

MACHINE LEARNING AND NATURAL LANGUAGE PROCESSING BASED COGNITIVE DISEASE PREDICTION MODEL FOR PREDICTION OF PSYCHOLOGICAL DISEASES

Vikas Kamra

Assistant Professor, Ajay Kumar Garg Engineering College, Ghaziabad, U.P India
kamravikas@akgec.ac.in

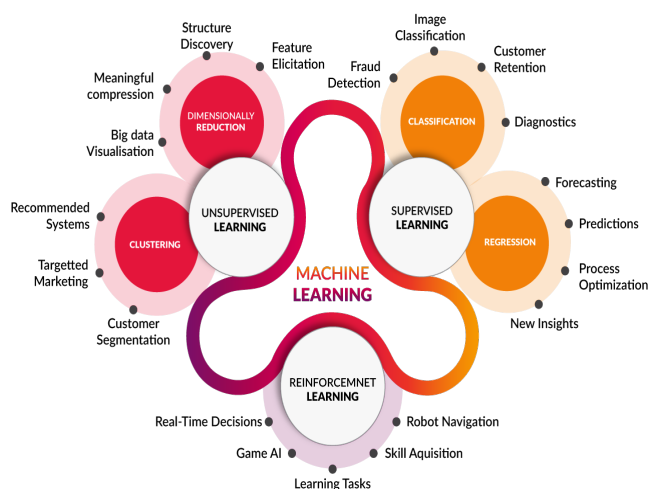
Abstract— Artificial Intelligence is advancing into domains that were previously thought to be the sole realm of domain analysts. Recent advancements in digital information processing, artificial learning, and computing resources have aided employees working in healthcare services across a variety of sectors. The constant growth of our country's population places a significant strain on our medical resources. As a result, efficient technology that improves the performance of our medical diagnosis system is critical. Both hospitals and patients will benefit from the technology. It lowers the chance of human error during patient registration in hospitals. This study presents a model for predicting cognitive pathology in psychological disorders such as bipolar disorder, obsessive-compulsive disorder, and schizophrenia. Providing cognitive health care in the mainstream medical business necessitates novel approaches to treating people with mental diseases.

Keywords— Machine Learning, Mental Illnesses, CDSS, Disease Recommendation System, Physiological Diseases

I. INTRODUCTION

Machine learning is a methodology based on AI that is able to develop a template that can benefit from and build on own interactions instantly. There is no provision for explicit scheduling. Associative memory and mathematical learning are the main idea behind computer practice. A deep learning methodology can be used to write an application that can learn and forecast data. For computer tasks where direct programming is complicated or not feasible, machine learning may be used [1]. The supervised learning trains the mathematical formalism with certain appropriate data sets and, accordingly, all practical inputs are correctly responded to by the proposed architectures. Regression and classifications are two types of controlled training. Unsupervised learning determines the probability of input data. It then categorizes

the data. Unsupervised learning is a stochastic illustration. Half-monitored learning uses unlabelled data that are used for template education. It consists of unsupervised and supervised training. The method of data mining is to find secret details from a huge volume. This process examines the available data from diverse sources [2]. It then turns the data into significant information. Then a reference set of data is prepared until any data mining technique is used. The set of data is also pre-processed. This mechanism can be represented in repositories as information discovery. In order to determine and generate information, data mining can be coupled by means of different hallucination tools, facts and natural language processing techniques [3].



Machine Learning Techniques [13]

Users can fully grasp weaknesses on the system using the aforementioned computer vision and data gathering techniques to create an effective medical diagnostic help program. Understanding from credible sources should

also be collected. The model suggested should be ideal for regional areas. Application consumers must receive adequate training. The integrated device should be tracked properly [4]. The program's foundation of knowledge should be routinely controlled. Together with artificial learning, conventional feature optimization approaches and profound learning techniques are two common strategies to supervised learning-based visual analyses. Conventional techniques of feature-engineering depend on experience and know how to handle certain functions in a certain context, on the essential discriminatory characteristics and methods. Machine based learning approaches combined with analytical interpretation provide strong resources for the analysis of objects. One benefit of these systems is that it is fairly easy to decipher, by linked to genetic information the characteristics of the classifier system [5]. Such conventional methods are still in use in various applications in multiple digital pathologies for years, such as cell identification and identification, human and mouse model bowel cancer graduation. Such traditional methods have also been used to explain histology features in quantitative terms. Nevertheless, earlier studies gave room for enhanced functional technology and performance identification [6].

A neural network comprises groups of neurons that are fundamentally related with each other. The relation to a single particle to many other cells is possible and there are a large number of connections in a network [7]. Connections known as synapses are usually formed from synapses to dendrites, although it is possible to produce dendrodendritic axons and other associations. Many types of signals emerge from neurotransmitter diffusion, including electrical signals [8]. Neural network is basically an associated group of naturally stirring neurons that use some conceptual or analytical model of knowledge treatment based on a connective approach to do calculations is called as artificial neural network [9]. An ANN is an evolutionary framework which modifies its architecture on the basis of explicit or implicit network data. Non-linear scientific data analysis and choice-making methods are more applicable for neural networks [10]. They could be used in modeling dynamic sources and outcomes associations, or in defining data trends [11].

Natural Language Processing is a branch of computer science and AI that works with computer-human interactions through speech recognition. The eventual aim of NLP is to translate, encode, comprehend, and render sense of human linguistic in a useful manner [12]. Neural networks are used in most NLP approaches to extract information from world speech. In data science, NLP is regarded as a challenging process. It is complex because of the essence of human expression. The rules governing the transmission of knowledge through colloquial expressions are difficult for programs to comprehend [13]. NLP is the process of using machines to define

and retrieve idiomatic phrases from disorganized data sets so that programs can interpret it. When the programme receives the text, it can use machines to retrieve the context associated with each word and gather the necessary data. Speech recognition is the potential of a computer or software to understand and translate enunciated syllables into written form [14]. It is a multidisciplinary concept that combines AI, computer programming and philology. Much modern equipment can provide speech recognition features to make using the computer simpler [15].

II. RELATED WORK

Several scientists have now investigated the complexity and diversity of accessible digital health data. An extensive analysis of models previously provided with a view to strengthening their shortcomings has been carried out. Here you'll find a short literary review relating to past work. In 2019, Angela and Yael stated that the reasoning of pattern recognition offers a comprehensive set of machine learning notions for sculpting human learning and taking decisions. In 2018 CDSS based on ANN was prepared by Hela LASSOUED and Raouf KETATA. For selecting ECG signals, the MIT-BIH database was used. The second major explanation for cancer deaths in females was prostate cancer, Dalal, Mohammad Ahmad and Zhang reported in 2018. You may classify it by means of a biopsy.

The predictive detection of Alzheimer's Disease [AD] was examined by Gunawardena, Rajapakse and Kodikara in 2017. The brain imaging results were used to diagnose Alzheimer's disease. SVM has a sweeping generalization failure that is usually less than that of the other classifications, so SVM is widely used to address object identification issues. They initially identified the most biased characteristics using the SVM criterion and then implemented the SVM-based classification system for diagnosing healthy subjects and individuals with AD using MRI neural scans. In this study, responsiveness of 93.7% and a precision of 88.5% have been obtained. In the year 2017, B. Che and Z. Jin and C. Jin and C. Liu find that EHRs clinical diagnostics records, doctor records, and medical records. In 2017, the main death causes is operator error in developing countries were explained by Jingzhi, Yang and Lida Xu.

The USA's health care system did not find it ideal to time, yet with a enormous funds for 2016, H. Yin and N. K. Jha noticed. They stated mobile clinical devices can collect up-to-date health-related information. Simple stochastic optimization algorithms are a group based on the Bayes theorem. The author suggested the Bayesian multiplied Kernelization technique that allows AD to be distinguished from NC, however, poor diagnostic results for MCI individuals and non-MCI issues. In 2016, Shuai Wang and Yang Cong stated computer-aided endoscopic diagnosis methods needs supervised machine

learning techniques. The endoscopic medical reports are used for diagnosis by the help of text and images, which are basically labeled data. In 2016, Hurouq Hijazi and Alex Page stated the use of personalized medicines and their functions and applications based on artificial intelligence and machine learning techniques.

In 2016 the benefits of machine learning algorithms are clarified by Zhu, Liu, Lu, Li, and the psychological evaluation can be made available to patients everywhere. The increased population of China was highly problematic in 2016 Yao, Yue and Jin explained. It puts more pressure on the country's health-care facilities. An alternative online diagnostic platform is therefore needed. Mansoor Baig, Gholam Hosseini and Maria Lindén developed a automatic decision-making support network in 2017 for diagnosis and treatment of core neurological actions based on Probabilistic reasoning simulation. In 2017, the software-aided percutaneous approaches need to supervise the machine's practice were stated by Shuai Wang and Yang Cong. Treatment of text and images based diseases that are essentially labeled information is carried out using endoscopic medical reports.

III. DATA MINING CLASSIFICATION TECHNIQUES

Different data mining classification techniques were used by various researchers to design an intelligent disease recommendation system [16]. These techniques include Bayesian Network, Decision Table, LAD Tree, and J48 Tree. Various performance parameters like, Mean Absolute Error, Root Mean Square Error and Relative Absolute Error were checked on Weka Data Mining Tool by using multiple data sets [17].

Results of various classification techniques for different set of errors are revealed beneath in Table 1.

Table 1: Errors Generated By Various Classification Techniques

Technique / Parameter	Mean Absolute Error	Root Mean Square Error	Relative Absolute Error
LAD Tree [7]	1.352	0.182	1.126
Bayesian Network [16]	0.981	0.176	1.105
J48 Tree [9]	1.258	0.180	0.985
Decision Table [23]	0.852	0.242	1.652

Table 1 shows that Decision Table Classifier has lowly mean absolute error of 0.852; Bayesian Network Classifier have lowly root mean square error of 0.176; and J48 Tree Classifier has lowest relative absolute error of 0.985.

IV. PSYCHOLOGICAL DISEASES

The term psychological illness is often used to refer to mental disorders which are better recognized. Mental conditions are characterized by a pattern of psychopathological symptoms that affect various aspects of one's life [18]. The person who is having these manifestations is distressed by these illnesses. Brain disorder not only lowers a person's quality of living, but it also leads to higher health-care costs. In reality, psychological state is currently the most costly aspect of our healthcare sector, surpassing heart disease, which was previously the far more costly [19]. Many people who suffer from depression symptoms do not access prompt psychiatric help due to the high costs of care. A lack of psychiatrists and the prevalence of mental disorders are two other factors that contribute to the high cost of therapy [20].

Artificial intelligence is now disrupting a variety of fields, including mental wellbeing. Machine learning and AI are creating a new form of treatment that focuses on delivering individualized support and encouragement. Despite the fact that AI for psychological health still faces numerous challenges, evidence suggests that behavioral health approaches benefit from consistency, and innovation appears to be providing a better user interface. Better cognitive wellbeing is now within reach [21].

Obsessive Compulsive Disorder (OCD) is a psychological illness that triggers intense emotions or thoughts (obsessions) or the need to replicate a behavior (compulsions). Obsessive thoughts and compulsive strains will coexist in certain people. Bipolar disorder is a chronic condition characterized by mood swings that are severe. Mania, or an excessively high mood, is one of the symptoms. These can however contain depressive episodes. It is also known as panic disorder or bipolar syndrome. Individuals with bipolar disorder can struggle to manage daily tasks at work, as well as sustain relations [22]. While there is no solution, there are many recovery services that will help relieve the symptoms. Schizophrenia is a chronic neurological illness in which patients have altered perceptions of reality. Schizophrenia may trigger hallucinations, paranoia, and highly dysfunctional thought and behavior, which may make it difficult to work in a regular basis. Schizophrenia patients are to be treated for the rest of their lives [23].

V. PROBLEM FORMULATION AND PROPOSED SYSTEM METHODOLOGY

A guideline template for illness treatments using data mining is created to help users get AI guided diagnosis before admission centered on valuable data produced by the online physician's activities. The system is suggested for the treatment of artificial intelligence-led diseases. Firstly, annotate the illness derived identified by users using a computational disease survey by defining the entity. Within this diagnosis model of the AI based diagnosis of diseases, we turn the abstraction from

the diagnostic disease into a classification problem, using a fairly hot trend in deep study research to overcome this classification concern. Smart health care systems face many obstacles. These include the confidentiality of information used in these programs, the disparity between pilot programs designed by study and implementation of initiatives, barriers to contact between medical staff and data academics, increased expenses to transition from limited to lengthy-term systems, etc. To artificially diagnosis the disease, a step-by-step procedure is defined in following paragraph:

We give instructions to the designed framework initially. The age, gender and other user questions of the patients are incorporated into the system. Using word segmentation processing, the word embedding vector technique is used to process the input data. After that, the list matrix is initialized and the cursor's initial value is set to zero. We execute a series of statements by using the control structure, like during for loop. We set the term of the i-th letter as the system's current phrase. By setting up the cursor to the i-th letter, we double the dimensions of the current word embedded vector. After one effective iteration of the control scheme, the counter is increased by one. Together with the gender and age of the users, the new created matrix is added to the CNN Network as the parameters for predicting the results from the developed framework. We're finally submitting the results to the device. We are thus developing a medical diagnosis support system based on artificial intelligence.

The General Disease Dataset, which was retrieved from the Kaggle Repository, was utilized in this study to assess diseases as well as to test and train the suggested disease diagnosis assistance system. A database of 4371 patients was created using similar knowledge extraction and data preparation procedures. The entire dataset is divided into two pieces. 75 percent of the data pieces in the provided dataset teach the suggested machine learning model. To test the system, the entire software is checked and 25% of the data items are left behind.

These are the steps which are used to design an algorithm for machine learning based medical diagnosis support system. Data collection from pharmacists or clinics can be used for assembling data sets. Statistics on illnesses and the effects of these disorders can be accessed from numerous reputable web pages. Configured data-set is split into two parts, one designed to train the projected system and the other to evaluate the established model in real-time.

VI. RESULT ANALYSIS

The Proposed model, which is an up-gradation over recurrent neural network, gives better accuracy as compared to Convolution Neural Network, Hopfield Network, and Recurrent Neural Network. It is implemented on Python 3.7. Fastai library of Python is used to train and test various available

neural networks and the proposed model. Matplotlib library is used to plot the graphs for learning rates of each network model. The dataset is divided into 24 batches of equal sizes. A pickle file has been generated after training and testing of datasets. AWD_LSTM language model with 70 hidden layers has been used to train the proposed model. 30% neurons kept off to stop over fitting and 5 epochs has been generated to measure performance of all models. The accuracy comparison of Proposed Model with other Neural Network Techniques is shown below in Table 2:

Table 2: Accuracy Comparison of Proposed Model with other Neural Network techniques

Model Implemented	Accuracy Achieved (%age)
Convolution Neural Network	91.73 %
Recurrent Neural Network	95.48 %
Hopfield Network	87.65 %
Proposed - CDPM	96.51 %

It is shown here that proposed model has achieved highest accuracy among all types of Neural Networks. Convolution Neural Network (CNN) has achieved 91.73% accuracy. Recurrent Neural Network (RNN) has achieved 95.48% accuracy. Hopfield Network has achieved 87.65% accuracy.

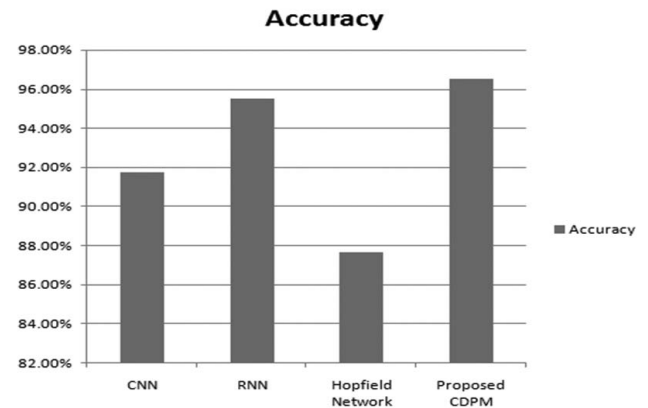


Fig. 2. Accuracy comparison of Proposed - CDPM with other Neural Network Techniques

Proposed model has achieved has achieved 96.51% accuracy. A confusion matrix has been created to calculate the performance of each model and to do the comparison of accuracy among all types of models.

VII. CONCLUSION

The indications and repercussions of mental illness might vary depending on the diagnosis, conditions, and a variety of other factors. Mental chronic disorders can affect one's thoughts, feelings, and actions. Mental diseases are thought to be influenced by genetics, chemical factors before birth,

and cognitive function. During this investigation, a thorough literature review was conducted. In the approaching years, Smart Healthcare Applications based on the designed systems will become increasingly prevalent. By delivering improved medical facilities to patients, the developed system would improve the future of the healthcare sector. It will mostly benefit people who live in distant places with limited access to doctors and medical facilities. The system designed will improve the world of healthcare by providing greater facilities for clinicians. People who live in distant areas with limited access to medicine and medical facilities will profit from this. This will also help to lessen workloads by providing smart healthcare testing services to medical staff, physicians, and various hospitals during the initial phase of enquiries.

REFERENCES

- [1] D. Malathi, R. Logesh, V. Subramaniaswamy, V. Vijayakumar, A. K. Sangaiah, "Hybrid reasoning-based privacy-aware disease prediction support system" in *Computers & Electrical Engineering*, vol. 1, pp. 114-127, 2019.
- [2] S. Yang, R. Wei, J. Guo and L. Xu, "Semantic Inference on Clinical Documents: Combining Machine Learning Algorithms With an Inference Engine for Effective Clinical Diagnosis and Treatment" in *IEEE Access*, vol. 5, pp. 3529-3546, 2017.
- [3] A. Radulescu, and Y. Niv, "State representation in mental illness", *Current opinion in neurobiology* in vol. 55, pp. 160–166, 2019.
- [4] J. H. Eom, S. C. Kim and B. T. Zhang, "AptaCDSS-E: A classifier ensemble-based clinical decision support system for cardiovascular disease level prediction" in *Expert Systems with Applications*, vol. 34, pp. 2465-2479, 2008.
- [5] M. Sun, T. Min, T. Zang and Y. Wang, "CDL4CDRP: A Collaborative Deep Learning Approach for Clinical Decision and Risk Prediction" in *Processes*, vol. 7(265), no. 5, pp. 1-22, 2019.
- [6] A. Awaysheh, J. Wilcke, F. Elvinger, L. Rees, W. Fan and K. L. Zimmerman, "Review of medical decision support and machine-learning methods" in *Veterinary pathology*, vol. 4, pp.512-525, 2019.
- [7] A. Umer et al., "A Decision Support System for Diagnostics and Treatment Planning in Traumatic Brain Injury," in *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 3, pp. 1261-1268, May 2019.
- [8] D. W. Bates et al., "Ten commandments for effective clinical decision support: making the practice of evidence-based medicine a reality" in *Journal of the American Medical Informatics Association*, vol. 10, no. 6, pp. 523-30, 2003.
- [9] M. P. Brown et al., "Knowledge-based analysis of microarray gene expression data by using support vector machines" in *Proceedings of the National Academy of Sciences*, vol. 97, no. 1 pp. 262-267, 2000.
- [10] K. T. Chui et al., "Disease diagnosis in smart healthcare: Innovation, technologies and applications" in *Sustainability*, vol. 9, no. 12, pp. 1-23, 2017.
- [11] R. Kaushal, K. G. Shojania, and D. W. Bates, "Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review" in *Archives of Internal Medicine*, vol. 163, no. 12, pp. 1409–1416, 2003.
- [12] K. Kawamoto, C. A. Houlihan, E. A. Balas and D. F. Lobach, "Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success" in *BMJ*, vol. 330(7494), no. 765, pp. 1-8, March 2005.
- [13] A. X. Garg et al., "Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review", *JAMA*, vol. 293, no. 10, pp. 1223–1238, 2005.
- [14] E. Caballero-Ruiz et al., "A web-based clinical decision support system for gestational diabetes: Automatic diet prescription and detection of insulin needs" in *International journal of medical informatics*, vol. 102, pp. 35–49, 2017.
- [15] M. Mandzuka et al., "Mobile Clinical Decision Support System for Acid-base Balance Diagnosis and Treatment Recommendation" in *Acta Informatica Medica*, vol. 25, no. 2, pp. 121–125, 2017.
- [16] P. Sajda, "Machine learning for detection and diagnosis of disease", *Annual review of biomedical engineering*, vol. 8, pp. 537–565, August 2006.
- [17] M. M. Baig, H. G. Hosseini and M. Lindén, "Machine learning-based clinical decision support system for early diagnosis from real-time physiological data" in *proceedings of IEEE Region 10 Conference*, Singapore, pp. 2943-2946, November 2016, doi: 10.1109/TENCON.2016.7848584.
- [18] H. Yan, Y. Jiang, J. Zheng, C. Peng, and Q. Li, "A multilayer perceptron-based medical decision support system for heart disease diagnosis" in *Expert Systems with Applications*, vol. 30, no. 2, pp.272-281, 2006.
- [19] H. Yin and N. K. Jha, "A Health Decision Support System for Disease Diagnosis Based on Wearable Medical Sensors and Machine Learning Ensembles," in *IEEE Transactions on Multi-Scale Computing Systems*, vol. 3, no. 4, pp. 228-241, December 2017, doi: 10.1109/TMSCS.2017.2710194.
- [20] S. Fong, S. Akashe, P. N. Mahalle, "Information and Communication Technology for Competitive Strategies" in *Proceedings of Springer Third International Conference on ICTCS*, India, pp. 21-29, 2017.
- [21] K. Polat and S. Güneş, "Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine" in *Applied Mathematics and Computation*, vol. 186, no. 1, pp.898-906, 2007.
- [22] M. Sun, F. Wang, T. Min, T. Zang and Y. Wang, "Prediction for high risk clinical symptoms of epilepsy based on deep learning algorithm" in *IEEE Access*, vol. 6, pp.77596-77605, 2018.
- [23] A. D. Sunny, S. Kulshreshtha, S. Singh, B. M. Srinabh and H. Sarojadevi, "Disease diagnosis system by exploring machine learning algorithms" in *International Journal of Innovations in Engineering and Technology*, vol. 10, no. 2, pp. 14-21, 2008.
- [24] X. Jiang, A. Wells, A. Brufsky and R. Neapolitan, "A clinical decision support system learned from data to personalize treatment recommendations towards preventing breast cancer metastasis" in *PloS one*, vol. 14, no. 3, pp. 213-292, 2019.
- [25] Jasdeep Singh Bhalla and Anmol Aggarwal, "A Novel Method for Medical Disease Diagnosis Using Artificial Neural Networks Based on Backpropagation Algorithm" *IEEE Xplore*, June - 2014, pp. 55-61.

ABOUT THE AUTHOR



Vikas Kamra received his Bachelor's and Master's Degree in Computer Science and Engineering. Presently he is pursuing his doctoral research in the field of machine learning and healthcare informatics. He is working as an Assistant Professor in Computer Science & Engineering Department at Ajay Kumar Garg Engineering College, Ghaziabad. His primary research area is machine learning and data mining.