Journal of Population Therapeutics & Clinical Pharmacology

RESEARCH ARTICLE
DOI: 10.47750/jptcp.2023.30.08.028

An Efficient Deep Learning Model for Disease Prediction Using Cnn

P. Sathishkumar[1*], Swathi R[2], Yukesh Balaji K[3], Ruthra Kumar V[4]

[1]Assistant Professor in Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode

[2,3,4]Students of Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode

***Corresponding author:** P. Sathishkumar, Assistant Professor in Computer Science and Engineering, K.S.Rangasamy College of Technology, Tiruchengode

Submitted: 10 February 2023; Accepted: 11 March 2023; Published: 19 April 2023

ABSTRACT

Predicting a patient's likelihood of acquiring a disease is an important area of study in healthcare. Because of the vast amount of data growth in the biomedical and healthcare fields, precise medical data analysis has become advantageous for early illness identification and patient treatment. The Load function is useful in a variety of scenarios, all of which revolve around retrieving data from a given data source and adding it to the current data container. Machine learning (ML) is critical in Computer Aided Diagnostic testing. Body organs, for example, cannot be appropriately identified using a basic equation

Thus, pattern recognition requires training via illustrations. Pattern recognition and artificial intelligence have the potential to increase the reliability of disease approaches and detection in the biomedical field. They also value the impartiality of the decision-making process. ML provides a credible technique for developing improved and automated algorithms for the analysis of large-dimension and multi-modal biomedical data. We create a unique temporal fusion CNN framework to train patient representations while also measuring pairwise similarity. Unlike typical CNNs, our time fusion CNN can learn local temporal associations and contributions from each time period.

Keywords: *Learning, Model, Value, Data, Analysis, Source*

1.INTRODUCTION

1.1 Deep Learning

Music has a huge effect on people's lives. It is an important kind of entertainment and is frequently utilized as a form of treatment. Complex music players with a wide range of functions, such as volume modulation and genre classification, have been made feasible by technical advancements and continual multimedia breakthroughs. Even if this tool meets their needs, users must still search their playlists for music that conveys their emotions.

A user of a traditional music player must individually look through his playlist and select tunes that will improve his emotional condition. Because this method of song selection is time-consuming and complicated, the user may have difficulty picking the appropriate music. The emotion module, the random music player module, and the queue-based module make up our music player. Using the CNN technique, the Emotion Module reliably determines the user's mood from an image of their face.

1.2 Artificial Intelligences

The Artificial Intelligence lesson gives an overview of AI that will assist you in understanding the ideas behind AI. We also explored deep learning, machine learning, natural language processing, reinforcement learning, Q-learning, intelligent agents, numerous search techniques, and so on in this lesson. Our AI course is written at an elementary level so that you may grasp the entire instruction from fundamental principles to high-level ones. In the modern world, technological advancements are occurring swiftly, and we are continuously exposed to new developments. Artificial intelligence is a rapidly developing field of computer science that has the potential to herald in a new age by enabling the development of intelligent machines. The use of artificial intelligence has become widespread.

1.3 Deep Learning Techniques

Deep learning is a method of teaching computers to learn by doing, just like people do naturally. Self-driving cars can identify a stop sign or distinguish between a pedestrian and a lamppost thanks to deep learning, a key technology supporting them. For voice control in consumer devices like phones, tablets, televisions, and hands-free speakers is necessary. Recent years have seen a lot of interest in deep learning and for good reason. It is achieving results that weren't imaginable before.

2. LITERATURE REVIEW

2.1 Comparison Of Svm Classification Techniques And Navies Bayes For Predicting Bacterial Diseases Shameem

FATHIMA and colleagues proposed in this study We give a performance study of several data mining algorithms for predicting the Arboviral disease-Dengue in this research. The data set utilized in the analysis uses current data gathered from super specialty hospitals and diagnostic centers where blood samples are processed for diagnosis-related inquiries during enrolment in a course and once more upon leaving the hospital. This data collection has 5000 records and 29 parameters. We studied two data mining approaches in this paper: SVM and Naive Bayes Classifier. A competent methodology - The identification of tiny groups of parameters for use in diagnostic procedures in clinical practices is made possible by the random forest classifier and its related Gini feature importance; this entails accumulating the fewest symptoms possible while still achieving satisfactory dengue illness prediction performance. We integrate both techniques and assess the performance of the evaluators. The comparison of the approaches revealed that SVM outperforms Nave Bayes in the diagnosis of Dengue sickness. Data mining classification algorithms may be valuable for clinical diagnostic decision assistance. Chikungunya is a devastating illness produced by the Otariidae virus family.

2.2 Malaysian Dengue Outbreak Detection Model Feature Selection Algorithms

HUSAMet.al. proposed in this research that one of the most common diseases spread by mosquitoes worldwide is dengue fever. Identification of dengue outbreaks can be very helpful for practical efforts to slow the disease's rapid spread by giving the knowledge to predict the next outbreak's emergence. Using various data mining approaches, several research has been undertaken to model recast dengue outbreaks. Particle swarm optimization (PSO), genetic algorithms (GA), and rank search were employed in this work to identify the best features that increase the forecast accuracy of dengue epidemics (RS). Based on the listed variables, three predictive modeling methods (J48, DTNB Naive Bayes) were employed to identify dengue outbreaks. The Public Health Department in Seremban, Negeri Sembilan, Malaysia supplied the dataset for this study. The experimental findings demonstrated that using a feature selection approach before predictive modeling enhanced predicted accuracy. The study also demonstrated a set of attributes for Malaysian health officials to use in detecting dengue outbreaks. Dengue fever is regarded as one of the most frequent and potentially fatal infections in the world (Edelman 2007).

J Popul Ther Clin Pharmacol Vol 30(8):e262–e267; 19 April 2023.

2.3 Effective Liver Disorder Analysis And Diagnosis By Data Mining

In this study, SANJAY KUMAR et al. offer several liver problems that require a practitioner or another healthcare expert. With this study, we shall provide a classification technique of liver dataset patients to predict or forecast proper diagnosis of the liver utilizing multiple algorithms for data mining. An actual patient dataset was utilized to create classification algorithms to predict the diagnosis of the liver. To assess the performance of these classifiers, Five classification algorithms were used for a specific dataset and examined characteristics like accuracy, recall, and precision. To improve the effectiveness of classification models, investigations have been made. Finally, the findings demonstrated encouraging outcomes in the early detection of a liver condition. One of the largest internal organs in the human body is the liver. It weighs roughly 3 pounds. The liver is located underneath the diaphragm on the abdominal cavity's right side. Blood is transported to the liver by two big veins: the Hepatic Artery and the Portal Vein. This artery carries blood from the arteria, which is oxygenated. (a major vessel in the heart). These blood arteries branches continuously throughout the liver, eventually terminating in very small capillaries. Every liver problem can cause inflammation that may be acute or chronic, weakened liver function, and potentially injury to other internal organs. Alcohol-related sickness is still a significant disadvantage. This document also covers the biopsy performed when someone is suffering from a liver ailment. Alcalescent phosphotase, alanine aminopherase, aspartate aminopherase, and gamma-glutamyl transpeptidase are examples of enzymes.

2.4 Classification Of Liver Patients Using Intelligent Techniques

In this research, Anju Gulia et al. suggested that for a more accurate categorization than a single classifier, classification algorithms have been frequently used in the medical field. This research describes computational intelligence strategies for classifying people with liver disease. This study evaluates the chosen classification algorithms for the categorization of liver patient datasets (J-48, Multi-Layer Perceptron, Support Vector Machine, Random Forest, and Bayesian Network). In three stages, this research employs hybrid model development and comparison analysis to improve the prediction accuracy of liver patients. The first phase involves applying classification algorithms to the initial patient datasets for the liver obtained from the UCI repository. The second phase involves the collection of a subset (data) of patients with liver disease from entire datasets of patients with liver disease using feature selection, followed by the application of certain classification algorithms to the subset of features that have been acquired. Before performing feature selection, the SVM method is believed to be the best-performing algorithm when compared to other classification algorithms, it offers more accuracy. However, The Random Forest approach is thought to be the best-performing algorithm after feature selection, nevertheless. The third phase compares the results of classification techniques that include and exclude feature selection. The results of our trials show that the Random Forest method surpassed all other strategies using feature selection, with an accuracy of 71.8696%. The largest internal organ in the human body, the liver is critical to several important bodily functions, including metabolism. The liver is the body's biggest glandular organ. It weighs around 3 lb (1.36 kg).

2.5 Prediction Of Liver Disease Using Svm And Nave Bayes Algorithms

In this study, Dr. S. Vijayarani et al. suggest In recent years, data mining has become a popular tool for illness prediction in the healthcare sector. Data mining is the process of extracting information from large databases, warehouses, or other sources. It is a difficult issue for academics to anticipate illnesses using massive medical datasets. To address this issue, researchers employ categorization, clustering, and association rules are examples of data mining approaches. The primary goal of this research is to use classification algorithms to anticipate liver disorders. The algorithms used in this paper are Nave Bayes and support vector machines (SVM). Performance parameters like classification accuracy and execution time are used to compare

these classifier algorithms. based on the results of the experiment, the SVM is a better classifier for predicting liver disorders. Researchers in the healthcare sector confront a more difficult problem in predicting illnesses using massive medical datasets. Data mining is becoming increasingly important in the healthcare business. Classification, grouping, and association rules are examples of data mining approaches that are used to uncover frequent patterns in medical data to forecast disease. Classification algorithms are widely used in data mining for medical diagnosis and illness prediction. In this study, the Support Vector Machine (SVM) and Nave Bayes classification techniques are utilized to predict liver illness. Several liver illnesses need the clinical treatment of a physician.

3. EXISTING SYSTEM

Current medical systems are concentrated on collecting and processing all medical data, such as hospital management and decision-making systems. Each and every factor is taken into account, and the entire patient data is loaded. The medical data cannot be simply analyzed since, to generate a probabilistic rating, elements such as test findings, epidemics currently circulating, medical history, environmental circumstances, and several additional aspects that possibly or unlikely not be contained in the report are necessary. We previously created a GNN (Graph Neural Networks) model. A data mining technique was created based on this new measure to mine the causal association between medications and their associated adverse drug responses (ADRs). To rank the putative causal links between each of the identified medications, the exclusive causal-leverage method was used. The algorithm may successfully rank known ADRs higher than all other symptoms in the database.

4. PROPOSED SYSTEM

In the suggested work, a user would utilize a search engine to use symptoms as a query to gather information on diseases and treatments to diagnose illnesses. These symptoms are pre-processed to make it simpler to find the symptoms keyword in the subsequent step, which helps in the quick identification of the ailment. The symptoms associated with that keyword are compared with the stored medical input database to find the various illnesses associated with that keyword. When many diseases are found, it will perform pattern matching on the numerous diseases and calculate disease probability. The disease will next do a differential diagnosis to determine disease accuracy. The keyword, which is a pre-processed symptom, is compared to the illnesses contained in the local database to find the disease that corresponds to the symptoms provided by the user. To categorize diseases into subgroups, CNN (Convolutional Neural Networks) classification was used to scan a record library of more than 20000 disorders and even more symptoms. Searching in the new, smaller subgroup with the most similar symptoms receives higher priority, which reduces database access. In pattern recognition, the k-nearest neighbor algorithm is a technique for classifying objects based on the nearby training examples in the feature space. Instance-based neural learning techniques like CNN only estimate the function locally and save all computation for classification. The existing system has been found to work best with this feature.

4.1 Data Preprocessing Semantic Word Analysis

Data pre-processing, a popular first step in the data mining process, transforms the data into a format that can be handled more quickly and effectively for the user's needs. This step removes noisy and irrelevant data from the retrieved data. Initially, the string tokenize class is used to separate the sentence with a space. Then stop terms like a, an, the, is, was, and so on are eliminated. Unwanted terms, such as filler words, were deleted after human mistakes were eliminated. Then stemming is performed, which is the act of separating morphological and inflexional ending words from root words. Finally, semantic word extraction is carried out and saved in the local database. Medline database articles use the same pre-processing procedures, such as stemming and stop word removal.

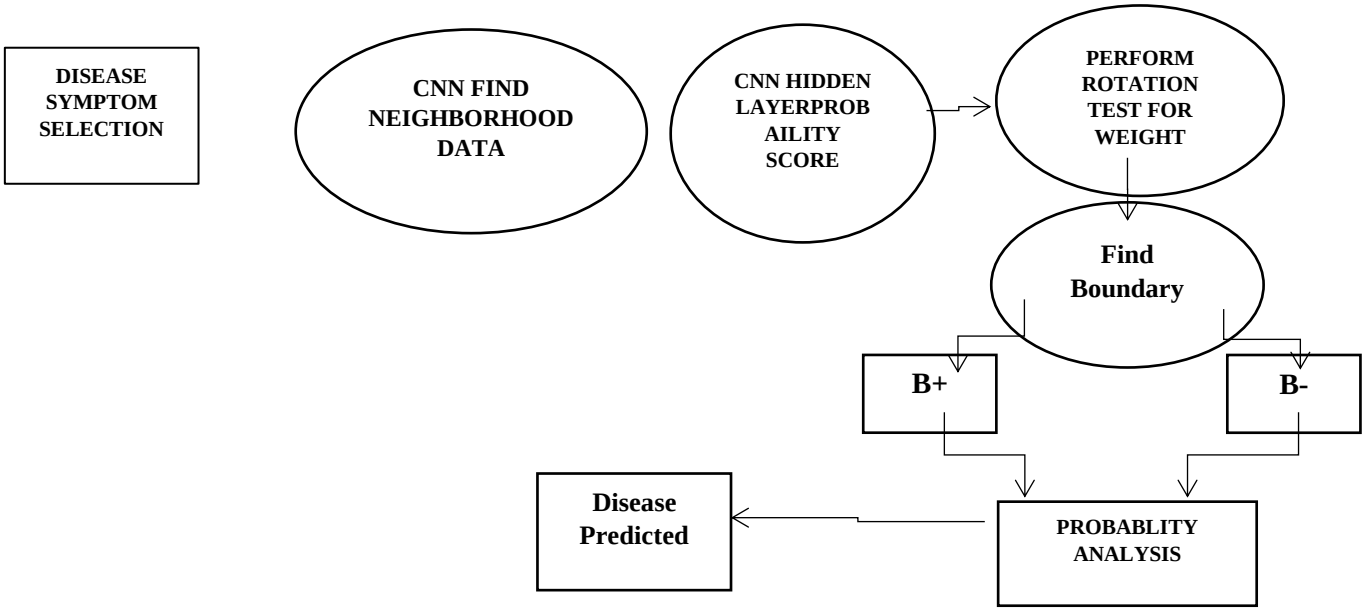
J Popul Ther Clin Pharmacol Vol 30(8):e262–e267; 19 April 2023.

4.2 Disease Identification

The keyword, which is a pre-processed symptom, is compared to the illnesses contained in the local database to find the disease that corresponds to the symptoms provided by the user. Since searching through a record database of more than 200 diseases and even more, symptoms are time-consuming, CNN classification was used to categorize disorders. Searching in the new, smaller subgroup with the most similar symptoms receives higher priority, which reduces database access. Weights have been assigned to specific disease symptoms, each disease as an individual, and the subclass disease group.

4.3 Weight Assigning

After collecting the closest matching symptoms, many diseases are identified. The required symptoms are then extracted from the final report. The derived correct symptom is then contrasted with the initially entered symptoms.



The LAMSTAR network is currently used to weigh this data. The word net network uses correlation links between individual neurons in different self-organizing map (SOM) modules to encode information. The complete input word is not used to create neurons in the word net. Instead, only individual sub-words(weights) are stored in SOM modules, and correlations between sub-word are noted in terms of establishing or modifying connections that link neurons in different SOM modules. Small changes in trend will lead to significant changes in the ratio as opposed to leaving negative weight in place. The variations should be minimal to keep the weight ratio as consistent and exact as feasible. As a result, only positive weights are taken into account.

4.4 Pattern Matching

By using prior knowledge or statistical data extrapolated from the patterns, pattern recognition aims to identify data (patterns).

Measurements or observations that identify points in multidimensional space are frequently grouped to form patterns that need to be classified. Iterative search for pattern matching makes use of previously stored data. The selection of the patient's symptoms serves as the first step in the algorithm. In turn, the algorithm produces a list of all likely diseases, ranked by the number of matching symptoms in the database. The list is produced once each symptom is entered. After the first iteration, the second iteration's list of symptoms will be

proposed based on the disease list created in the first iteration. Only symptoms of illnesses on the previous list will be included in the new list of symptoms. This technique is repeated repeatedly, with illnesses ranked according to their likelihood.

4.5 Differential Diagnosis

A Hopfield network is used by the system to do a differential diagnosis. They function as binary threshold units in content-addressable memory systems. However, they are not guaranteed to converge to one of the observed patterns; only they will converge to a local minimum. The value is dependent on whether the input to the unit reaches its threshold. They are also rigorous, but they suffer from the curse of dimensionality in terms of the number of categories they can efficiently manage, and they are susceptible to missing data sets.

5. CONCLUSION

Medical diagnosis is an essential field of study that aids in the identification of disease occurrence. Using the numerous strategies outlined, the system will uncover the underlying condition as well as a list of the most likely diseases with comparable symptoms. Because the database is a description database, tokenization, filtering, and stemming are employed to decrease the dataset size. The system's key advantage is that it can be used for any type of dataset, whether it is a description dataset or not. In this study, the CNN classifier is utilized to categorize the semantic connections between illnesses, and the keyword searching algorithm is used to collect relevant healthcare information for the related user symptoms. In the future, prediction accuracy will be improved by employing various machine learning algorithms, which will aid in medical diagnosis.

REFERENCES

1. SVM Classification Predictive Analysis for the Dengue Virus by Fathima, A.S., and Manimeglai, D. 521 to 527 in the International Journal of Engineering and Technology.
2. N.D.A. Tarmizi et al., 2018) two approaches to selecting features for a dengue epidemic detection model in Malaysia. Journals of Next Generation Information Technology (JNIT), pages 96-107.
3. Rajeswari, P., and G.S. Reena published Data Mining Algorithms for the Analysis of Liver Disorders in 2019. Global Journal of Computer Science and Technology, 10, 48-52.
4. Gulia and others 2014) Grouping of Liver Patients Utilizing a Keen Method. International Journal of Computer Science and Information Technology, pages 5110-5115 (IJCSIT).
5. Liver diseases can be predicted using SVM and the Naive Bayes Algorithm, according to S. Vijayarani and S. Dhayanand (2015). 816-820 in IJSETR, the International Journal of Science, Engineering, and Technology Research.