



# **Integration of Bidirectional Long Short-Term Memory (BiLSTM) and Two-Dimensional Convolutional Neural Networks (CONV2D) for Improved Hepatitis A and B Prediction**

*Oyinlola Dupe Sharon*

Department of Information Technology, Faculty of Science

National Open University of Nigeria

E-mail: [engineersharon79@gmail.com](mailto:engineersharon79@gmail.com), Mobile: +2348063833141

DOI: <https://doi.org/10.55248/gengpi.5.0124.0122>

## **ABSTRACT**

Hepatitis A and B pose significant global health challenges, necessitating accurate predictive models for timely identification and intervention. This study proposes an innovative approach by integrating Bidirectional Long Short Term Memory (BiLSTM) and Two-Dimensional Convolutional Neural Networks (Conv2D) to enhance the precision of predictive models for hepatitis A and B. The research aims to bridge existing gaps in the literature by thoroughly investigating the synergies between BiLSTM and Conv2D, with a focus on optimizing their combined impact on predictive performance. The research objectives encompass optimizing the architectural integration of BiLSTM and Conv2D, exploring diverse hyperparameter tuning strategies, and assessing the model's efficacy in capturing spatial and temporal dependencies within heterogeneous datasets. Additionally, emphasis is placed on enhancing the interpretability of model outputs, ensuring its practical applicability in clinical settings. Conducting an extensive literature review reveals a scarcity of studies addressing the joint integration of BiLSTM and Conv2D for hepatitis A and B prediction. The novelty of this research lies in addressing this gap and contributing insights into the optimal configuration of these architectures for heightened predictive accuracy. The methodology section details the proposed approach, including a comprehensive discussion of its components, detailed explanations of each aspect, and a thorough examination of any crucial algorithms or frameworks utilized. The research aligns with the imperative need for precise hepatitis diagnosis, considering the high mortality rates associated with hepatitis-related complications. Employing a bidirectional long short-term memory and a two-dimensional convolutional neural network model on the UCI hepatitis dataset, the study demonstrates promising results. Performance measures, including accuracy (98.3%), precision (0.65), recall (0.95), and F1 score (0.69), underscore the effectiveness of the proposed model. Noteworthy is the innovative learning strategy applied to the dataset before prediction, contributing to enhanced performance. While the model exhibits notable success, challenges persist in differentiating overlapping hepatitis cells. Future research avenues include investigating issues related to cell overlap and uneven color distribution in pathological data obtained through various staining methods. Techniques such as area thickness analysis and fractal diameter measurement will be vital for distinguishing overlapping cells from single ones. This research contributes to advancing hepatitis disease prediction, providing a robust model with potential implications for clinical applications. Future endeavors can address challenges related to cell overlap and color distribution, ensuring the continued refinement and applicability of the proposed model in diverse medical settings.

## **Introduction**

Hepatitis B virus HBV is a partially double helix DNA virus with a unique genomic structure and replication mechanism. Due to low reverse transcriptase fidelity and rapid replication, a progeny virus genome in an infected cell may not be the same as the parental genome (Andino, & Domingo, 2015). This results in a viral ecosystem consisting of swarms of mutants, or "near-species", which are populations of genetically diverse but closely related viral variants (Caligiuri et al. events, 2016). Viral hepatitis is defined as "inflammation of the liver". The liver can become inflamed for many reasons, including drinking too much alcohol, sustaining physical trauma, developing an inflammatory response, or reacting to bacteria or viruses. A, B, C, D and E are the five most common hepatitis viruses. Hepatitis virus can cause fibrosis, cirrhosis, liver failure and even liver cancer. Your liver's ability to function is impaired, making it harder for your body to filter out toxins (Spearman C.W, hep.org/blog/). Although hepatitis A and B both affect the liver, the two viruses are very different. Hepatitis B is a blood-borne disease, transmitted mainly through direct blood-to-blood contact with an infected person. On the other hand, hepatitis A can be transmitted by the fecal-oral route or by ingestion of contaminated food or drink. HBV infection is still a great public health concern around the world. HBV infected about 257 million people and more than 350 million people lived with CHB (Tian et al, 2019). In general, know that Seroclarance HBsAg is a significant precious marker when treating ECHB (Lee et al, 2014). By persistent patients with HBVinfed, HBSAG's annual spontaneous serum ratio ranges from 0.45% to 2.38%, prove that Seroclarance HBsAg is a rare appearance (Wu et al, 2015 ). The previous search told HBsAg seroclarance, whether spontaneous or caused by drugs, better prognosis, improved liver histology, lower liver cell carcinoma (HCC) and longer longevity (Yarasuri et al, 2019). Therefore, HBsAg serum clearance is an important target for better antiviral outcomes. Previous studies have found

evidence of important viral variables and host characteristics of HBsAg serum clearance. Low blood HBsAg levels have been identified by researchers as a major contributor to HBsAg serum clearance, either alone or in combination with low serum HBV DNA loading (Chong & Hao, 2019). In terms of host attributes, one of the main determinants of HBsAg antibody seroconversion is age, trailed by gender, fatty liver, cirrhosis at baseline or during follow-up, and baseline alanine aminotransferase (ALT) levels (Chu & Liew, 2010). In recent years, machine learning algorithms in healthcare have received a lot of attention. It is effectively used as a powerful classification method for extracting useful information from high-dimensional, correlated, non-linear, and imbalanced clinical datasets, enabling accurate diagnostic and predictive decisions (Austin et al, 2013). However, the existing HBsAg serum clearance prediction model with maximum power has not been identified. The utilization of deep learning in imaging has erupted in notoriety as of late (LeCun, Bengio, Hinton, 2015). Convolutional neural networks (CNNs) have demonstrated to be a solid instrument for general object identification, studying high-level image attributes, and improving image classification performance, (Li et al 2012). These accomplishments, utilizing deep learning have provoked the interest of specialists in clinical image examination throughout the years (Wolberg, Street, Mangasarian, 2019). Notwithstanding, the training of deep CNNs without any preparation is a troublesome undertaking, especially in the medical imaging field, since it requires colossal medical imaging datasets with suitable human-outlined explanations, which has been proven tedious and laborious to acquire. Nevertheless, an innovation known as "transfer learning" has emerged and is now being utilized in medical imaging analysis (Spanhol et al 2016). Pretrained CNNs modeled with either a nonmedical picture dataset or a medical picture dataset from an alternate methodology are utilized to perform clinical decision-making tasks with a restricted medical image dataset in these occasions. The output obtained from the network's layers can be utilized as elements for an assortment of clinical exercises. Pretrained profound CNNs based on mammograms were utilized in crafted by Samala et al. for breast mass lesion discovery on computerized breast tomosynthesis pictures. Huynh et al. utilized deep CNNs with transfer learning to group breast cancers on computerized mammograms, with discoveries indicating comparable performance to current computer-aided diagnosis (CADx) approaches.

---

## Statement of the Problem

Hepatitis A and B are viral infections affecting the liver, posing significant public health challenges worldwide. Early and accurate prediction of these infections is crucial for effective preventive measures and timely medical interventions. Traditional prediction models often face limitations in capturing complex patterns and dependencies within heterogeneous datasets, hindering their ability to provide accurate and timely predictions. The integration of advanced deep learning techniques, specifically Bidirectional Long Short Term Memory (BiLSTM) and Two-Dimensional Convolutional Neural Networks (Conv2D), offers a promising avenue for enhancing the predictive accuracy of hepatitis A and B models. However, there remains a gap in the research regarding the optimal integration of these two architectures and their combined impact on predictive performance.

This study aims to address the following key issues:

**Limited Integration Studies:** There is a lack of comprehensive studies exploring the synergies between BiLSTM and Conv2D in the context of hepatitis A and B prediction. Understanding how these architectures complement each other could lead to a more robust and accurate predictive model.

**Feature Representation Challenges:** Traditional models often struggle to effectively represent complex spatial and temporal patterns present in hepatitis A and B datasets. Exploring how the combination of BiLSTM and Conv2D can capture both sequential and spatial dependencies is crucial for improving feature representation and, consequently, prediction accuracy.

**Optimal Hyperparameter Tuning:** The integration of BiLSTM and Conv2D introduces a multitude of hyperparameters that need to be optimized for optimal performance. Finding the optimal set of hyperparameters for the joint architecture remains an open challenge, and this study aims to contribute insights into achieving this optimization.

**Interpretable Model Outputs:** The black-box nature of deep learning models often hinders their interpretability. This research seeks to enhance the interpretability of the combined BiLSTM and Conv2D model outputs, providing valuable insights into the factors influencing hepatitis A and B predictions.

By addressing these challenges, this study aims to advance the understanding of the application of deep learning techniques for hepatitis A and B prediction and contribute towards the development of more accurate and interpretable predictive models in the domain of infectious disease epidemiology.

---

## Research Objective

The goal of this research is to build Bidirectional Long Short Term Memory (BiLSTM) and Two-Dimensional Convolutional Neural Networks for Hepatitis A and B Prediction (Conv2D). The objectives are as follows: To develop a method for detecting and predicting hepatitis A and B with the aid of a Bidirectional Long Short Term Memory (BiLSTM) in conjunction with a Two-Dimensional Convolutional Neural Network (Conv2D).

---

## Research question

1. How does the integration of Bidirectional Long Short Term Memory (BiLSTM) and Two-Dimensional Convolutional Neural Networks (Conv2D) enhance the predictive accuracy of models for hepatitis A and B, and what are the specific contributions of each architecture in capturing temporal and spatial dependencies within heterogeneous datasets?

2. What is the impact of hyper parameter tuning on the performance of the joint BiLSTM and Conv2D model for hepatitis A and B prediction, and how can an optimal set of hyper parameters be identified to maximize predictive accuracy while ensuring computational efficiency?

---

### Significant of research

This research will be limited to the examination and analysis of Hepatitis A and B risk assessment and prediction activities. Other elements of the domain aren't taken into account. Hepatitis A and B are becoming more common over the world. In most rich and emerging countries, the rate of occurrence has been steadily increasing. Hepatitis A and B-related mortality are also on the rise, in addition to the rising incidence. Around 7.6 million individuals died worldwide from Hepatitis A and B in 2008, with about 70% of these deaths happening in underdeveloped nations (Tian et al, 2019). Every year, more than 250,000 new cases of Hepatitis A and B are projected to be detected in Nigeria, with up to 10,000 Nigerians dying from Hepatitis A and B-related causes (FBS). These assessments or estimates may not mirror the genuine picture since they are regularly hinged on hospital-generated data, which overlooks the many cases that are not present in medical clinics, those oversaw by traditional medication specialists, and the various instances of misdiagnosis in our various neighborhood hospitals. As of 2015, the United States had roughly 230,000 new instances of hepatitis A and B each year. In the United States, this sort of Hepatitis A and B causes around 40,000 deaths every year. As a result, providing a mechanism for assessing and forecasting this risk would be extremely beneficial to the general public, particularly Nigeria as a country. Machine learning algorithms will undoubtedly provide us with access to the available processing power of recent advances in computing technology, as well as the artificial intelligence with which to appropriately utilize this computing power for solving beneficial problems such as Hepatitis A and B.

---

### Literature Review

Hepatitis is an ailment wherein the liver tissue becomes swollen. Hepatitis causes yellow staining of the skin and whites of the eyes (jaundice), as well as lack of appetite, regurgitation, exhaustion, stomach and abdominal agony, and in certain people or creatures, diarrhea (WHO, 2016). Hepatitis is declared acute assuming it clears up in a half year or chronic or persistent assuming it endures longer than that (Vos et al, 2016). Intense hepatitis can disappear all alone, progress to persistent hepatitis, or lead to acute liver failure (once in a while). Cirrhosis, liver failure, and liver cancer are generally potential results of persistent or chronic hepatitis (Khalimi and Burman, 2019). Hepatitis A, B, C, D, and E are the most predominant viruses that cause hepatitis. Weighty liquor ingestion, a few medications, poisons, different contaminations, immune system diseases, and non-alcoholic steatohepatitis, (NASH) are among different reasons. Hepatitis A and E are essentially sent by corrupted food and water. Hepatitis B is essentially transmitted through sexual contact, yet it can likewise be given from mother to youngster during pregnancy or labor, and it can likewise be passed on through tainted blood (Manns et al, 2019). Hepatitis C is commonly spread by tainted blood, which can happen when intravenous medication clients share needles. Just people who have proactively been contaminated with hepatitis B can get hepatitis D. Inoculation against hepatitis A, B, and D is conceivable. Persistent viral hepatitis can be treated with prescribed drugs. Aside from those with infections that limit their average life span, antiviral prescriptions are exhorted for all individuals with persistent hepatitis C. (Thurz et al, 2018). Despite the fact that there is no particular treatment for NASH, it is suggested that patients participate in actual physical work, consume a nutritious eating regimen, and get more fit. Prescription to suppress the immune system might be utilized to treat immune system hepatitis. In both intense and constant liver failure, a liver transplant might be plausible (Lawrence, 2015). Hepatitis A tainted around 114 million individuals worldwide in 2015, persistent hepatitis B contaminated roughly 343 million individuals, and persistent hepatitis C infected roughly 142 million individuals. NASH contaminates approximately 11 million individuals in the United States, while alcoholic hepatitis affects around 5 million. Hepatitis causes in excess of 1,000,000 fatalities every year, the vast majority of which are brought about by liver scarring or malignancy.

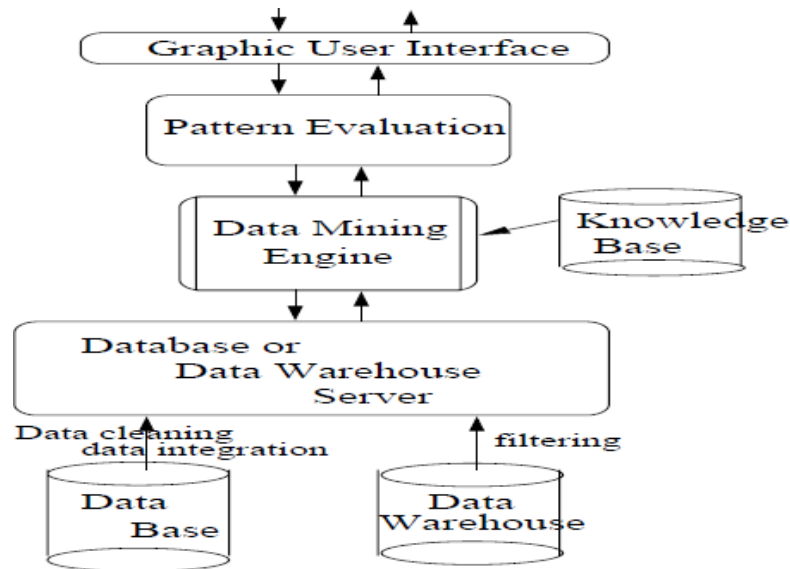
---

### Data Mining

As per Manjunath et al., (2010) and Kimball R. (2000), the utilization of choice emotionally supportive mechanisms is a method that effectively utilizes information to more readily handle, operate, and improve the assessment of educational institutions. Dependent on the attribute and availability of the fundamental data, an answer like that could handle various issues by refining data from any mix of student records management systems. Mining information from an information store can be an immediately accessible and productive system for decision makers. Manjunath et al. (2010) found that an information store is a problem engaged, unified, stable, and time-variant assortment of information used to help the executives decisions. In light of this perspective, Jiawei et al., (2000) depicted the engineering of an ordinary information mining system as displayed in Figure 2.1.

- 1 **Database, data stores, or other information warehouse.** This is a database, one or a variety of data sets, information stores, spreadsheets, or different kinds of data warehouses. Techniques, for example, Data cleansing and information integration can be applied to the data.
- 2 **Database or data warehouse server.** A user's data mining requirements determine whether the database server or the data warehouse server is responsible for retrieving the necessary data.
- 3 **Knowledge base.** Domain knowledge can be used to guide research or to measure whether patterns discovered are useful. Concept hierarchies can be used, for example, to organize attributes at different levels of abstraction. In addition, user confidence, which is used for judging the attractiveness of a model based on its unexpectedness, can also be calculated. Furthermore, domain knowledge can also refer to additional boundaries or thresholds of interest, as well as metadata (e.g., description of data from multiple heterogeneous sources).

- 4 **Data mining engine.** The characterization, association analysis, classification, analysis, and deviation analysis module is an important component of a data mining system.
- 5 **Pattern evaluation module.** Often, this component uses interestingness measures and works with the data mining modules to extract interesting data. It has access to the knowledge base storing the interestingness thresholds. Depending on the architecture of the data extraction method, the sample evaluation module can additionally be integrated with the extraction module. To limit mining to only interestingness patterns, we recommend pushing the evaluation of pattern interestingness as deep as possible into the data mining process.
- 6 **Graphical user interface.** In this module users are able to specify queries or data mining tasks in order to interact with the data mining system, providing information to help them focus on their research and operations activities. This module uses intermediate data mining to determine what's worth investigating. With this component, users can also browse databases and database programs, evaluate mined patterns, and view patterns in a variety of ways.



**Figure 1: Architecture of a typical data mining system. (Jiawei et. al., 2000)**

Among the procedures utilized in information, mining are database technology, statistics, machine learning, elite performance computing, pattern recognition, neural networks, data representation, data recovery, image and signal handling, and information analysis Geographic data.

Figure 2.2 illustrates the general data mining process as described by Zarki (2003). It includes the steps below, some of which are optional depending on the situation in question:

1. **Grasp the application domain:** To appreciate user-requested data mining results, you must first understand scope. To increase your chances of success, you need to content and use all the information you have before.
2. **Collect and construct the target dataset:** Data mining relies on the availability of relevant data that reflects the variety, order, and underlying structure of the problem. Therefore, it is essential to collect a data set that includes all conceivable scenarios relevant to the topic under study (Jiawei et al., 2000).
3. **Clean up and change the objective dataset:** Raw information has numerous blunders and irregularities, like noise, exceptions or outliers, and missing qualities. Copying information records to build a non-repetitive dataset is a significant piece of this cycle (Troyanskaya et al., 2001).
4. **Select elements, limit dimensions:** Even after information has been cleaned of copies, irregularities, missing qualities, and so forth, there can in any case be noise inconsequential to the subject being studied. These noisy elements can prompt disarray in later phases of information investigation, as well as the creation of irrelative rules and connections and expanded computational expenses. Therefore, it is very reasonable to perform dimensional reduction or feature selection stage to distinguish useful attributes from irrelevant attributes (Jiawei et al., 2000). Fisher's criteria, Wilcoxon's rank sum test, principal components analysis, entropy analysis, and other statistical or heuristic techniques are commonly used to perform this stage.
5. **Data mining algorithms can be used to examine data,** such as association rule discovery, sequence mining, classification tree induction, clustering, etc.

- 6 Patterns for interpretation, evaluation, and visualization: Check output to interpret and evaluate extracted patterns, rules, and patterns. New perspectives on the situation can only be analyzed through this process of interpretation and evaluation.

## Machine Learning

The scientific investigation of statistical models and algorithms that are used by computer systems to achieve undertakings coherently without utilizing express guidelines, rather depending on patterns and deduction, is Machine Learning (ML). It is considered a subset of artificial intelligence. Machine learning algorithms manufacture a mathematical model of test information, known as "training data," to deliver estimates or decisions without being explicitly modified to do thusly (Koza, 1996). Machine Learning techniques are utilized in applications, for example, email separating, network intrusion discovery, and computer vision, where fostering an algorithm with explicit guidelines for task completion is not feasible. Machine Learning is nearly connected with computational statistics, which is focused on forecasting or foretelling with the aid of computers. The study of mathematical optimization helps the discipline of machine learning by giving devices, hypothesis, and application domains. Information mining is a subfield of machine learning that spotlights on unaided learning for exploratory information analysis (Friedman, 1998). When used to tackle business issues, machine learning is additionally alluded to as prescient analytics.

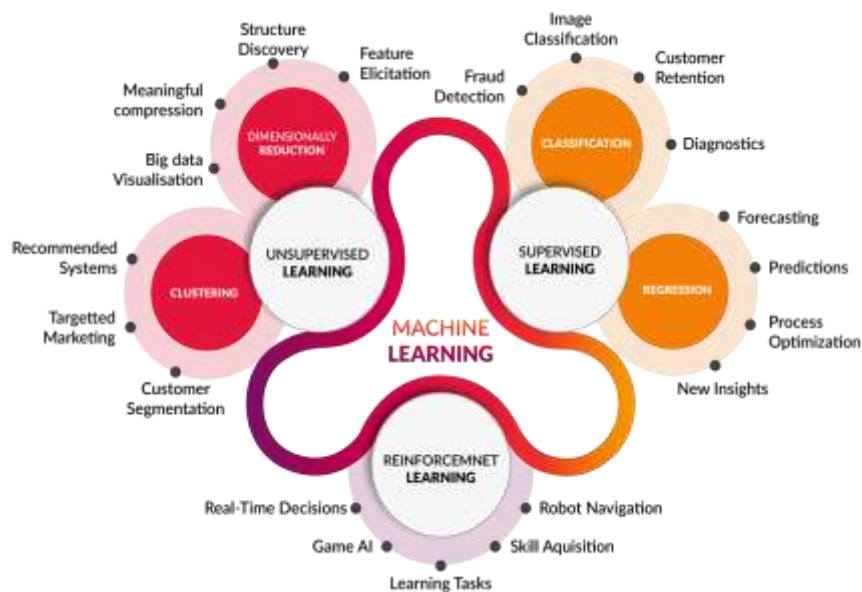


Figure 2. Machine learning breakdown (towarddatascience.com)

Machine Learning is a branch of computer science that uses a statistical technique to enable computers to "learn" from data. (For example, consistently improve proficiency on a certain activity) without being explicitly coded (Sarah et al 2017). Arthur Samuel coined the phrase "machine learning" in 1959. Samuel (1988). Machine learning evolved from the study of pattern recognition and computational learning theory in artificial intelligence, and it entails the innovative work of calculations that can gain from and anticipate information (Ron, 1998) - such calculations avoid adhering to rigorously static program guidelines by settling on information driven expectations or choices (Dickson, 2017) by developing a Machine learning is used in a variety of computing tasks where planning and programming unambiguous algorithms with excellent execution is difficult or impossible; examples include email sifting, detection of network intruders or malicious insiders attempting an information breach, optical character recognition (OCR), learning to rank, and computer vision. (He and Garcia 2019). Computational statistics, which focuses on making predictions utilizing computers, are very nearly related (frequently duplicated) to machine learning. It is very intimately related to mathematical optimization and gives equipments, hypothesis, and application domains to the field. Machine Learning and data mining are regularly befuddled (Aswathy, Jagannath, 2017), however the later subfield centers on exploratory data examination and analysis and is called unsupervised machine learning. You can likewise utilize unsupervised machine learning to learn and construct basic behavioral profiles for various entities. It very well may be utilized to distinguish related oddities (Tillmann, 2015).

Machine learning is a method for creating complex models and algorithms that are appropriate for forecast in the field of data analysis. For business use, this is referred to as predictive analytics. By studying past relationships and data trends, these analytical models permit specialists, information researchers, data scientists, engineers, and experts to "produce reliable and reproducible decisions and results" and unravel "hidden insights" (Bostrom, 2011). Clustering is a non-hierarchical approach for moving data points across clusters until similar item clusters emerge or a desired set is obtained. Assumptions regarding the data set are made by clustering algorithms. If that assumption is correct, the cluster will be good. However, satisfying all assumptions is a simple task. It may be helpful to use a combination of different clustering methods as well as altering input parameters. Association rule mining is a technique for identifying common patterns among various variables in a dataset. When compared to other classification methods, associative classification usually provides a better classification. Developing a model for assessing and accurately predicting Hepatitis B in individuals who are at risk is the primary goal of this study. The bidirectional last short term (BiLSTM) and two-dimensional convolutional neural network will be used in this study (Conv2D). Medical diagnosis is still considered an art, despite all of the standardization attempts, because it involves knowledge in dealing with

ambiguity, which is not present in today's computing gear. Artificial intelligence is a relatively recent technique in computer science, despite the fact that the notion is not new. Seroclearance of the hepatitis B surface antigen (HBsAg) after treatment is linked to a better prediction in patients with chronic hepatitis B (CHB). We still have a lot of work to do to figure out how to forecast HBsAg seroclearance reliably and efficiently using available clinical data. Previous studies that developed prediction models, on the other hand, relied mostly on long-term tracking of a few key variables and standard statistical approaches, which could lead to estimates that are skewed owing to the potential for collinearity in high-dimensional medical information. While Hepatitis A is not as deadly as Hepatitis B, it is similarly significant and, due to its route of transmission, is quite easy to get. In this paper, we present an active learning-based technique (AL) for training deep neural networks for hepatitis A and B classification.

---

## History of Machine Learning

In 1959, Arthur Samuel invented the term "machine learning". In a more formal explanation of the algorithms explored in the machine learning field, Tom M. Mitchell stated:

*"A computer program is said to learn from experience  $E$  with respect to a particular class of tasks  $T$  and a particular performance measure  $P$  if the program improves its performance at those tasks  $P$  as measured by  $E$ ." Mitchell, T. (1997).*

As opposed to portraying the subject in analytical terms, this meaning of machine learning tasks gives a very basic level functional depiction. This is in accordance with Alan Turing's methodology in his paper "Computing Machinery and Intelligence," which replaces the inquiry "Can machines think?" with "Can machines do what we (as thinking creatures) can accomplish?" (Harnad 2008). The various properties that could be held by a reasoning machine, as well as the different ramifications in creating one, are introduced in Turing's proposition. A Support vector machine is a kind of directed learning model that parts data into areas portrayed by a linear boundary. Here, the linear boundary isolates the black and white contrasting circle. Machine learning occupations are isolated into general classes. In controlled or supervised learning, the algorithm develops a numerical model from a lot of data that consolidates the two information sources and expected outcomes. For example, suppose the goal is to conclude whether an image contains a specific item, then, the training information for the controlled learning system will include pictures with and without the item (the input) and each image will have a name (the output) showing whether it contained the item. In various cases, the input may be confined to unequivocal reactions or just somewhat available. Semi-controlled learning procedure makes numerical models from inadequate training information, where some example inputs are unlabeled. Supervised learning techniques incorporate classification and regression algorithms. Whenever the results are restricted to a little arrangement of values, classification techniques are applied. An incoming email will serve as the input to an email sifting classification algorithm and the folder's name to classify the mail will serve as the output. The output of a spam email recognition algorithm would be a declaration of "spam" or "not spam," depicted by the Boolean qualities true and false. Regression algorithms obtain their name from the way that their results are continuous, meaning they can be any value within a range. Temperature, length, and item cost are generally instances of continuous qualities.

The system makes a mathematical model from an informational set that incorporates just data sources or inputs and no ideal or desired output marks for unaided learning. Unaided learning procedures are utilized to investigate the structure of information, for example, bunching or grouping of data points. Unaided learning, similar to feature learning, can find patterns in information and can sort the inputs into different categories. The process of scaling down the number of "features", or inputs, in a set of data is referred to as dimensionality reduction. The choice inputs for which training labels are acquired are optimized by active learning algorithms. This enhancement is achieved by obtaining access to desired outputs (training labels) for a definite amount of inputs in view budget. When used collaboratively, they can be given to a human client for labeling. In autonomous vehicles or while sorting out some way to play a game against a human adversary, Reinforcement learning algorithms are used, and they receive feedback as positive or negative support in an evolving environment (Bishop, 2006). Topic Modeling is a kind of Machine learning methodology wherein a computer program is given a cluster of natural language texts and asked to find other compositions or documents that pertains to related topics. Machine learning strategies may be used in density estimation problems, to acquire the imperceptible probability density function. Meta learning algorithms gain proficiency with their own inductive bias taking into account previous happenings. In developmental robotics, robot learning algorithms produce their own sequence of learning experience, for the most part insinuated as an instructive arrangement or educational program, to learn new limits over the long haul through autonomous investigation and social communication with people. Active learning, maturation, motor synergies, and imitation, are the means through which these robots are guided.

---

## Artificial Neural Network

An artificial neural network is a combination of interrelated nodes that works almost the same way as the brain's immense collection of neurons. Each circular node portrays an artificial neuron, and the link between the input of another artificial neuron and the output of one artificial neuron is portrayed by each arrow. Artificial neural Network (ANNs), otherwise called connectionist systems, are computational systems in view of natural neural networks detected in animal brains (Scott, 2010). The neural network is a system that permits different machine learning algorithms to team up and investigate intricate data inputs. It isn't, all by itself, an algorithm. Such frameworks "learn" to tackle undertakings by looking at samples, as often as possible with practically no task explicit instructions being programmed. An artificial neural network (ANN) is a model that is based on a network of linked units or nodes known as "artificial neurons," which are roughly modeled after neurons in the human brain. Every link, like synapses in the human brain, may transfer data or a "signal," beginning with one artificial neuron and progressing to the next. When an artificial neuron receives a signal, it can decode it before sending it on to other artificial neurons. In ordinary ANN implementations, the signal at a link between artificial neurons is a real number, and the output of each artificial neuron is generated by some non-linear function of the sum of its inputs. "Edges" refer to the connections between artificial neurons. The weight of artificial neurons and edges is regularly modified as learning progresses. The weight affects signal strength at a link by expanding

or contracting it. Artificial neurons may have a baseline that prevents signal transmission until the entire signal exceeds the baseline. Artificial neuron layers are frequently used. Various layers may make distinct adjustments to their feedback bits. Signals are guided from the first (input layer) to the final (output layer), maybe many times. The basic goal of the ANN technique was to approach problems in the same manner that a human mind would. However, after a while, the attention shifted to certain tasks, resulting in biological abnormalities. Computer vision, voice recognition, machine interpretation, social network sorting, board and video games, and clinical diagnostics are all examples of how artificial neural networks have been employed. Deep learning is comprised of a few secret layers in an artificial neural network. This strategy endeavors to reenact how the human brain processes light and sound into vision and hearing. Deep learning has been utilized effectively in many aspects, including computer vision and speech recognition (Honglak, 1995).

### ***Support vector machines.***

Support vector networks, or SVMs, are supervised learning techniques used for categorization and regression. A model that predicts whether a new case falls into one of the two categories given a progression of training scenarios that are labeled as having a place with one of the two classifications is assembled by an SVM training strategy (Cortes, 1995). Platt scaling, for example, is a probabilistic technique to employing SVM in a probabilistic classification environment, yet an SVM training algorithm is a non-probabilistic, paired, linear classifier. The kernel technique enables SVMs to perform both non-linear and linear characterisation by verifiably interpreting their inputs into high-dimensional feature spaces.

### ***Bayesian networks***

This is a simple straightforward Bayesian network. Whether or not the sprinklers are turned on is influenced by the rain, and both the rain and sprinklers affect how moist the grass is. A Bayesian network, otherwise called a belief network or a directed non-cyclic graphical model, is a probabilistic graphical model that portrays a bunch of irregular factors and their contingent independence utilizing a directed acyclic chart (DAG). A Bayesian network, for instance, may be utilized to show the probability relationship among ailments and symptoms. In view of side effects, the network might be utilized to decide the probability of specific diseases being present. There are productive induction and learning algorithms. Dynamic Bayesian networks are Bayesian networks that model variable arrangements like voice signals or protein successions. Influence diagrams are Bayesian network generalization that might depict and tackle decision issues under uncertainty.

### ***Genetic algorithms***

A Genetic algorithm (GA) is a search algorithm and heuristic instrument that recreates natural selection by creating new genotypes through techniques such as mutation and crossover with expectations of uncovering ideal solutions to a specific issue. Genetic algorithms were utilized in machine learning during the 1980s and 1990s (Zhang, 2011). Conversely, machine learning algorithms have been utilized to enhance the performance of genetic and evolutionary algorithms.

---

## **Empirical Review**

### **Works Related to Breast Cancer Classification**

Classification is one of the most significant and critical jobs in machine learning and data mining. A lot of research has been done on various medical datasets utilizing data mining and machine learning to categorize Hepatitis A and B. Many of them have a high level of precision in categorization. Bandyopadhyay et al (2020) suggested an automated system that distinguishes patients with hepatitis syndromes in their work A Voting Ensemble Approach for Hepatitis Disease Detection. They subsequently developed an automated method based on a multi-phased classification technique. Several classifiers, including the Support Vector Machine, Multi-layer Perceptron, naive Bayes, k-Nearest Neighbor, and Decision Tree, were utilized in the early phase. Ada Boost, Gradient Boost, and Random Forest were employed as phase-2 classifiers. These implemented classifiers were analyzed and compared in terms of prediction performance. They created a voting-based ensemble technique for the final phase classification, which uses the top two classifier models from the first and second phase classifications, respectively. They argue that using the proposed classifier would increase prediction performance, allowing patients with hepatitis to be properly identified.

Yarasuri, Indukuri, and Nair's paper "Prediction of Hepatitis Disease Using Machine Learning Technique" (2019). They contrasted and compared several machine learning technologies and neural networks. They used the accuracy rate and mean square error as performance measures. Machine Learning (ML) methods such as Support Vector Machines (SVM), K Nearest Neighbor (KNN), and Artificial Neural Network (ANN) have been used to categorize and predict Hepatitis disease. They performed a brief examination into the aforementioned problem based on illness diagnostic prediction accuracy. All of the machine learning approaches were implemented and validated using the MATLAB program. Hepatitis Disease Prediction Oishi et al. advocated using an Effective Deep Neural Network (2020). Their study objective was to build a Deep Neural Network model that could predict the illness with accuracy. Their study was compared with activation functions to assure the best performance. The study made use of a standard dataset on hepatitis infections gathered from 154 people in India. They claimed that their model's accuracy for predicting hepatitis was 92.3 percent, which they said was higher than previous research or studies using other approaches.

Julker et al (2021) employed a multi-model method in their work Prediction of Hepatitis Disease. Some of the approaches employed include K-Nearest Neighbors, Naive Bayes, Support Vector Machine, Multi-Layer Perceptron, and Random Forest. They predicted hepatitis disease using a variety of data mining techniques. In addition, they proposed a viable strategy for increasing the performance of their prediction models. They dealt with missing values in their dataset by eliminating missing value observations. They were able to discover extraneous attributes by combining the info-gain feature selection approach with ranker search. To calculate prediction accuracy on the hepatitis disease dataset, classification methods such as K-Nearest Neighbors (KNN), Naive Bayes Support Vector Machine (SVM), Multi-Layer Perceptron (MLP), and Random Forest were utilized. They compared the classification models' performance by assessing accuracy, precision, recall, F1-score, and ROC. They claimed that removing observations with missing values and employing the info-gain feature selection technique in their prediction models improved their accuracy. Julker et al found that the Random Forest model performed the best, with a classification accuracy of 92.41 percent. Gowtham et al (2021) proposed a deep learning strategy to classifying Paddy disease. Their research aimed to assist a large number of farmers, particularly those involved in paddy farming, in understanding and forecasting the illness that would affect their crop. Their study demonstrated the accuracy of using deep neural networks to classify paddy leaf disease. On their dataset, they employed pre-processing techniques such as data augmentation and the median filter to reduce over fitting and improve model performance and accuracy.

PinarYildirim created Filter Based Feature Selection Methods for Risk Prediction in Hepatitis Disease (2015). A comparison of filter-based feature selection techniques was carried out using a well-known dataset (the hepatitis dataset), and four classification algorithms were used to test the performance of the algorithms. Following the application of feature selection methods, they claimed that Nave Bayes and Decision Table classifiers outperformed the others on the hepatitis dataset. According to them, the study indicated that feature selection techniques can improve the performance of learning algorithms. However, they claimed that no single feature selection technique based on filters is the best. Consistency Subset, Info Gain Attribute Eval, One-R Attribute Eval, and Relief Attribute Eval outperformed the others in general.

Mohamed et al. (2014) investigated the use of non-invasive liver fibrosis models to predict morbidity and mortality in chronic hepatitis C patients. They investigated Compensated CHC patients for seven years. Using Hepascore, FIB-4, APRI, and liver biopsy data, a list of non-invasive indicators was created. The Western Australian Data Linkage System was used to acquire follow-up morbidity and mortality data. The Kaplan–Meier approach was used to examine the prognostic relevance of baseline noninvasive indicators and biopsies. A total of 406 persons (64 percent male, mean age 48 11 years) were tracked for 2385 person-years, during which 22 (5.4 percent) died, 14 (3.4 percent) died from liver disease or required a liver transplant. Sixteen of the participants had hepatic decompensation (3.9 percent). Hepascore and liver biopsy ( $P = 0.005$ ), but not APRI or FIB-4, predicted overall and liver-related mortality and decompensation. A Hepascore greater than 0.5 was associated with an increased risk of overall mortality [HazardRatio (95percentCI)6.7(2.6–17), $P = 0.001$ ], liver-related mortality [32.8 (4.3–250),  $P = 0.001$ ], and future decompensation [11.8 (3.3–41),  $P = 0.001$ ], whereas a Hepascore less than 0.5 was associated with a 99 percent chance of not dying from liver-related causes over 10. With an AUROC of 0.86 (95 percent CI 0.80–0.90) and 0.87 (0.79–0.96), Hepascore was as accurate as liver biopsy in predicting liver-related mortality.

## Research Methodology

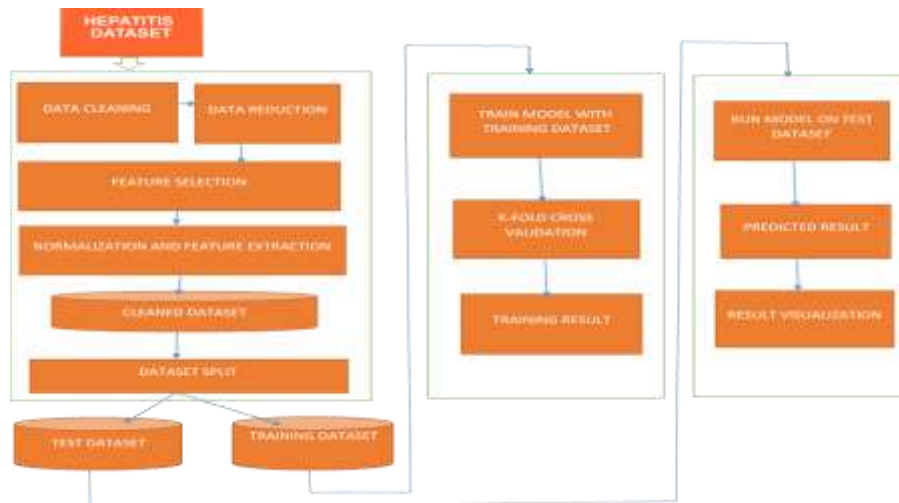
### Research Design

The methodology utilized in carrying out the project will be talked about in this part of the research study. It will be partitioned into three areas: a discourse on the proposed methodology and an overall discussion of its parts, point by point explanations of every aspect of the proposed methodology, including any pivotal algorithms, diagrams, and a discussion of the assessment strategy of the proposed methodology. Hepatitis is one of the most common liver illnesses in the world, accounting for the majority of new sickness cases and hepatitis-related mortality, making it a critical public health concern in today's society, according to worldwide data. Early hepatitis identification can help improve a patient's chances of survival since it allows medical providers to deliver focused therapy based on an accurate diagnosis. To minimize persons obtaining unnecessary drugs, hepatitis A and B should be categorized more precisely. As a result, the purpose of this research is to offer a reliable diagnosis of hepatitis A and B, whether present or absent. Because of its particular benefits in recognizing crucial characteristics from complicated hepatitis datasets, machine learning (ML) is generally recognized as the technique of choice in hepatitis pattern categorization and forecast modeling. A methodology is proposed that will obtain the UCI hepatitis dataset, carry out the required information preprocessing tasks on it in order to eliminate and data irregularities to prevent dimensionality thereby preventing model over fitting, partition the purified dataset into testing dataset and training dataset, utilizing the training dataset to train the model, and afterward validate the trained model utilizing the test dataset. The model will be built utilizing a two-dimensional convolutional neural networking conjunction with a bidirectional long-short term memory encoder. The bidirectional long-short term memory will effectively become familiar with the dataset and the two-dimensional convolutional neural network and will classify it into the suitable hepatitis A or B class.

The proposed methodology is shown in figure 3.1 below. The methodology is divided into three major parts namely:

- 1 Data gathering and preprocessing
- 2 Model plan, training and validation, and
- 3 Hepatitis discovery and assessment technique'





Sources, Researcher 2023

Figure 3 Research Design

## Data Collection and Preprocessing

The dataset utilized for the research activity coupled with the means through which the dataset will be processed to ensure its availability for usage in our analysis, is amply discussed in this phase of the methodology. The hepatitis dataset from UCI Wisconsin is being used. The next section delves into the dataset's intricacies.

### Data Preprocessing

After we've discovered, extracted, and purified the material we'll need for our use case, the next step is to understand it. These processes are used to preprocess the dataset. The first important problem is the enormous number of duplicated records in the Wisconsin (UCI Repository) dataset. We determined that around 68 percent and 65 percent of the entries in the Wisconsin (UCI Repository) dataset train and test sets are duplicated. Because the train set contains a large number of redundant records, learning algorithms will be biased toward more frequent records, preventing them from being learnt. As a result, the purpose of this research is to offer a reliable diagnosis of hepatitis A and B, whether present or absent. Because of its particular benefits in recognizing crucial characteristics from complicated hepatitis datasets, machine learning (ML) is generally recognized as the technique of choice in hepatitis pattern categorization and forecast modeling. Many research projects employ random subsets of the train set as test sets, which are how we got these statistics on both the train and test sets. As a result, they get a classification rate of roughly 98 percent utilizing fairly simple machine learning algorithms. Even utilizing the test set will yield in a minimum classification rate of 86 percent, making it difficult to compare Hepatitis disease detection algorithms because they all fall somewhere between 86 and 100 percent. We proposed a solution to the two problems mentioned in this study, resulting in new train and test sets comprised of chosen records from the full data set. In the data set supplied, none of the aforementioned concerns occur. Furthermore, the train and test sets have a sufficient number of records. This advantage allows the tests to be run on the complete collection rather than just a small sample at random. We used active learning to help us preprocess the dataset so that the deep learning system could grasp it. This is crucial since an accurate detection result requires a clean and non-redundant dataset in both the train and test sets. The objective of this part is to ensure that the dataset is in a format that the machine learning algorithm can understand. We also want to avoid over fitting or under fitting the model with data, and we want to reduce the dimensionality of the dataset so that each stated parameter contributes at least equally to the proposed model's prediction.

## Model Design, Training and Validation

In the approach, the creation of the model follows the selection and cleaning of the dataset. The machine learning methods BiLSTM and 2dConv were used. They're algorithms for supervised data processing. The model is explained further in the section that follows.

### Auto encoders

A form of unsupervised artificial neural network that learns appropriate data coding strategies is an auto encoder. An auto encoder aims to train a network to ignore signal "noise" in order to learn a representation (encoding) for a set of data, usually to reduce its dimensionality. Aside from the reduction component of the autoencoder, it also learns the reconstructing component. This component aims to build a representation as close to the original input as possible using the reduced encoding component, hence the name.

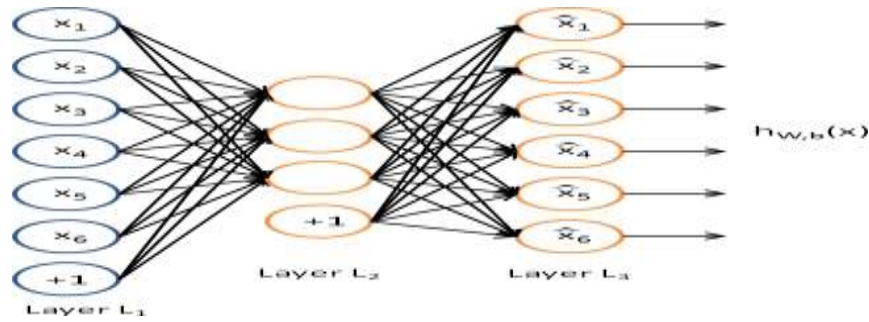


Figure 4 autoencoder

An autoencoder learns to compress data from the input layer into a short code, which it then decompresses into data that is virtually identical to the original data. As a result, the autoencoder is compelled to participate in dimensionality reduction activities such as learning to ignore noise.

**Bi-Lstm-Crf**

Another typical sequence concept is conditional random fields (CRF). They are a type of discriminative undirected probabilistic graphical model that exhibits a single log-linear dispersion over structured outputs as a function of an input perception sequence. The model should predict the values of arbitrary variables Y given conspicuous factors X, arbitrary variables Y, and an undirected organization G in which Y are connected by undirected edges exhibiting conditions. CRF defines the conditional probability of a set of output values yY given a set of input values xX as the outcome of various capacity on graph cliques.

$$p(y|x) = \frac{1}{Z_x} \prod_{s \in S(y,x)} \phi_s(y_s, x_s) \tag{2}$$

Where  $Z_x$  represents the overall output value normalization factor,  $S(y, x)$  represents the set of G cliques, and  $s(y_s, x_s)$  represents the clique potential on clique s. Following that, in the Bi-LSTM-CRF model, a softmax across all conceivable tag sequences yields a probability for the sequence y. The following is how the output sequence prediction is computed:

$$y^* = \text{argmax}_{y \in Y} \sigma(X, y) \tag{3}$$

Where  $\sigma(X, y)$  is the score function defined as follows:

$$\sigma(X, y) = \sum_{i=0}^n A_{y_i, y_{i+1}} + \sum_{i=0}^n P_{i, y_i} \tag{4}$$

The score of a transition from the tag  $y_i$  to  $y_{i+1}$  is denoted as  $A_{y_i, y_{i+1}}$ , where A is a matrix of transition scores.  $P_{i, y_i}$  is the score of the  $y_i$ th tag of the  $i$ th word in a sentence, P is the Bi-LSTM network's score matrix, and n is the length of a phrase.

**The Work Flow of the Proposed Model**

We've finished putting together our model and are ready to present it to you. Although the BiLSTM layer has been shown to be successful for dealing with temporal correlation, it has much too much redundancy with regards to spatial data. To resolve this issue, we propose an expansion of BiLSTM that incorporates convolutional structures in both the input-to-state and state to-state transitions. By stacking a numerous Convolutional layers and building an encoding-predictive structure utilizing the autoencoder, we might make a network model for hepatitis prediction as well as for more broad spatiotemporal sequence predicting issues. BiLSTM's essential shortcoming in taking care of spatiotemporal information is the use of whole connections in input-to-state and state to-state transitions when no topographical data is recorded. To address this trouble, we developed the BiLSTM and 2dCNN as 3D tensors with all inputs  $X_1, \dots, X_t$ , cell outputs  $C_1, \dots, C_t$ , hidden states  $H_1, \dots, H_t$ , and gates, as spatial dimensions (rows and columns). To easily comprehend the inputs and states, we can show them as vectors on a spatial grid. The BiLSTM and 2dCNN utilize the inputs and past states a cell's local neighbors to gauge its future state in the grid. A simple way to deal with do this is to involve a convolution operator in the state-to-state and input-to-state transitions. The key equations for the BiLSTM and 2dCNN are as per the following:

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)})$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)})$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)})$$

$$u_t = \tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)})$$

$$P(\mathcal{Y} | \mathcal{X}) = \frac{1}{Z_x} \prod_{s \in S(\mathcal{Y}, \mathcal{X})} \phi_s(y_s, x_s) \tag{2}$$

$$h_t = o_t \cdot \tanh(c_t)$$

Hepatitis predictions might be designed as a lineally isolated parametric function issue that can be solved using the traditional sequence-sequence learning framework and the aforementioned equation in conjunction with the 2dCNN. To adequately replicate or simulate the spatiotemporal connections, the concept of FC-LSTM is expanded to Bi-LSTM and 2d-CNN, which incorporate convolutional structures in both the input-to-state and state-to-state

transitions. By stacking many Bi-LSTM and 2d-CNN layers and building an encoding-forecasting framework, an end-to-end trainable model for hepatitis forecast may be obtained. We evaluate the Wisconsin hepatitis (UCI Repository) dataset, which might aid future research, particularly in developing machine learning techniques for the problem. To build a BiLSTM network for sequence-to-label classification, we need to do the following:

- 1 Create a layer array with a sequence input layer, a BiLSTM layer, a fully connected layer, a softmax layer, and a classification output layer.
- 2 Make the sequence input layer the same size as the number of features in the input data.
- 3 Set the number of classes as the size of the completely connected layer. The length of the sequence does not need to be specified.
- 4 We select the number of hidden units and the output mode 'last' for the BiLSTM layer.

The output from the BiLSTM network is utilized to train the one-dimensional neural network to group sequence data. We might utilize the BiLSTM organization cater for sequence data into a two-dimensional neural network and make forecasts in light of the sequence data's discrete time steps. The training data is then stacked from the BiLSTM output. Suppose that XTrain is an array variable. XTrain is a cell array that possesses N 32-dimensional sequences. The 32 features are denoted by Y, which is a category vector of labels. XTrain passages are matrices with 32 rows (one for each feature) and a variation of columns (one column for each time step). The two-dimensional convolutional neural network can then possess the capabilities to access and classify the data to make a prognosis.

## Data Analysis

### Implementation and Evaluation Experiment Analysis

This portion of this article dives into the details of the tests that were carried out. We present the results of the analyses and emphasize the work's prediction technique.

### Data and Parameters

The UCI hepatitis dataset was used to train the BiLSTM and 2d-CNN hepatitis classification models. The dataset used in this work is freely available to the public. Table 1 displays a sampling of the dataset:

**Table 1 Sample Dataset**

Category	Age	Sex	ALB	ALP	ALT	AST	BIL	CHE	CHOL	CREA	GGT	PROT
0=Blood Donor	32	m	38.5	52.5	7.7	22.1	7.5	6.93	3.23	106	12.1	69
0=Blood Donor	32	m	38.5	70.3	18	24.7	3.9	11.17	4.8	74	15.6	76.5
0=Blood Donor	32	m	46.9	74.7	36.2	52.6	6.1	8.84	5.2	86	33.2	79.3
0=Blood Donor	32	m	43.2	52	30.6	22.6	18.9	7.33	4.74	80	33.8	75.7
0=Blood Donor	32	m	39.2	74.1	32.6	24.8	9.6	9.15	4.32	76	29.9	68.7

**Source: Author 2023**

As displayed in the third chapter of this study, a BiLSTM and 2-dimensional convolutional neural network architecture is utilized in the prediction of the existence of either Hepatitis A or B. The design comprises of 2 networks which are: an autoencoder network, in conjunction with a multilayer perceptron network. The network feature selector is operated by the autoencoder, which as aforementioned in the third chapter of this study, is a BiLSTM coupled with a carry forward attribute to aid in treating feature in the time series component of the dataset. A multi-facet perceptron network handles classification and forecast. The autoencoder is the primary or original component of the article design. An autoencoder is a sort of unassisted network that extricates non-linear properties from input data. An autoencoder possesses three layers: an input layer, a secret layer that utilizes the sigmoid activation function, and an output layer. The autoencoder is setup so that the output layer endeavors to match the input layer as intently as achievable. The secret layer produces a non-linear compressed portrayal of the input layer because of the sigmoid activation function. The thought behind this change is that the information will be more compressed (i.e., less inclined to over-approximation) and that a few intriguing non-linear associations might be found, maybe expanding the clarification of the output variable. The autoencoder's hidden layer's non-linear compressed portrayal of the initial input is directly associated with a Multilayer perceptron, for example the BiLSTM is directly associated with the convolutional neural network. This convolutional network is responsible for making forecasts in our problem by utilizing the new issue portrayal as an input. To accomplish the reason for this research study, a few prerequisites for executing the proposed arrangement should be tended to.

We examined the age distribution of the dataset and this is shown in a histogram shown in figure 5

Figure 5 A Distribution of Age

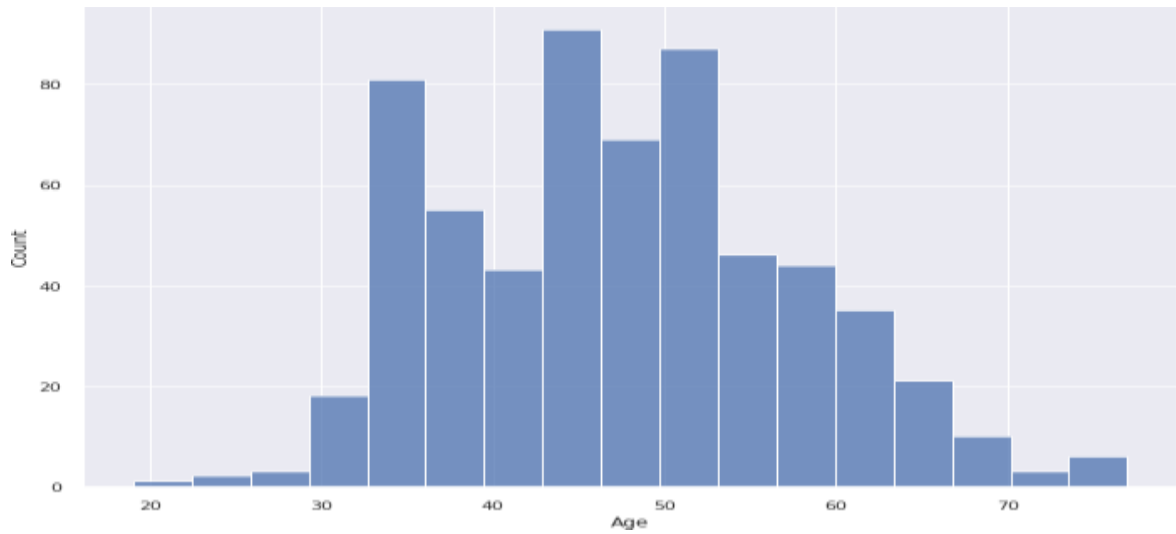


Figure 6 shows the distribution plot for the categories.

Sources, Researcher 2023

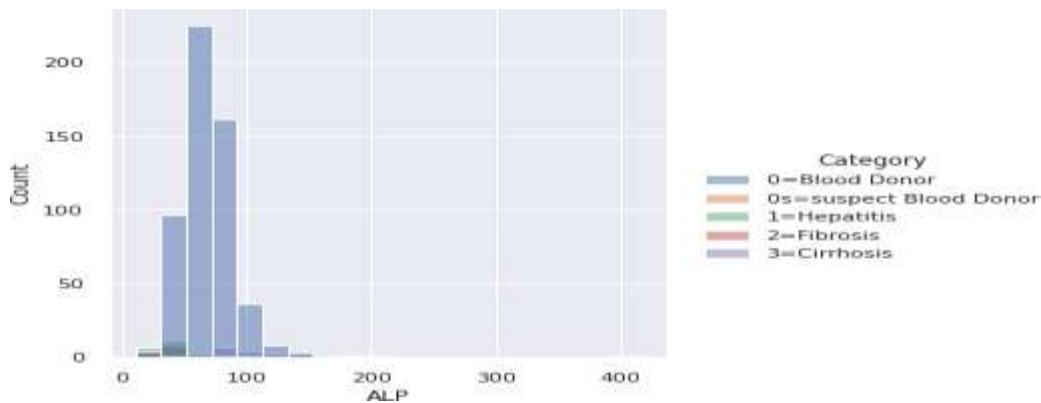


Figure 6 distribution plot of category

Figure 7 depicts a heat map of the dataset's correlations to highlight how they affect the prediction's outcome.



Sources, Researcher 2023

Heat map of correlation of parameters

Every cell in the correlation matrix typifies a 'correlation coefficient' between the two factors that symbolize the cell's row and column. A connection coefficient is referred to as a number that demonstrates how solid a relationship exists between two factors. There are various correlation coefficients to

take into account. Pearson's coefficient  $\rho$  (condensed as (rho)) is the most broadly used. It's evaluated by obtaining the quotient of the covariance between two factors and the product of the two factors' standard deviations.

#### Formula for correlation coefficient between two variables

Where  $COV(X, Y)$  is described as the "expected value of the product of X and Y's deviations from their particular means."

- The value of  $\rho$  ranges from -1 to+1.
- Values around +1 suggest that X and Y have a strong positive relationship, whereas values near -1 indicate that X and Y have a strong negative relationship.
- Values close to 0 indicate that there is no link between X and Y.

There appear to be some relationships in Biomarker levels, as seen by the heat map.

#### Optimizing the proposed network architecture

The autoencoder and neural network were implemented in a variety of network designs with varied settings. Trainings and tests were carried simultaneously on a cluster that possessed 34 nodes. Originally, various value criterion or guideline combinations were looked into in order to locate the optimum potential architecture configuration. Specifically, a grid search was run for the following set of parameters listed in Table 2

Table2	Parameter	Values
Autoencoder BiLSTM-CRF	Hidden neurons	[11, 13, 19, 23, 29, 35, 40, 45]
	Learning rate	[0.1, 0.3, 0.9]
	Corruption level	[0.1, 0.3, 0.9]
	Iterations	[33, 100, 300, 1000, 3000, 5000, 9000]
2dCNN	Hidden layers	[10, 13, 19,34]
	Learning rate	[0.1, 0.3, 0.9]
	Momentum factor	[0.1, 0.3, 0.9]
	Iterations	[33, 100, 300, 1000, 3000, 5000, 9000]

Source: Author 2023

We discovered that the following setup for the autoencoder was the best after conducting the tests: 5000 training cycles, 34 hidden neurons, 0.9 learning rate, 0.1 corruption level The best result was achieved with the following values for the 2dCNN configuration: 19 hidden layers, 0.3 learning rate, 0.9 momentum factor, and 1000 training iterations. Figure 4.8 depicts the model's whole set of parameters, both trainable and non-trainable, as well as its forms.

Layer (type)	Output Shape	Param #
densenet201 (Model)	(None, 7, 7, 1920)	18321984
global_average_pooling2d_1 (	(None, 1920)	0
dropout_1 (Dropout)	(None, 1920)	0
batch_normalization_1 (Batch	(None, 1920)	7680
dense_1 (Dense)	(None, 2)	3842
-----		
Total params: 18,333,506		
Trainable params: 18,100,610		
Non-trainable params: 232,896		

Figure 8 the baseline of the model

In this demonstration, 70% of the samples were utilized for training and 30% of the samples were used for testing. Steps were taken to ensure that the individuals (Patients) that were selected for training were not reused during testing so as to generalize the task of classification and perform satisfactorily when testing new patients.

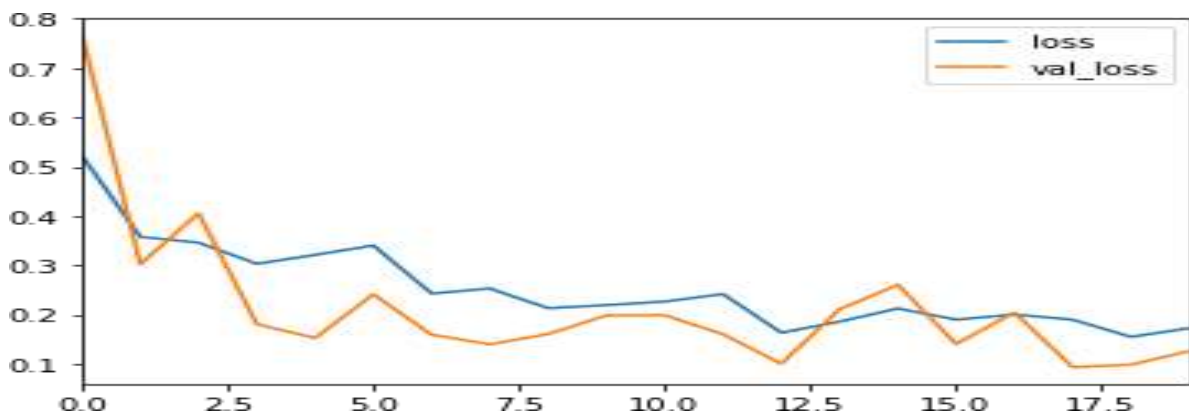
**Source: Author 2023**

The training dataset's loss function is shown in Figure 4.9. Both row-wise and column-wise evaluations were used in our experiments. This is done to supplement the training and test datasets, which must be degraded in order for the dataset to be processed efficiently. We used three alternative approaches for row-wise classification:

**STEP 1:** We converted the input samples to a vector, which dramatically reduced the amount of information in the samples.

**STEP 2:** For each sample, multiple data augmentation techniques were used to alter the dataset, resulting in 20 different augmented samples.

**STEP 3:** To generate a parameter database for training and testing the model, 200 random columns were retrieved. To get the results, we employed a procedure in which the final class was decided based on the class with the most parameters nominated. The parameter labels are assumed to be of the same class as the original set of parameters.

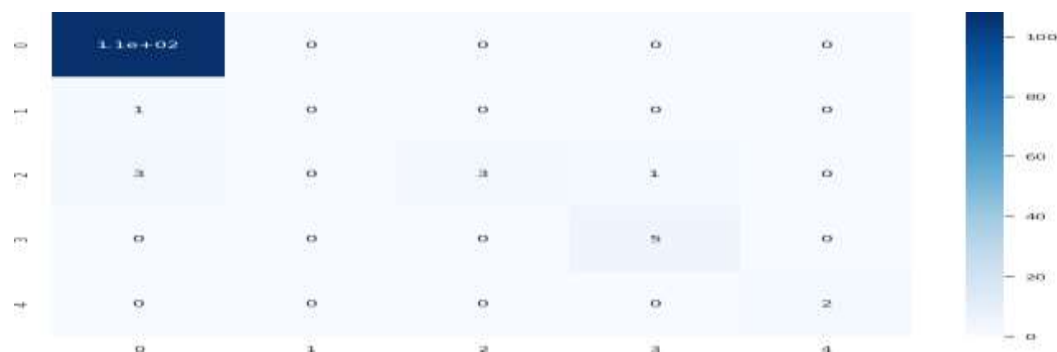


**Figure 9** loss function of the training dataset

The accuracy value for the model on the test dataset is shown in Figure 4.10. We used a variety of data augmentation techniques in each dataset, including 40-degree sequential rotation, 0.2-factor width shift, 0.2-factor height shift, 0.2-factor shear, 0.2-factor zooming, horizontal flipping, and vertical flipping. In the first experiment, we used the stochastic gradient descent (SGD) optimization function to train the proposed model. We set the momentum to 0.9, and the decay is determined using the trial's starting rate of learning and amount of epochs. We tried for three trials, using 500 epochs each time. The rate of learning reduces by a factor of 10 after 500 epochs.

**Figure 10** accuracy prediction of the test dataset

The most common metric for evaluating model performance is the accuracy.



**Source: Author 2023**

**Confusion matrix for the predictions**

The proposed model's confusion matrix is shown in Figure 4.11. With regards to misclassification, the Confusion Matrix is an essential measure to consider. The examples in an anticipated class are depicted by the rows of the matrix, while the occurrences in a genuine class are depicted by the columns. The diagonals show which classes have been effectively classified. The Key Performance Indicators (KPIs) that were used to evaluate the system. KPIs which are listed below, are ascertained with the aid of the confusion network

		Predicted class	
		Class = Yes	Class = No
Actual Class	Class = Yes	True Positive	False Negative
	Class = No	False Positive	True Negative

Figure 9 confusion matrix

The perceptions that were accurately predicted and henceforth depicted in green are valid or true positives negatives. Since we aim to diminish misleading or false positives and negatives to the least, they're set apart in red. These articulations are somewhat puzzling. So we should go through each phase individually and ensure we comprehend it well.

**Valid Positives (TP)** - These are actually expected positive characteristics, showing that the value of the real class is 'yes,' very much like the value of the expected class.

**Valid Negatives (TN)**- These are unequivocally anticipated negative characteristics, inferring that the value of the real class is 'no' and the worth of the predicted class is also 'no.'

Values that occur when the actual and predicted classes do not match are known as False positives and negatives.

**False Positives (FP)**- This happens once the genuine class is "no" but the normal class is "yes."

**False Negatives (FN)**- This happens once the genuine class is "yes," yet the projected class is "no."

Our model's Accuracy, Precision, Recall, and F1 score can all be evaluated once these four parameters are known.

**Accuracy** - The most essential and direct performance statistic is accuracy, which is only the proportion of foreseen observations that have been accurately forecasted to total observations. It's a far and wide misperception that more exactness equates to a superior model; nevertheless, this is quite true if the dataset is symmetric, and that implies that false positives negatives have essentially identical values. The formula for evaluating accuracy is as per the following:

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN})$$

**Precision**- The proportion of accurately anticipated positive perceptions to total anticipated positive perceptions is known as precision. This matrix attempts to resolve the issue of the number of true positives are actually positive and by how much. The following is the formula:

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

**Recall (Sensitivity)** - The proportion of precisely anticipated positive perceptions to the total perceptions in the real class is known as review. The equation can be located beneath.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

**F1 score** - The weighted average of Precision and Recall is known as the F1 Score. Therefore, this score considers both false positives and negatives. Albeit not as easily comprehend when compared to accuracy, F1 is regularly more valuable than precision, especially when the class distribution is uneven. At the point when the cost of false positives and negatives are equivalent, accuracy functions best. On the off chance that the costs of both false positives and false negatives are unique, looking at both Precision and Recall is ideal. The recipe is given below:

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

Table 4 shows the results of the Precision, F-Score, and Recall tests:

Table 4 result of Precision, F - Score and Recall for the proposed solution

Accuracy	Precision	Recall	F1 score	ROC-AUC
98.3%	0.65	0.95	0.77	0.692

Source: Author 2023

The best results for each model execution are recorded in the table above. This includes the main, second, third, fourth, and last runs.

---

## Summary of findings

Seroclearance of HBsAg sometime been respected to be a huge indicator of CHB prognosis. Numerous past examinations on significant subjects have exhibited that applying machine learning algorithms to foresee sickness status or results with clinical datasets is acquiring expanding awareness in the clinical and health arena. In this paper, we proposed a functioning learning method on the hepatitis dataset utilizing a bidirectional long short term memory and a two-dimensional convolutional neural network model. The trials were carried out on the UCI hepatitis dataset utilizing the proposed model, and the outcomes were surveyed utilizing an assortment of performance measures. The performance of the proposed procedure was assessed utilizing data level, patient level, row based, and column based investigation. Numerous boundaries were considered in this execution, including class, age, sex, steroid, antivirals, tiredness, malaise, anorexia, liver big, liver firm, spleen palpable, spiders, ascites, varices, bilirubin, alk phosphate, sgot, albumin, protime, and histology

---

## Conclusion and Recommendation

The proposed model worked well, with an accuracy of 98.3 percent, precision of 0.65, recall of 0.95, and an F1 score of 0.69. The fact that we adopted a new strategy of learning the dataset before conducting the prediction was largely responsible for the boost in performance. It shows that if a model comprehends a problem domain before doing any type of classification, it will perform better than if it is introduced to a new problem domain. This is a type of transfer learning where a model actively learns the dataset before passing it on to another model for categorization. The model's enhanced performance could be due to a variety of variables. It's worth noting that the model was only tested on the UCI dataset and not on any others, and the accuracy reported is only dependent on this fact.

This research has enhanced hepatitis disease prediction, but there will be a difficulty when it comes to overlapping hepatitis cells. It will be fascinating to see if it can tell the difference between overlapping hepatitis cells and single hepatitis cells. In the future, this paper recommend looking into issues such cell overlap and uneven color distribution in hepatitis pathological data obtained using various staining methods. Area thickness analysis employing a known overlapping dataset with well-defined perimeter and fractal diameter will be required to distinguish overlapping cells from single ones. The thickness and color distribution analyses will also present important issues that will be need to be addressed in future research.

---

## References

- A. K. Waljee and P. D. R. Higgins, "Machine learning in medicine: a primer for physicians," *American Journal of Gastroenterology*, vol. 105, no. 6, pp. 1224–1226, 2010.
- Andino R, Domingo E. Viral quasispecies. *Virology*. 2015 May;479-480:46-51. doi: 10.1016/j.virol.2015.03.022. Epub 2015 Mar 29. PMID: 25824477; PMCID: PMC4826558.
- Andronie, M. (2010). Economic process optimization deployed by companies with complex activity production–trade–services using informatics systems with integrated data mining techniques. *Annals of SpiruHaret University*, 1(1), 77-84.
- Aswathy, M. A., & Jagannath, M. (2017). Detection of breast cancer on digital histopathology images: Present status and future possibilities. *Informatics in Medicine Unlocked*, 8, 74-79.
- Bandyopadhyay, Samir & Dutta, Shawni. (2020). A Voting Ensemble Approach for Hepatitis Disease Detection. *International Journal For Research in Applied Sciences and Biotechnology*. 7. 56-62. 10.31033/ijrasb.7.5.6.
- Bengio, Y., Courville, A., and Vincent, P. 2013. "Representation Learning: A Review and New Perspectives". *IEEE Trans. PAMI, Special Issue Learning Deep Architectures*. 35 (8): 1798–1828. arXiv:1206.5538. doi:10.1109/tpami.2013.50. PMID 23787338.
- Bishop, C. M. 2006., *Pattern Recognition and Machine Learning*, Springer, ISBN 978-0-387-31073-2 breast cancer histopathological image classification," *IEEE Transactions*
- C.-M. Chu and Y.-F. Liaw, "HBsAgseroclearance in asymptomatic carriers of high endemic areas: appreciably high rates during a long-term follow-up," *Hepatology*, vol. 45, no. 5, pp. 1187–1192, 2007.
- C.-M. Chu and Y.-F. Liaw, "Hepatitis B surface antigen seroclearance during chronic HBV infection," *Antiviral Therapy*, vol. 15, no. 2, pp. 133–143, 2010.
- Caligiuri, P., Cerruti, R., Icardi, G., & Bruzzone, B. (2016). Overview of hepatitis B virus mutations and their implications in the management of infection. *World journal of gastroenterology*, 22(1), 145–154. <https://doi.org/10.3748/wjg.v22.i1.145>
- Evans, J., Pinato, D., Ward, C., THURZ, M., & Sharma, R. (2019). Longitudinal monitoring of cell-free DNA predicts for transarterial chemoembolization failure in hepatocellular carcinoma. *J Hepatol*, 70, e362.
- F. Bortolotti, M. Guido, S. Bartolacci et al., "Chronic hepatitis B in children after e antigen seroclearance: final report of a 29-year longitudinal study," *Hepatology*, vol. 43, no. 3, pp. 556–562, 2006.
- F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for



- F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for breast cancer histopathological image classification," *IEEE Transactions on Biomedical Engineering*, 2016.
- Fraser, M., Elgamal, A., Conte, J. P., Masri, S., Fountain, T., Gupta, A., ... & El Zarki, M. (2003, August). Elements of an integrated health monitoring framework. In *Smart Nondestructive Evaluation and Health Monitoring of Structural and Biological Systems II* (Vol. 5047, pp. 231-242). International Society for Optics and Photonics.
- Friedman, Jerome H. 1998. "Data Mining and Statistics: What's the connection?". *Computing Science and Statistics*. 29 (1): 3–9.
- G. Marx, S. R. Martin, J. F. Chicoine, and F. Alvarez, "Long-term follow-up of chronic hepatitis B virus infection in children of different ethnic origins," *Journal of Infectious Diseases*, vol. 186, no. 3, pp. 295–301, 2002.
- González García, C., NúñezValdéz, E. R., GarcíaDíaz, V., PelayoGarcía-Bustelo, B. C., & Cueva Lovelle, J. M. (2019). A review of artificial intelligence in the internet of things. *International Journal Of Interactive Multimedia And Artificial Intelligence*, 5.
- Gowtham Kishore Indukuri, Vedha Krishna Yarasuri, Aswathy K Nair, "Paddy Disease Classifier using Deep learning Techniques", *Trends in Electronics and Informatics (ICOEI) 2021 5th International Conference on*, pp. 1602-1607, 2021.
- H.-Y. Hsu, M.-H. Chang, C.-Y. Lee, J.-S. Chen, H.-C. Hsu, and D.-S. Chen, "Spontaneous loss of HBsAg in children with chronic hepatitis B virus infection," *Hepatology*, vol. 15, no. 3, pp. 382–386, 1992.
- Harnad, Steva. 2008. "The Annotation Game: On Turing (1950) on Computing, Machinery, and Intelligence", in Epstein, Robert; Peters, Grace, *The Turing Test Sourcebook: Philosophical and Methodological Issues in the Quest for the Thinking Computer*, Kluwer
- J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.
- J. Liu, H. I. Yang, M. H. Lee et al., "Incidence and determinants of spontaneous hepatitis B surface antigen seroclearance: a community-based follow-up study," *Gastroenterology*, vol. 139, no. 2, pp. 474–482, 2010.
- J. Liu, H.-I. Yang, M.-H. Lee et al., "Spontaneous seroclearance of hepatitis B seromarkers and subsequent risk of hepatocellular carcinoma," *Gut*, vol. 63, no. 10, pp. 1648–1657, 2014.
- J.-F. Wu, H.-Y. Hsu, Y.-C. Chiu, H.-L. Chen, Y.-H. Ni, and M.-H. Chang, "The effects of cytokines on spontaneous hepatitis B surface antigen seroconversion in chronic hepatitis B virus infection," *Journal of Immunology*, vol. 194, no. 2, pp. 690–696, 2015.
- Kimball, R., & Merz, R. (2000). The data webhouse toolkit: Building the Web-enabled data warehouse. *Industrial Management & Data Systems*.
- Li, J., Bioucas-Dias, J. M., & Plaza, A. (2012). Semisupervised hyperspectral image classification using soft sparse multinomial logistic regression. *IEEE Geoscience and Remote Sensing Letters*, 10(2), 318-322.
- Liang, D., Liu, Q., Zhou, K., Jia, W., Xie, G., & Chen, T. (2020). IP4M: an integrated platform for mass spectrometry-based metabolomics data mining. *BMC bioinformatics*, 21(1), 1-16.
- Liu J, Yang HI, Lee MH, Lu SN, Jen CL, Batrla-Utermann R, Wang LY, You SL, Hsiao CK, Chen PJ, Chen CJ; R.E.V.E.A.L.-HBV Study Group. Spontaneous seroclearance of hepatitis B seromarkers and subsequent risk of hepatocellular carcinoma. *Gut*. 2014 Oct;63(10):1648-57. doi: 10.1136/gutjnl-2013-305785. Epub 2013 Nov 13. PMID: 24225939.
- M. F. Yuen, D. K. H. Wong, J. Fung et al., "HBsAg Seroclearance in chronic hepatitis B in Asian patients: replicative level and risk of hepatocellular carcinoma," *Gastroenterology*, vol. 135, no. 4, pp. 1192–1199, 2008.
- M. H. Zheng, W. K. Seto, K. Q. Shi et al., "Artificial neural network accurately predicts hepatitis B surface antigen seroclearance," *PLoS One*, vol. 9, no. 6, Article ID e99422, 2014.
- Manjunath, T. N., Hegadi, R. S., & Ravikumar, G. K. (2010). A survey on multimedia data mining and its relevance today. *IJCSNS*, 10(11), 165-170.
- Manjunath, T. N., Hegadi, R. S., & Ravikumar, G. K. (2010). A survey on multimedia data mining and its relevance today. *IJCSNS*, 10(11), 165-170. F. Spanhol, L. S. Oliveira, C. Petitjean, and L. Heutte, "A dataset for
- Md. JulkerNayeem, Sohel Rana, FarjanaAlam, Md. Ataur Rahman, "Prediction of Hepatitis Disease Using K-Nearest Neighbors Naive Bayes Support Vector Machine Multi-Layer Perceptron and Random Forest", *Information and Communication Technology for Sustainable Development (ICICT4SD) 2021 International Conference on*, pp. 280-284, 2021.
- Mesquita, R. D., Vionette-Amaral, R. J., Lowenberger, C., Rivera-Pomar, R., Monteiro, F. A., Minx, P., ... & Oliveira, P. L. (2015). Genome of *Rhodnius prolixus*, an insect vector of Chagas disease, reveals unique adaptations to hematophagy and parasite infection. *Proceedings of the National Academy of Sciences*, 112(48), 14936-14941.
- Mitchell, T. 1997. *Machine Learning*. McGraw Hill. p. 2. ISBN 978-0-07-042807-2.

- Muir, A. J., Levy, C., Janssen, H. L., Montano-Loza, A. J., Shiffman, M. L., Caldwell, S., ... & GS-US-321-0102 Investigators. (2019). Simtuzumab for primary sclerosing cholangitis: phase 2 study results with insights on the natural history of the disease. *Hepatology*, 69(2), 684-698.
- Narendra Mohan, Vinod Jain, "Performance Analysis of Support Vector Machine in Diabetes Prediction", *Electronics Communication and Aerospace Technology (ICECA) 2020 4th International Conference on*, pp. 1-3, 2020. of Biomedical Engineering, 2016.
- OishiJyoti, Nazmin Islam, Fahim MD. SifnatulHasnain, "Prediction of Hepatitis Disease Using Effective Deep Neural Network", *Innovation in Technology (INOCON) 2020 IEEE International Conference for*, pp. 1-5, 2020.
- P. C. Austin, J. V. Tu, J. E. Ho, D. Levy, and D. S. Lee, "Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes," *Journal of Clinical Epidemiology*, vol. 66, no. 4, pp. 398-407, 2013.
- Perry, G. L., & Dickson, M. E. (2018). Using machine learning to predict geomorphic disturbance: The effects of sample size, sample prevalence, and sampling strategy. *Journal of Geophysical Research: Earth Surface*, 123(11), 2954-2970.
- R. Idilman, K. Cinar, G. Seven et al., "Hepatitis B surface antigen seroconversion is associated with favourable long-term clinical outcomes during lamivudine treatment in HBeAg-negative chronic hepatitis B patients," *Journal of Viral Hepatitis*, vol. 19, no. 3, pp. 220-226, 2012.
- R. Moucari, A. Korevaar, O. Lada et al., "High rates of HBsAg seroconversion in HBeAg-positive chronic hepatitis B patients responding to interferon: a long-term follow-up study," *Journal of Hepatology*, vol. 50, no. 6, pp. 1084-1092, 2009.
- S. H. Walker and D. B. Duncan, "Estimation of the probability of an event as a function of several independent variables," *Biometrika*, vol. 54, no. 1-2, pp. 167-179, 1967.
- Samuel, A. L. (1988). Some studies in machine learning using the game of checkers. II—recent progress. *Computer Games I*, 366-400.
- Scott Patterson. 2010. "Letting the Machines Decide". The Wall Street Journal. Retrieved 24 June 2018. [Show Article Full Text: PDF \(150KB\)](#) [Google Scholar](#)
- T. C. Tseng and J. H. Kao, "HBsAgseroclearance: the more and earlier, the better," *Gastroenterology*, vol. 136, no. 5, pp. 1843-1844, 2009.
- T. C. Tseng, C. J. Liu, T. H. Su et al., "Serum hepatitis B surface antigen levels predict surface antigen loss in hepatitis B e antigen seroconverters," *Gastroenterology*, vol. 141, no. 2, pp. 517-525, 2011.
- T. Chen and C. Guestrin, "XGBoost: a scalable tree boosting system," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785-794, ACM, San Francisco, CA, USA, August 2016.
- T. K. Ho, "Random decision forests," in *Proceedings of the 3rd International Conference on Document Analysis and Recognition*, p. 278, Montreal, Canada, August 1995.
- Tian, Z., Shen, C., Chen, H., & He, T. (2019). Fcos: Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 9627-9636).
- Tillmann, A. M. 2015. "On the Computational Intractability of Exact and Approximate Dictionary Learning". *IEEE Signal Processing Letters*. 22 (1): 45-49. arXiv:1405.6664. Bibcode:2015ISPL...22...45T. doi:10.1109/LSP.2014.2345761.
- Toivonen, J., Montoya Perez, I., Movahedi, P., Merisaari, H., Pesola, M., Taimen, P., ... & Jambor, I. (2019). Radiomics and machine learning of multisequencemultiparametric prostate MRI: Towards improved non-invasive prostate cancer characterization. *PLoS One*, 14(7), e0217702.
- Troyanskaya O, Cantor M, Sherlock G, Brown P, Hastie T, Tibshirani R, Botstein D, Altman RB (2001) Missing value estimation methods for DNA microarrays. *Bioinformatics* 17(6):520-525
- V. K. Yarasuri, G. K. Indukuri and A. K. Nair, "Prediction of Hepatitis Disease Using Machine Learning Technique," *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 2019, pp. 265-269, doi: 10.1109/I-SMAC47947.2019.9032585.
- Vos, T., Allen, C., Arora, M., Barber, R. M., Bhutta, Z. A., Brown, A., ... & Boufous, S. (2016). Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The lancet*, 388(10053), 1545-1602.
- W. H. Wolberg, W. N. Street, O. L. Mangasarian, 2019. Breast cancer Wisconsin (diagnostic) data set, UCI Machine Learning Repository [\[http://archive.ics.uci.edu/ml/\]](http://archive.ics.uci.edu/ml/).
- Wu D, Chou HH, Chien WH, Wu LL, Cheng CH, Chung CH, Horng JH, Ni YH, Tseng HT, Wu D, Lu X, Wang HY, Chen PJ, Chen DS. Age-related immune clearance of hepatitis B virus infection requires the establishment of gut microbiota. *Proc Natl AcadSci U S A*. 2015 Feb 17;112(7):2175-80. doi: 10.1073/pnas.1424775112. Epub 2015 Feb 2. PMID: 25646429; PMCID: PMC4343154.
- Y.-F. Liaw and C.-M. Chu, "Hepatitis B virus infection," *The Lancet*, vol. 373, no. 9663, pp. 582-592, 2009.

Zhang, J., Zhan, Z., Lin, Y., Chen, N., Gong, Y., Zhong, J., Chung, H. S.H., Li, Y., and Shi, Y. 2011. "Evolutionary Computation Meets Machine Learning: A Survey" (PDF). *Computational Intelligence Magazine*. 6 (4): 68–75. doi:10.1109/mci.2011.942584.