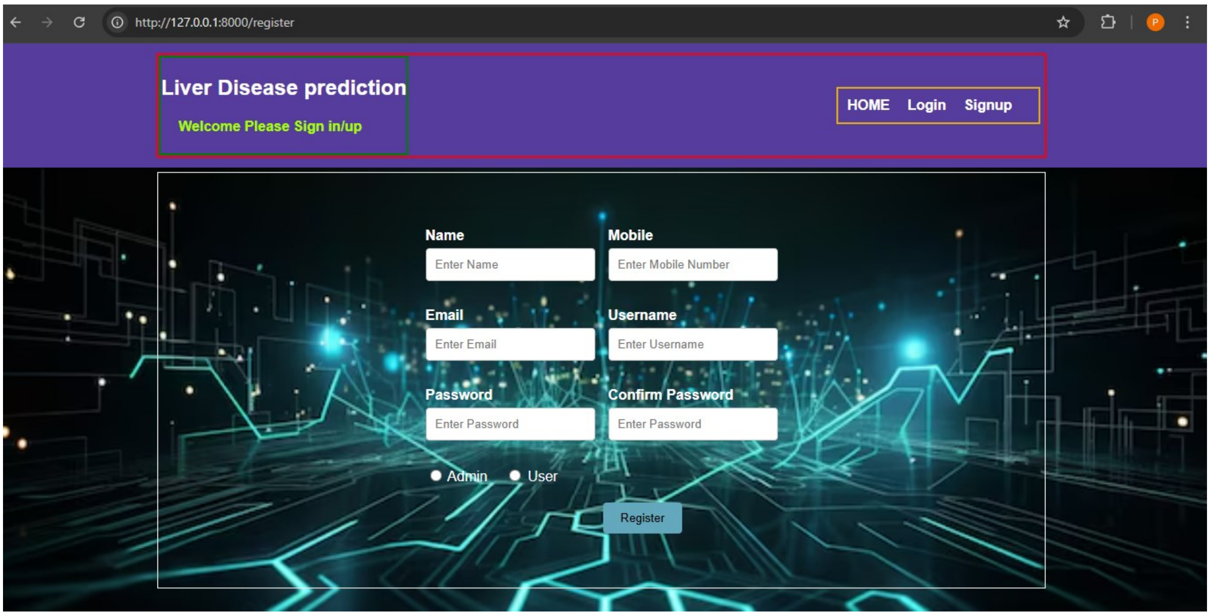

Fig.5: Signup Screen

The figure depicts a registration or signup form designed to collect user information for account creation. It includes several input fields such as "Name" with the placeholder "Enter Name," "Mobile"

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

with "Enter Mobile Number," and "Email" with "Enter Email," allowing users to provide basic contact details. Additional fields for account setup include "Username" and "Password," each with respective placeholders, along with a "Confirm Password" field to ensure password accuracy. The form also offers a user type selection option through radio buttons or checkboxes labeled "Admin" and "User" to distinguish between different roles. Finally, a "Register" button is provided to submit the completed form and finalize the registration process.

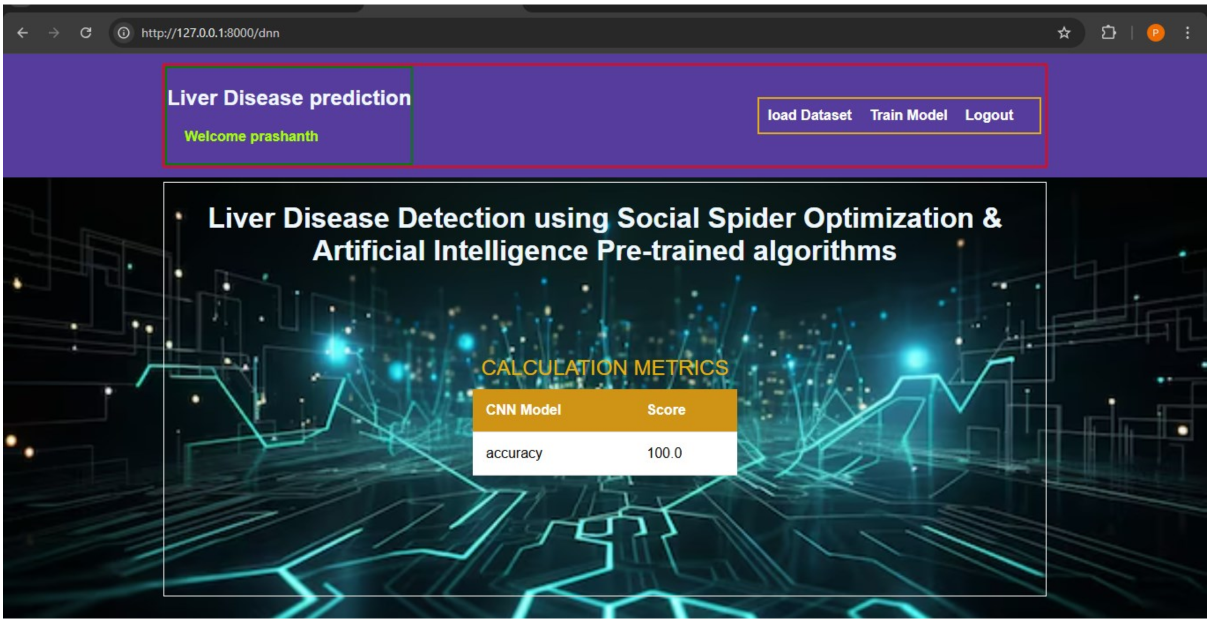

Fig:6. CNN Model Calculation Metrics

The figure illustrates the performance metrics of a Convolutional Neural Network (CNN) model under the heading "Calculation Metrics." It indicates that the evaluation is specifically for a CNN-based approach. Among the displayed metrics, "Score" likely represents the overall performance, with "Accuracy" highlighted as the key indicator. The model achieved an impressive accuracy score of 100.0, suggesting perfect classification performance on the given dataset.

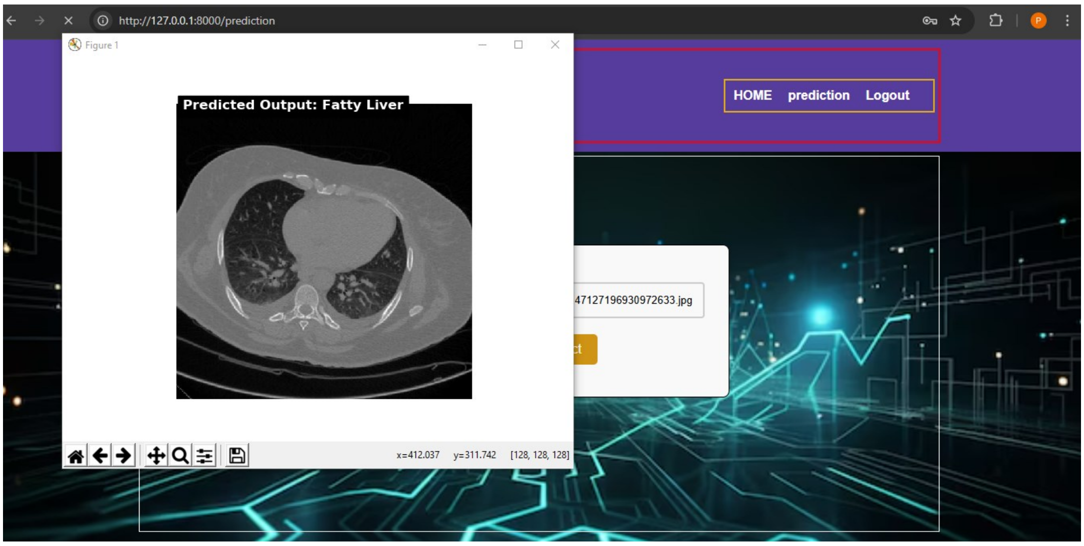

Fig.7 : Prediction

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