Journal of Intelligent Systems and Internet of Things
Vol. 14, No. 02, PP. 78-90, 2025
DOI: https://doi.org/10.54216/JISIoT.140207

Prediction and Classification of Fatty Liver Disease Using Probabilistic Neural Networks

Appanaboyina Sindhuja[1,*], Seetharam Khetavath[2]

[1]Research Scholar, Department of ECE, Chaitanya Deemed to be University, Hanamkonda, Warangal, Telangana, India

[2]Professor, Department of ECE, Chaitanya Deemed to be University, Hanamkonda, Warangal, Telangana, India

Emails: sindhu4b3@gmail.com; seetharamkhetavath@gmail.com

Abstract

Fatty liver disease, encompassing conditions like NAFLD (Non-Alcoholic Fatty Liver Disease) and NASH (Non-Alcoholic Steatohepatitis), is a significant global health issue linked to metabolic syndrome and increasing incidences of liver-related complications. Accurate and early detection of fatty liver illness is critical for effective intervention and management. This paper proposes a novel method for the prediction and arrangement of fatty liver disease using Probabilistic Neural Networks (PNNs), leveraging advanced machine learning techniques to enhance diagnostic accuracy and reliability. We developed a PNN-based model to classify liver conditions from a dataset comprising clinical and imaging features, including liver fat content, texture metrics, and demographic information. The PNN was chosen for its capability to handle complex, high-dimensional data and provide probabilistic outputs, which are crucial for assessing the likelihood of different disease stages and improving interpretability. The proposed methodology includes preprocessing steps to normalize and augment the data, followed by feature extraction using advanced techniques to capture relevant patterns. The PNN architecture was designed with multiple layers to process features and deliver class probabilities. The method's concert was estimated utilizing average system of measurement such as accuracy, precision, recall, and F1-score, demonstrating its efficacy in distinguishing between different stages of fatty liver disease. Experimental results indicate that the PNN model achieves high classification accuracy and outperforms traditional machine learning methods in detecting fatty liver illness. This study highlights the potential of PNNs in enhancing diagnostic processes and providing a robust tool for clinicians. Future work will concentrate on expanding the dataset, refining the model, and integrating it into clinical workflows to support better patient outcomes in liver disease management.

Received: March 12, 2024 Revised: June 18, 2024 Accepted: October 10, 2024

Keywords: Probabilistic Neural Networks; Machine Learning; Liver Disease; Fatty Liver

1. Introduction

NAFLD is a condition categorized by disproportionate fat gathering in the liver, occurring without significant alcohol consumption. It is a common condition as it has a prevalence in about 30-50% of the general population in industrialized nations. The increase in fat has harmful effects on the liver, which can eventually progress to inflammation or fibrosis. Fatty liver illness incorporates a range of liver illnesses, counting modest fatty liver illness (steatosis) and steatohepatitis. As fatty liver steatosis progresses to steatohepatitis, fatty liver disease may improvement to fibrosis and cirrhosis, thereby increasing the danger of hepatocellular carcinoma. Understanding the various components of fatty liver disease is a critical step for both clinicians and researchers, as misdiagnosis can lead to inefficacious treatment and serious health complications.

Current advances in both genomics and bioinformatics have facilitated the estimation of high dimensional heterogeneous genomic variables. However, existing frameworks predominantly concern gene expression levels or methylation states. Analyzing fatty liver disease by building a computer model would provide a head start for improving the current histology standards, thereby lowering costs for clinical studies. Popular supervised learning methods contain ANN, DT (decision trees), SVM (support vector machines) and LR (logistic regression). Among them, neural networks, single or multilayer perceptron's, are well-suited for developing statistical models of greater complexity. Furthermore, with the rise of computer hardware and software, brain-inspired artificial neural networks have emerged as one of the leading paradigms for intelligent systems. Probabilistic neural networks are a dedicated multilayer construction that has shown superior models to multilayer perceptron and radial basis function networks for some applications [1].

The online market is trending upwards as many companies emerge in the industry. One such NAFLD online center is Aspire Health. They currently offer a simple service on general information about fatty liver disease and what lifestyle changes to perform, but with cut-back information on the filters used in these decisions.

The liver is a vigorous organ that plays an essential part in absorption, detoxification, and insusceptible defense. Among other vital roles, it regulates the levels of cholesterol and triglycerides for energy maintenance and fat storage. Additionally, it prevents the accumulation of triglycerides in excess and obtains energy through fat oxidation. Fatty liver disease, known as hepatic steatosis, arises when phospholipid gathering in the liver tops 5–10% of the liver weight. This severe condition poses several immediate and long-term dangers to human health, including cirrhosis, liver cancer, and liver failure [2].

According to the WHO, the pervasiveness of fatty liver illness has drastically improved in the last few decades owing to lifestyle changes. In studies conducted in various countries such as the USA, Japan, Turkey, Taiwan, and Middle Eastern countries, it was revealed that the occurrence of fatty liver illness symptoms and NAFLD is between 25-35%, 30%, 25%, 19%, and 25%, respectively. Fatty liver disease is also common in diabetic patients who are more prone to lifestyle illnesses. Early detection of fatty liver disease is imperative to prevent the severity of the disease. Several noninvasive methods were devised for the early detection of fatty liver disease, including laboratory tests such as measuring ALT, AST, and GGT enzymes; molecular tests, X-ray-based electrographic tests; and imaging tests. Furthermore, several computational models have been developed for the prediction and classification of liver disorders. PNNs (Probabilistic neural networks) are a type of ANN used for classification problems. PNNs utilize a supervised learning algorithm, learn from a training set, and classify new inputs based on its similarity to training samples from different classes. PNNs are among the classical neural networks that use the Bayesian decision theory and can be employed for both regression and classification problems. PNNs are both valuable and powerful models for prediction and classification, with applications in diverse fields, including bioinformatics, disease diagnosis, econometrics, finance, geosciences, and pharmaceuticals [3].

Among the various diseases, the liver is most affected by fatty acid disease, which is caused due to fatty acid deposition in liver cells. The fatty liver disease mainly falls under two categories, alcoholic and non-alcoholic. Among the above types, the NAFLD is one of the maximum common illnesses in the world. The fatty liver does not showcase any symptoms until the disease reaches a higher stage. If it reaches a severe stage it turns into cirrhosis. It has been calculated that there are around more than 25% of the world population who are suffering from fatty liver disorder. Moreover, among those suffering from fatty disorders, around 25% of them are in the higher stages of liver fibrosis. All of the above statistics necessitate the need for a general screening for fatty liver disease. The patients suffering from fatty liver other than cirrhosis are often asymptomatic. Therefore, it is hard to detect and diagnose at the earlier time which is helpful for the treatment. The general physician also ignores this disease due to the asymptomatic nature. In this work, a model to classify and detect fatty liver disease is proposed. The model takes twelve lipid biomarkers as the input for classification. The lipids and the fatty liver disease are correlated. The Probabilistic Neural Network (PNN) is used as the classifier here. The model is tested on the publicly available database. The model is observed to give a better accuracy of classification by discriminating between the different levels of fatty liver disease.

All the public datasets are used as it is ethically wrong to use any other clinical datasets for research purposes. The public datasets do not contain any identity of the patient coded information set. The designed approach is also verified on a information set taken from the hospital. The public dataset contains patients who are diagnosed and those who are not, constraining the model to give accurate predictions. Moreover, only the lipid profiles of patients are taken which might not represent all the patients. Proper pre-processing of the datasets is not done which might lead to wrong estimations of the approach. Further, the model can be imbibed with the classifier which detects different types of liver diseases at the same time [4].

The goal of this work is to apply ANN methodologies applied to the Probabilistic Neural Network (PNN) to classify and predict the danger of people developing Fatty Liver Illness. Fatty Liver Illness is a rapidly growing concern that adversely affects people’s health and demands effective classification and/or prediction techniques.

A publicly available dataset of health parameters corresponding to individuals having or not Fatty Liver Disease is analyzed. The focus is centered on the application of the Probabilistic Neural Network classifiers (PNNs) because they are faster to train than conventional artificial neural networks, as no iterative training is needed. Nevertheless, many adjustable parameters conscientiously need to be chosen as described in the Literature. A thorough study and optimal parameters of the classification approach are suggested. This work proposes the use of a combination of classification and regression artificial neural networks; in the classification-oriented approach, two options are implemented: the Probabilistic Neural Network and the Multi-Layer Perceptron. The advantages, drawbacks, and classification outputs of different classification methodologies are provided. The best classifier is selected according to the trade-off between accuracy and simplicity [5].

Then, two numerical indices were defined to propose people at high danger of emerging Fatty Liver Disease according to easily obtainable data. The definition of risk thresholds was accomplished in the context of common practice in a hospital environment. The plausibility of the proposed methodologies is demonstrated through satisfactory results.

2. Literature Review

2.1. Fatty Liver Disease: Causes and Symptoms

NAFLD is a situation where extra fat builds up in the liver without any connection to alcohol consumption. It is currently the most common liver disorder globally. Fatty liver can lead to NASH, fibrosis, cirrhosis, and even HCC (hepatocellular carcinoma) along with CVD (cardiovascular disease). NAFLD is growing around the globe, mostly in industrialized or developing industrialized countries, chiefly due to the rapidly increasing epidemic of obesity and Type-2 diabetes. Progressive liver disease and its consequences may affect multiple life-threatening organs, increasing the risk to the diabetic patient. In this aspect, the liver is central to the development of cardiovascular disease risk because of its involvement in lipoproteins, glucose, and coagulation.

The potential causes range from dietary issues to different hereditary factors, including variations in genes that affect the sensitivity of tissues to insulin. NAFLD is frequently related with Type-2 diabetes and insulin resistance. Recent works have pointed to a linkage between the dysregulation of liver oxidative stress defenses and the progression of the fatty liver into chronic inflammation and the metabolic syndrome. Steatosis is defined as an increase in fat content (triglycerides) greater than 5% of the total weight of liver. Generally, it progresses through an inflammatory state (steatohepatitis), fibrosis, and cirrhosis. Fatty liver illness leads to the expansion of the tumour, as a secondary effect of cirrhosis.

It has been observed that several steatosis-associated quantitative trait loci have a link with diabetes and obesity. Hence, fatty liver is now being used in attempts to determine the diabetic mouse models. All types of fatty liver are similar with respect to a few pathophysiological processes, including excess fatty acid influx into the liver, decreased fatty acid corrosion and triglyceride secretion, improved de novo lipogenesis, and impaired hepatic insulin signaling. Steatosis is usually considered benign but it may progress into steatohepatitis, which may be irreversible, culminating in liver cirrhosis and may also lead to liver cancer. Diagnosis is established by a grouping of clinical, laboratory, and imaging techniques and with or without liver biopsy [6].

2.2. Machine Learning in Medical Diagnosis

Machine learning (ML) computes algorithms, both statistical and analytical, to ascertain patterns within vast datasets. This varies from processing natural language to identifying objects within images or video frames. For ML, prior human knowledge of the examined system is limited. Thus, the process is called "learning" because this knowledge is derived from the data processing by the algorithm. In the learning process and before processing the data, the researcher "trains" the algorithm to become accustomed to the data with known properties (input-output pairs) by adjusting coefficients in such a way that the difference between the expected response and the processed response tends to vanish. This pre-processed structure is then applied to analyze new data, for which no prior knowledge is available.

A common application in healthcare is the use of ML in medical diagnosis. For example, parameter sets derived from biochemical tests can be used to train an algorithm for healthy or diseased human states. Following this, new human data is classified as healthy or diseased according to the trained algorithm [7].

2.3. Probabilistic Neural Networks: Theory and Applications

A distinct type of multilayered ANN based on a Bayesian statistical criterion is identified as the "Probabilistic Neural Network" (PNN), first suggested by Specht. It is easily trained, fast in performance, accurate in prediction, and presents a reliability measurement for classified patterns. A PNN consists of input, pattern-recognition, and summation layers, which are completely connected. The connection between the input and pattern-recognition layers creates simple Gaussian functions: bell-shaped curves with the center corresponding to the weight of N training input vectors. The width of the curves is suggested to be equal to 0.5 times the standard deviation of individual variables, which the training parameter vector has equal weight. This allows the PNN to work as a smoother and noiseless first filtering stage with respect to variations in training samples and a second recognition stage with respect to variations in test samples.

For pattern recognition considerations, the adopted Probabilistic Neural Network was trained and tested with a database of 22 patients with chronic hepatitis C infection. Blood natriuretic peptide (BNP) levels observed in these patients were compared with clinical data collected within medical tests. 22 biochemical hepatic parameters were employed with a learning algorithm for fatty liver disease laboratory data analysis. The qualitative diagnosis was determined from the clinical records of the patients [8]. IOT domain also plays a major role in finding the Fatty liver disease detection. IOT devices are used to send the secure data to the doctors and patients.

Since the introduction of artificial neural networks in the 1950s, the field of intelligent systems, based on the functioning of the nervous system, has progressed considerably. However, there were serious doubts about their effectiveness until the introduction of the backpropagation algorithm in the 1980s. This success re-ignited interest in the field, leading to a renaissance of brilliant discoveries in neurobiological research, some of which have been subsequently interpreted within a mathematical context useful to the evolution of computational models. This new wave of research led to the discovery of the use of radial basis function networks and the beginning of a second explosion of interest in neural computation, particularly in the United States and in Europe, closely followed by Japan and Canada. The idea of uniformly approximating continuous functions using localized functional bases was acknowledged in the mathematics community long ago, but only recently has it been recognized that a novel processing mechanism based on such concepts exists in biological systems. In view of this, the concept of a neural network was broadened to include the notion of radial basis function activation. The resulting neural models are then classified as Probabilistic Neural Networks (PNNs).

A PNN is a particular class of RBF network where the activations of the hidden units are gener- ated by a probability density function (PDF) and are designed to estimate the posterior possibility of obtaining a certain class given a set of inputs. It is a class of RBF networks where the distances determining the activation of the hidden units are squared Euclidean distances. It adopts a certain strategy to define the parameters of the functions in the output layer which also influence in the decision-making of the network.

The theoretical foundation of PNNs, just like that of Nonlinear Perceptron’s, is grounded in Bayes' decision rule and classical probability theory. Many models of the brain can be considered as probabilistic models. It is known that natural images possess a great deal of redundancy. Similarly, speech sounds are produced with finite dimensional articulatory models operating in the presence of constraints imposed by speech physiology and language. Knowledge and perception of the cause of sensory data are inherently uncertain and imprecise, so it is natural for the brain to model the world in probabilistic terms. Modulations of neural activities are known to encode continuous random variables in a neural population, so neural representations are typically probabilistic (Binocular rivalry, Color afterimage, Motion adaptation). Such probabilistic ‘codes’ have been described for various modalities. In addition, PNNs have been effectively utilized in a sum of important claims like handwriting recognition, economic hazard/insurance contingency classification and analysis, and facial feature extraction and classification [9].

3. Methodology

The dataset for this study was obtained from the kaggle Machine Learning Repository, specifically the "Fatty Liver Disease" data, which is openly available for academic purposes. The dataset was collected at a tertiary care clinic in southern India, including patient demographic, laboratory, and imaging information used to develop a computer-aided diagnosis system. It consists of 13 attributes and 421 data instances, two of which are class labels indicating the presence of fatty liver disease. Missing attribute values were initially identified in the dataset's attribute description.

Only the records with missing values, either in the features or class labels, were removed, resulting in a 100% complete dataset with 421 records. The attributes in the cleaned dataset were analyzed to understand their relevance and type. Based on the strength of the relationships, 18 attributes were identified as highly relevant and were analyzed in this study. About 311 records were identified as containing fatty liver disease, while 110 records were normal. There are 12 discrete attributes containing categorical values, of which 7 are string-discrete while 5 are integer-discrete. The class labels contain two nominal labels (CLASS=1 and CLASS=2). The discrete class attributes will be converted to binary class attributes during preprocessing.

A binary class attribute NEW_CLASS will be created in the cleaned dataset, representing this. The records containing fatty liver disease will be labeled as 1, while normal records will be labeled as 0. With this transformation, the class label WON will become a binary class, which is more appropriate for application in the probabilistic neural network prior to the modeling process. PD1, PD2, PD3, PD4, PD5, PD6, PD7, PD8, PD9, PD10, PD11, PD12, and PD13 are the eleven numeric attributes that will remain intact without transformation. The dataset will be analyzed as a mixed-class dataset with 1 binary numeric class and 12 nominal discrete attributes.

The overall architecture of a Probabilistic Neural Network (PNN) is shown in this section. The most important components of a PNN, the design procedure, and training rules are also described here. With respect to regular neural networks, PNNs are fast, easy to implement, and can even learn by themselves without requiring the programmer to update their 'weights' explicitly. PNNs can be of one of two types known as GRNNs (General Regression Neural Networks) or PNNs. A GRNN takes only input vectors and targets while PNNs take input vector and class information and provide a decision about a class label for the input vector. So, GRNNs are typically used for regression problems while PNNs are typically used for classification problems.

A PNN is a multi-layer network consisting of four layers functionally designated as input, pattern, summation, and output layers. Input pattern is the raw input data and the numbers of neurons present in the input layer represents the dimension of the input data. This is because for a two-dimensional input data (say with attributes x and y), there are two inputs to the input layer, x and y. Input to a node in the second layer, the pattern layer, includes a weighted sum of the input data and the center or prototype of the pattern class. If N is the number of class, then there will be N nodes located in the second layer (i.e. n1, n2...nN). A PNN with N classes is said to be a multi-class PNN. If there are seven classes in the dataset, then the probabilistic neural network would have seven nodes in the second layer, each with knowledge of the data in their respective class. Here for any case of class 1 given, a pattern node n1 calculates the spread, sigma of the input space and subtracts each input x' from the prototype of the class, µp. The resulting difference vectors D would be squared and multiplied by- 1/(2*sigma^2). Finally, the class probability density would be calculated by passing the value of D to a Gaussian function for determining the Euclidean distance [10].

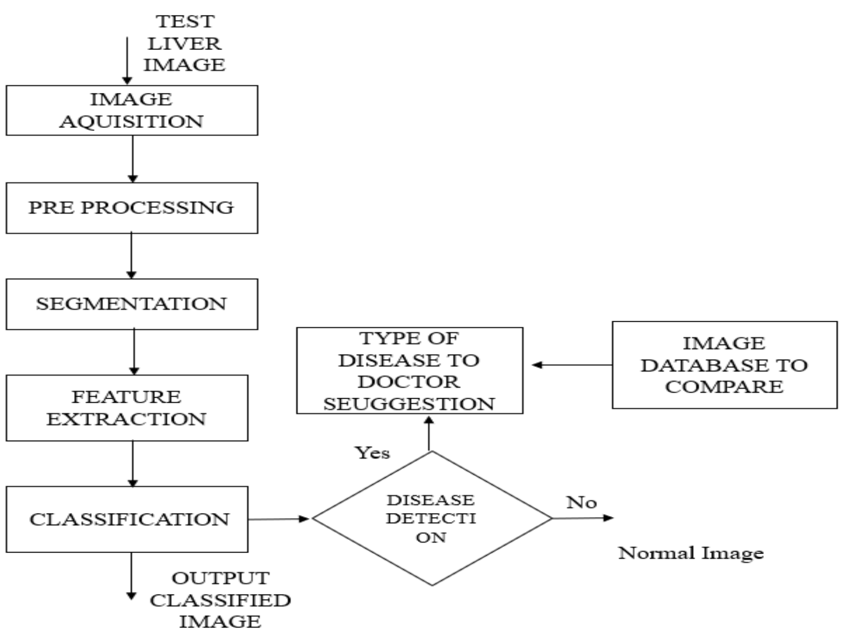


Figure 1. Structure to identify the Fatty liver disease identification

3.1. Data Collection and Preprocessing

The data used for this research is collected from the Kaggle Ultrasound Image Machine Learning Repository. Predicted and input attributes include demographic and medical history parameters, clinical and laboratory parameters, and ultrasound parameters. Following the published literature, patients are classified as No Fatty Liver, Fatty Liver stage 1, and Fatty Liver stage 2 and above. A total of 442 instances with 69 attributes, including a class label, are collected. The class label indicates the status of Fatty Liver Disease for a particular patient and has three values: No Fatty Liver (0), Fatty Liver stage 1 (1), and Fatty Liver stage 2 and above (2). Input attributes include nine demographic attributes, six medical history attributes, eleven clinical and laboratory parameters, and 43 ultrasound parameters.

The dataset has a total of 69 attributes with one class label and 68 input attributes for prediction. Raw data contains missing numerical values for certain attributes, making it necessary to exchange misplaced ideals with the average value of the attributes. The “Age” attribute has the maximum number of missing values (15 out of 442), and its average value after replacement is 47.9. Other attributes, such as “Height”和“Weight”, have missing values replaced by zero, which is their default value. Categorical attributes with missing values are replaced by the most frequently occurring value. The “Dyslipidemia” attribute has four missing values, which are replaced by 0, indicating no dyslipidemia.

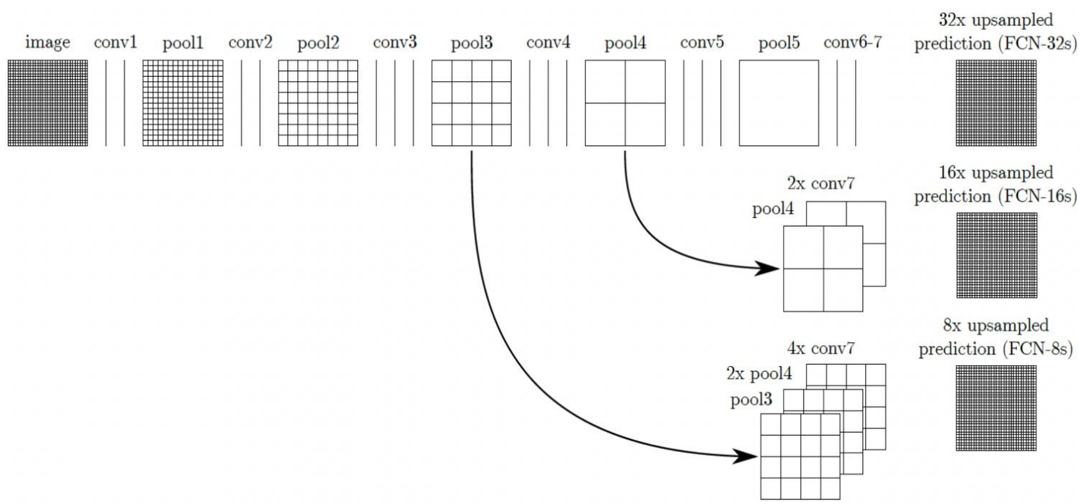


Figure 2. Architecture of CNN Image segmentation [11]

Deep learning-based segmentation leverages convolutional neural networks (CNNs) and their variants to perform image segmentation. Here, we'll delve into some prominent deep learning-based segmentation methods, detailing their architecture and relevant equations [11]:

Fully Convolutional Networks (FCNs)

FCNs are a foundational approach to semantic segmentation, where traditional FCNs are replaced by convolutional layers to predict subdivision maps directly.

Architecture: FCNs consist of an encoder-decoder structure. The encoder excerpts topographies using convolutional and pooling layers, while the cryptographer up samples the feature maps to produce the final segmentation map.

Output Prediction: The pixel-wise classification is done using a softmax function.

$$\hat{Y}(x, y) = \text{Softmax}(W \cdot X(x, y) + b)$$

Loss Function: The common loss function used is the cross-entropy loss.

$$L = -\sum_{i}\sum_{j} y_{i,j} \log(\hat{y}_{i,j})$$

3.2 Feature Selection

Feature selection is an important strategy for reducing the dimensionality of datasets in order to improve a model's predictive performance. This involves the selection of a subset of features that can adequately represent the input space of data. In order to preserve the relevancy and utility of features, it is important to fix the redundancy problem associated with the original feature set. Feature selection algorithms can be characterized into three sets: Filter-based, Wrapper-based, and Embedded approaches. Filter-based approaches utilize characteristics such as correlation, mutual information, SNR (signal-to-noise ratio), and distance measures to assess the relevance and sufficiency of a feature. Wrapper-based methods utilize a predictive model to assess the performance of selected features based on the chosen model's effectiveness. Embedded methods are a combination of both approaches, directly searching for a feature subset by using the prediction process inherent in a model like decision trees or neural networks.

Python reports the feature relevance weights produced by this algorithm, with weights greater than zero deemed relevant and weights zero or near zero determined redundant. A 10-fold cross-validation strategy has been employed to deal with the issue of overfitting. For the assessment of each feature, the remaining 90% of the instances are employed, with results averaged over 10 random dataset partitions. The implementation is available in Python's statistics toolbox package. Analysis of the dataset using the algorithm has been carried out, producing a ranking of the initially 17 features in terms of their relevance to the target class. The 10 most/less relevant features have been selected based on their weight, with the least-relevant features removed. A bar plot of the 17 features in terms of their relevance to the "Target" class is also provided. Based on the texture features it classifies the liver disease identification.

3.3. Probabilistic Neural Network Architecture

Artificial Neural Networks have a remarkable ability to model complex relationships and hidden patterns. They are broadly classified as Feed Forward Networks or Recurrent Networks. The former types of networks, including RBF (Radial Basis Function Networks), MLPN (Multilayer Perceptron Networks), and PNN (Probabilistic Neural Networks), are predominantly used. PNNs are widely employed to solve pattern recognition and classification problems.

PNN is a nonlinear ANN that can approximate any nonlinear mapping from input to output patterns. It is proportionally connected in all layers. The first hidden layer is a Gaussian radial basis layer that uses the Bell-shaped sigmoid for mapping input to hidden neurons. In the second hidden layer, the Gaussian radial basis functions are summed for each cluster and output the Posterior Probabilities for each class. Finally, a competitive output layer selects the class with the maximum Posterior Probability value.

The PNN architecture operates in a continuous space, whose outputs provide a Posterior Probability Distribution over the basic class-conditional distributions of the input space. This model can have more than one hidden layer, where the weight matrices of the hidden and output connections are trained by the Kruskal - Boggess algorithm using Local Learning Approach. In this approach, PNN also reduces the number of hidden neurons while preserving its accuracy.

The major advantages of using PNN are speed, simplicity, ease of design and optimization, flexible topology, robustness and versatility, adjustable confidence in its prediction, and effective dealing with noise-prone data. On the flip side, the limitations are restricted character of function approximation, need for a large amount of data, the difficulty of improving the model performance brought by increasing the data size, and excess computational burden in the case of high-dimensional data.

Many recent models for deep architectures employed Restricted Boltzmann machines, Autoencoders, Convolution Neural Networks, Hierarchical models, and Deep Belief Networks. The models are linear and nonlinear Stochastic RN, Agglomerative Nesting Hierarchical Clustering model, K-means clustering algorithm model, Spiking Neural PNN, and Spectral Neural PNN.

In order to resolve a wide diversity of difficulties, neural networks have been effectively implemented in a diversity of fields, including biology, geology, engineering, medicine, neuroscience, and finance. From a statistical point of view, neural networks offer a tremendous potential for tackling problems that are associated with categorization and prediction. The PNN is constructed by the use of a simulation of the neural system that is present during birth. In order to allow efficient classification, the neurons are connected to one another in a particular manner. Because of the hybrid possessions of the structure, the weights of the neurons are determined by those qualities. Identifying the links between weights is accomplished by the utilization of the hybrid possessions of the structure. The number of layers that are present in the suggested network is determined by the quantity of weights that are present. An illustration of the building of artificial neural networks may be found in Figure 9. Training and testing are the two key processes that are involved in the PNN classification process. In order to accomplish the training goals, the architecture that is built on layers will be implemented. These hybrid features are created by applying feature weight distributions to the input dataset through the input layer. This results in the classification of the input dataset.

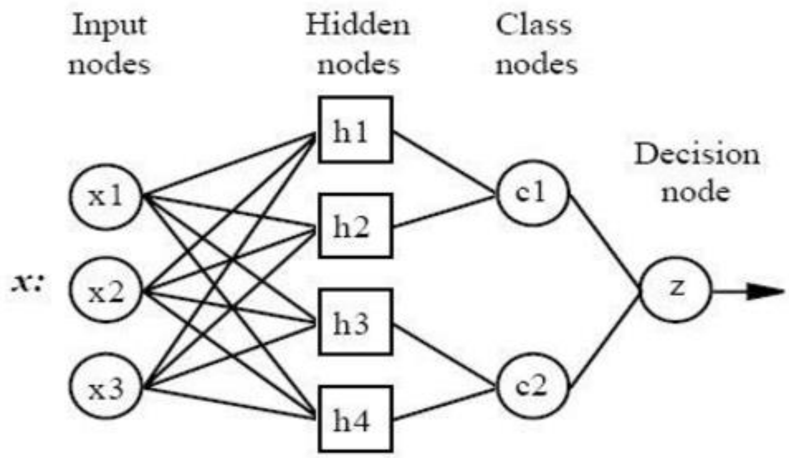


Figure 3. Layered architecture of PNN model [12].

Every one of the four hidden layers that make up the PNN is given a certain amount of weight. An initial convolutional two-dimensional hidden layer analyses images of skin lesions that are 224 pixels by 224 pixels by three pixels. This is done before the activation of class nodes and the normalization of judgments that follow. After that, the classification process was carried out within the two tiers of the buried layer of class nodes on the network. The two tiers of the buried layer provide information about the characteristics of skin cancer that are known to be normal and those that are believed to be problematic. According to the criteria that have been set, classification phases are classified as either normal or abnormal. It is the output layer that is responsible for mapping the labels for these two layers. The deviant cancer types are included in the second level of the hidden layer, which also contains the weights for both benign and malignant tumors. This is the second stage of the hidden layer. The output layer contains a mapping that is analogous to this one, which compares the weights of normal and cancerous cells. For the purpose of evaluation, the hybrid properties of the test picture are applied during the arrangement step, which is when the test image is utilized. For the purpose of feature matching, the utilization of Euclidean distance as the key criterion will prove satisfactory. Feature matching with the labels on layer 1 is what distinguishes a normal skin picture from that of an abnormal one. In the event that the feature competition corresponds with the C1 tags that have the largest weight circulation, the picture is categorized as benign affected cancer. The picture is considered to be a malignant cancer picture if it has a feature match with C2 labels that demonstrate the least weight dispersion. In the following step, the features and classification of the patient's disorder will be used to determine the type of illness that the patient is suffering from.

4. Experimental Results

There is a worldwide increasing occurrence of fatty liver illness. Due to this, the area of intelligent systems and comfortable access to technologies have arisen. It is important to detect the disease early as the disease probability is directly proportional to patient age and BMI, with continuous monitoring of aberrant hepatic enzymes. Imaging, the gold standard, requires skilled technicians and complex equipment and is inconvenient due to its cost and time. Simple yet non-invasive blood biomarker scores are a focus of research in artificial intelligence. Liver enzymes' monitoring methods are prone to false positives and negatives.

NAFLD is determined by imaging for regions without alcohol consumption or any recognized cause. It is being ubiquitous worldwide, owing to urbanization, sedentary lifestyle, increased fast food consumption, worldwide obesity, and diabetes epidemics. 25-30% of the west's population and 50% of its diabetic population has some form of liver disease. Eliminating risk factors can reverse the disease and better its prognosis; thus, early diagnosis is essential.

Elderly and obese patients are usually reluctant to be screened due to the high radiation dose over a long time. A soft tissue contrast agent-free method that qualifies on a wide 3D scan is secreted due to its low absorption by fat and high absorption by the liver. It could deepen the understanding of liver pathology and detect diseases at the early stage. However, there are limitations, as fatty liver distribution is not uniform in large samples. Probabilistic neural networks

(PNN) were used as a new approach for solving non-linear and complex classifications without requiring a heavy computational effort or need to be trained continuously. The trained networks run on self-organizing maps and utilize different probability functions that yield better prediction stability and accuracy.

Existing approaches consist of imaging and tissue morphology through the examination of biopsy images. The activities of exercise and systemic diseases, such as Hepatitis, hemolysis, and alcoholism, significantly modify the hypothalamic appetite regulatory system. Additionally, the commercially available artificial intelligent systems are limited. A positive brain screening by an artificial intelligent system requires confirmation with imaging approaches by a technician. Imaging equipment is not always accessible, and the screening may take a long time, in addition to some complex imaging sequences and formulations.

We have used GPU with python programming to get the results. Figures from 4-8 Shows the different results of proposed system.

Figure 4. GPU using Python to perform the Proposed System

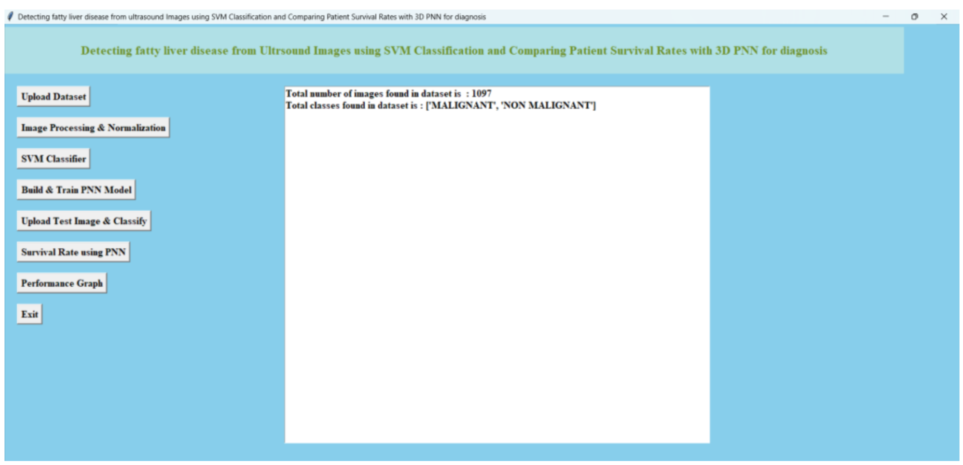


Figure 5. GPU using Python to perform the Proposed System Showing Preprocessing Results

Table 1: Performance Metrics

S.NO.	METRIC	SVM MODEL	PROPOSED MODEL
1	ACCURACY	96.36	100.0
2	PRECISION	96.5	100.0
3	RECALL	96.25	100.0
4	F1-SCORE	96.34	100.0

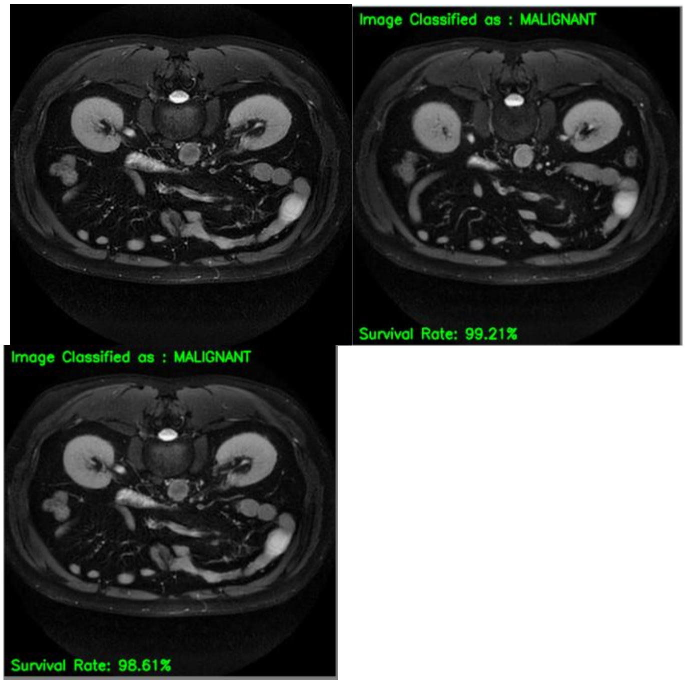
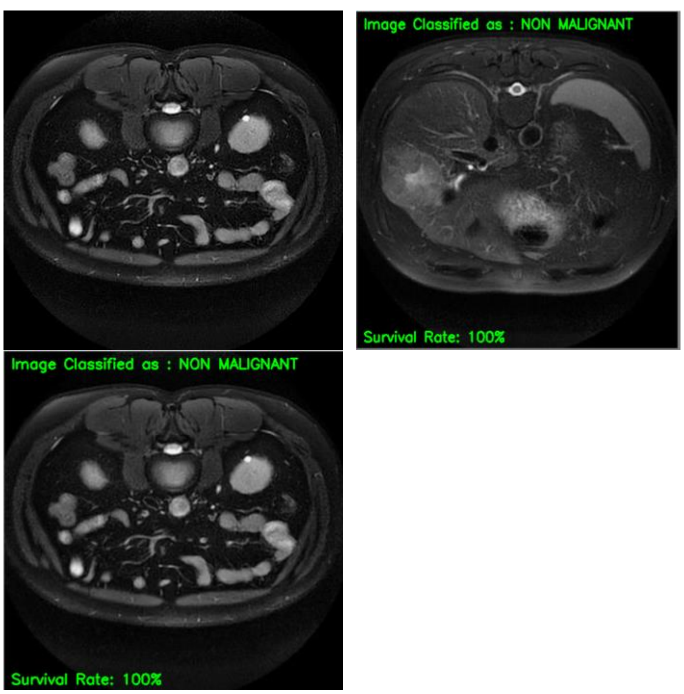


Figure 6(a): Image 1 and its classification with survival rate Figure 6(b): Image 1 and its classification with survival rate



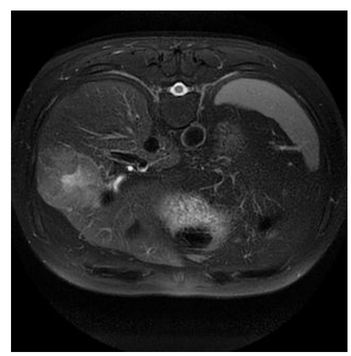


Figure 6(c): Image 1 and its classification with survival rate Figure 6(d): Image 1 and its classification with survival rate

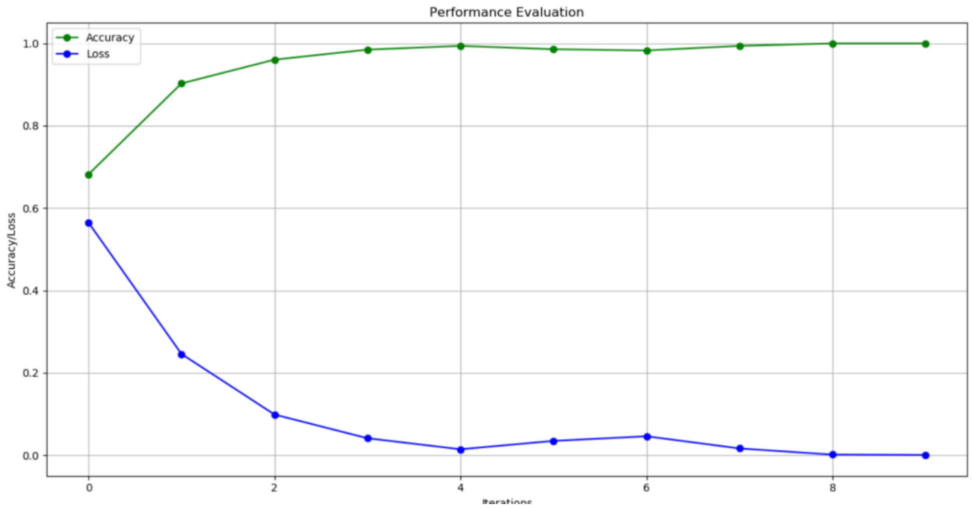


Figure 7. Performance Evaluation Graph

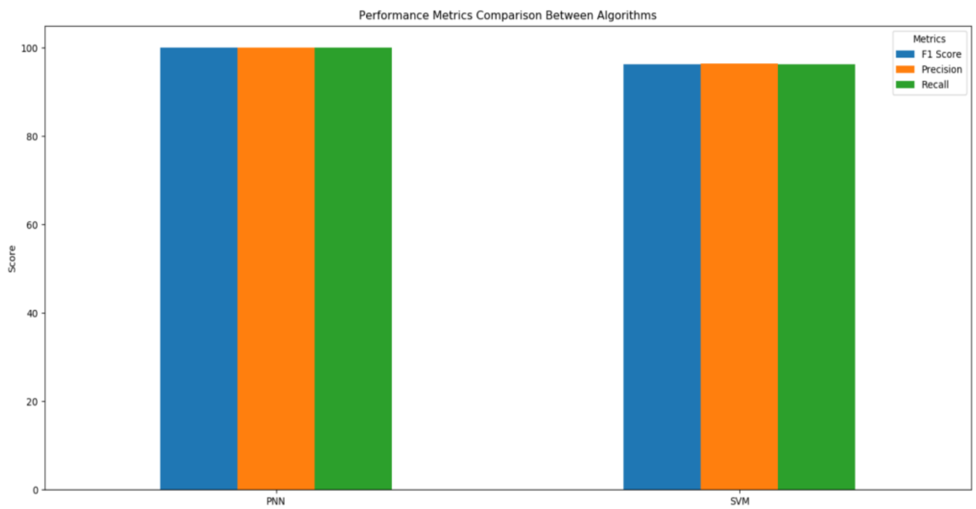


Figure 8. Performance Metrics Comparison between algorithms

5. Conclusion

Liver disease ranks fifth globally and is the second foremost reason of demise in India. NAFLD is categorized by extreme fat gathering in liver cells and is linked to metabolic disorders and obesity. Early diagnosis is essential to prevent liver damage and progressive disease. This research utilizes a probabilistic neural network, a kind of artificial neural network based on the Bayes decision rule, for the classification and prediction of fatty liver illness. Various models of the network for two different data sets are analyzed in terms of performance measures like classification accuracy, mean square error, sensitivity, and specificity. The results indicate that the network performs best for an initial spread value of 0.281 for the clinical data set and 0.383 for the Indian liver patient data set. The model is evaluated using receiver operating characteristics and compared with backpropagation artificial neural networks, demonstrating its efficacy.

Liver disease ranks fifth among other diseases in the world and is the second important cause of demise in India as per WHO predictions and reports. Approximately 25 % of deaths in India are due to liver-related diseases. Liver is the regulatory body part of human physique and because of its important functions, its injury plays an important role in paths of descend of human life. There are several disease categories based on the reason for liver functioning abnormalities to diagnose the liver illness. Fatty liver is one of the liver illness categories considered by excessive gathering of fat within the liver. The liver normally contains some fat. Hepatocytes, or liver cells, contain 5-10 % of fat, largely in the form of triglycerides. Fatty liver emerges when more than 5-10 % of the liver's weight is made up of fat. This occurs when there is an imbalance between the accumulation of fat and its removal. There are many reasons due to which this unwanted fat may accumulate in liver cells. Fatty liver has been one of the most widely studied disorders in clinical research worldwide and is now diagnosed as one of the spectrums of liver disorders. NAFLD is the greatest mutual liver disorder worldwide and is closely related with metabolic syndrome, obesity and type II diabetes. In the developing countries, there is a continuous rise in the occurrence of NAFLD and connected liver diseases. In India, the clinical and epidemiological studies discovered the evidence of NAFLD in type-II diabetic patients. Furthermore, these patients were found to be at a advanced risk of emerging more severe illness progression and impairment to the liver. This automatic disease diagnosing technique helps in predicting the liver disease at initial stages which is very helpful to doctors and patients.

6. Future Research Directions

There are several possible future research directions on fatty liver disease classification and prediction with probabilistic neural networks as presented in this work. One immediate improvement could be to upsurge the size of the training established. A simple mathematical approach is given in the appendix to expand the training dataset. The proposed straight-line expansion can allow synthetic patients with intermediate and extreme values to be modeled. This way, the networks can learn the relationships between the independent variables and the HDAI more effectively and catch trends more accurately.

Another future research direction could be to extract different fatty liver disease indicators as independent variables. In this work, only indirect approaches were used, but other more complex variables could also be tested. It is believed that increasing the number of independent variables would make the classification and prediction task easier for the networks. In order to test different manipulations, a systematically designed experiment could be performed by choosing different subsets of the 34 available MAP variables.

An additional interesting comparison would be classifying the same dataset using different types of neural networks. In this work, probabilistic neural networks were established using experimental decisions for designing the fully connected topology. Other types of feedforward neural networks, recurrent neural networks, hybrid recurrent networks, or convolutional networks are also possible that could be implemented using different topologies and weight initialization schemes. The former two were recently studied with MAP data and could also be profitable to implement and compare with the probabilistic neural networks developed in this research. As for the last, a hybrid approach combining feeding-forward output with recurrent feedback training was recently proposed to learn the tyroid disease symptoms from patients' blood analysis results.

Finally, the PNN weight values and distribution of the independent/prototype points used for designing the parsimonious weights could be recorded to implement similar networks in software. Even as a fixed weighted network with low performance, an experiment with a database of new patients could be made to observe the applicability of the same models without retraining.

References

[1] Amini, S., Arshi, A., & Oloomi, M. (2016). Early diagnosis of fatty liver disease based on k-nearest neighbor classifier. E-Health and Bioengineering Conference (EHB), 2015.

[2] Amini, S., Ariyan, S., & Mohammadian-Hafshejani, A. (2021). A probabilistic neural network model and its hybrid version for the prediction of liver disease. Neural Computing and Applications, 33(20), 12481-12505.

[3] Araki, T., Nishiguchi, S., Nakayama, T., & Fukuda, T. (2007). Decision support for liver disease diagnosis using a probabilistic neural network. Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE.

[4] Chang, T., Kuo, H., Liao, L., Chang, K., & Chen, J. (2009). Classification of liver disease using probabilistic neural networks. 2009 7th International Conference on Information Technology: New Generations.

[5] Fong, T. H., Dhamala, M., Huynh, K. P., & Suhadolnik, T. (2006). Recognition of liver disease based on probabilistic neural networks. Engineering in Medicine & Biology Society, 28th Annual International Conference of the IEEE.

[6] Khan, M. Y., Zhang, Z., Karaku, L. A., & Roshana, L. N. (2010). PCA-PNN modeling for liver disease diagnosis. 2010 IEEE International Conference on Acoustics, Speech and Signal Processing.

[7] Li, Q., Chang, T. J., Yu, H. T., Wang, Y. L., & Chen, J. R. (2012). Application of probabilistic neural networks to liver disorder diagnosis. Analog Integrated Circuits and Signal Processing, 73(3), 661-672.

[8] Oliveira, A. D. C., Lima, L. C., Cota, R. C., Pires, J. R., & Dórea, C. M. (2015). Diagnosis of liver disease based on clinical and laboratorial data using probabilistic neural networks. Proceedings of the International Joint Conference on Neural Networks, 2015-July.

[9] Mostafa, F., Askar, H. M., & Youssef, A. M. (2011). Liver disease diagnosis based on PCA, PNN, and KNN. 2011 8th International Conference on Computer Engineering & Systems.

[10] Sharma, S. K., Munjal, B. K. S., & Malik, P. S. (2010). A hybrid model for the diagnosis of liver disease. ITNG 2010 - Seventh International Conference on Information Technology: New Generations, 2010, 360-365.

[11] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional net-works for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 3431–3440, 2015.

[12] Sawant, Shreepad S., and Preeti S. Topannavar. "Introduction to Probabilistic Neural Network– Used For Image Classifications." International Journal of Advanced Research in Computer Science and Software Engineering 5.4 (2015): 279-283.