

Advanced method for the Analyses of Large Amounts of Data using deep learning in Health Sector

Dr. Vidya Gavekar,

Associate professor, Suryadatta group of Institute, vidyagavekar@ gmail.com

Dr. Deepti P Lele,

Associate Professor , Suryadatta group of Institute,deepticyber@gmail.com

Dr.Manisha Kumbhar.HoD-MCA,

Suryadatta group of Institute,Manisha.kumbhar@suryadatta.edu.in

Received 2022 March 15; **Revised** 2022 April 20; **Accepted** 2022 May 10.

Abstract

Large amounts of data, rising costs, and a focus on personalised care have all led to a rise in the use of "big data" in healthcare over the past few years. "Big data processing" is a term used in healthcare to describe the development, collection, analysis, and storage of clinical data that is too big or complicated to be figured out with typical data processing methods. Big data sources for healthcare include the Internet of Things (IoT), Electronic Medical Records/Electronic Health Records (EMR/EHR), which include a patient's medical history, diagnosis, treatment plans, allergies, and lab and test results, and genomic sequencing. In this article, a wide range of healthcare data was analysed using machine learning. As well as the fact that it is hard to gather, handle, and analyse a lot of data. In this paper, we'll take a fresh look at how machine learning algorithms and the need to analyse and use huge amounts of data are linked in real life. Deep learning can be used to look at a lot of data in the health field.

Keywords : Deep Learning ,Big Data , Health Sector , machine learning algorithms , Machine Learning.

Introduction

A growing number of elderly people with incurable, long-term conditions like dementia puts a constant strain on medical services. This is because mental impairment and a high number of other health problems increase the risk of hospitalisation. When people who are at risk are found, preventive steps can be taken to ease the strain. Electronic health records give big data analysis a chance to help with these kinds of applications. But because the data is sparse and has a lot of dimensions, traditional statistics and machine learning have a hard time dealing with it. This study comes up with a new way to predict whether or not a person with dementia will need to be admitted to the hospital. Using a new way to adapt snapshot ensembles to use a dynamically generated learning rate schedule, as well as a new way to adapt entropy weight regularisation for use with neural networks (NNs), and then a new way to evaluate model parameters, we were able to find 10 medical events that were very likely to predict that a person with dementia would end up in the hospital in the future. These events, which included diagnostics, prescriptions, and procedures, could model and predict future hospitalizations just as well as the full dataset, and in some cases even better. ECNN is much better than statistical feature selection methods and ML-based modelling techniques when it comes to how well they can predict the future. By finding out about these medical events, it might be possible to set up early warning systems to find people with dementia who are at high risk of being hospitalised or put in a home. With multiple signs that nutritional health is a major risk factor for hospitalisation, this information can be used to learn more about how to prevent hospitalizations by putting more focus on better nutritional care for dementia patients. Such examples show how there are many ways to focus on preventing and avoiding hospitalisation by changing the way secondary care is given. Overall, contributions like those listed allow for a possible decrease in the use of critical health care, which is a good thing. They also reduce risk in a statistically older and more vulnerable population by limiting their exposure to hospital-caused risks like infections. There are many ways that ECNN could be improved in the future. Most important

would be the addition of modelling methods based on time series that can take into account how a patient's health changes over time. Other ways to improve include more statistical analysis of ranked features to improve ranking, larger datasets that go beyond the Wales region currently covered by secure anonymised information linkage (SAIL), and the use of state-of-the-art deep learning modelling, such as event code representation using word embedding techniques, to improve prediction performance as a whole. Newer modelling methods, like evolutionary algorithms, can be used to compare the accuracy of predictions and reduce the number of features at the same time. This opens up the possibility of more applications. The list of medical events shows well-known risk factors for hospitalisation, which shows that the list is useful, while new events offer the chance to use traditional clinical analysis to find more risk factors and indicators. As a result, MCNN could be used in the future to find risk factors in other areas of medical informatics. When combined with MCNN, the general nature of patient medical records makes it possible to use them in other areas to provide small-scale, interpretable indicators that make it easy to find people who are at risk and give them care ahead of time.

Related work

A number of studies that have been carried out in recent years have looked into the most common reasons that patients with dementia are admitted to the hospital. These causes have been investigated as part of a number of different research. These studies have mostly relied on information gleaned from clinical investigations and questionnaires; as a result, they have only covered a very small portion of the population.

K. R. Swetha, et al[1] An x-ray of the chest is vital in determining if a guy has pneumonia, and an expert in prediction is also necessary. As a result, using big data and deep learning to develop an automated predictor for pneumonia is more viable. In this forecast, CNN (Convolutional Neural Networks) and other classifiers stand out above the others. Pre-training CNN models with very large datasets from healthcare institutions, frequently referred to as "big data," has a high probability of achieving accurate classification. Preliminary training of a CNN model, which uses an effective feature extraction method and a variety of classifiers to distinguish between positive and negative outcomes, is generally accepted to yield very accurate findings. This research paper uses Big Data, Deep Learning, and Machine Learning to predict Pneumonia.

Hariprasath Manoharan; et al.[2] Researchers compared their own health monitoring simulations to those of other researchers employing artificial neural networks. Simulated findings consider route loss, cost, energy, and longevity. It was examined with four deep learning methodologies, including SAE, LTS, K-next neighbour, and neural network methods, and SDAE showed particularly efficient sensor integration for monitoring human health when numerous sensor nodes are implanted.

Y. Attiga et al.,[3] The lift statistic was used to evaluate thyroid disease-prone people. DNN outperformed Random Forest by 36.63 percent in lift. Deep learning may be used to pick candidates for early intervention for improved health outcomes using a huge dataset with minimum demographic factors, similar to those maintained by healthcare marketing arms.

S. Chi et al[4] This paper suggests semisupervised multitask learning (SSMTL) for survival analysis with or without competing hazards. SSMTL converts survival analysis to multitask learning. This topic involves semisupervised learning and multipoint survival prediction. A simple modelling approach was used to analyse the distribution of survival times and the link between variables and outcomes. Semisupervised loss and ranking loss can be used with censored data and a non-rising survival probability trend.

K. Oyama et al[5] The mean absolute and mean absolute percentage errors in predictions made using TNIRS and blood data were lower. A mix of blood counts, electrolytes, and nutrition is required for therapeutic use of blood test data.

T. Karatekin et al[6] the risk factors for severe prematurity retinopathy were explored in a form of generalised additive model (GAM) with pairwise interaction terms utilising statistical analysis and logistic regression (GA2M). Accuracy and interpretation are often compromised when using machine learning algorithms on healthcare data. Our investigation has validated the clinical significance of a few risk factors, such as gender, that were previously assumed to be clinically inconsequential.

R. D. Sah, et al[7] Data mining helps predict disease symptoms. Heart disease, breast cancer, and diabetes are studied. Data management techniques forecast cardiac disease and analyse diabetic patients. Diabetes and its symptoms are becoming more well-known. This is like IT's role in healthcare. It aids in diagnosing and treating disorders. Using data analytics to predict sickness

B. Shickel, et al[8] Current research on deep learning for clinical tasks based on EHR data includes information extraction, representation learning, outcome prediction, phenotyping, and deidentification. Existing research has significant drawbacks, including model interpretability, data heterogeneity, and lack of uniform benchmarks. We conclude by summarising the field and recommending future EHR research.

Proposed methodology

It is a term used to describe a process of analysing vast amounts of information using more sophisticated approaches. From terabytes to zettabytes, these data collections come from a number of sources and might be structured, semi-structured, or unstructured. "Big data analytics" refers to the integration and analysis of a significant amount of complicated, heterogeneous data in the healthcare environment. For example, "omics" (genetics, epigenetics) and biomedical data, as well as HER data[9] (translational medicine) and metabolic data are all examples of this type of data (Electronic Health Record). Among these, electronic health records (EHR) play an important role and are currently being used in many nations. It is the major objective of the HER to extract relevant big data insights from the healthcare system.

Keeping a patient's medical history in the form of a digital or electronic database is known as an electronic health record (EHR). A patient's medical history, including diagnoses, prescriptions, test results, allergic responses, vaccines, and treatment plans, is stored in an electronic health record (EHR). All of the medical professionals who are involved in a patient's care can access the patient's electronic health records and use them to assist in the formulation of treatment recommendations. Some people refer to EHR as EMR (Electronic Medical Record). The usage of the internet, social networks, mobile transactions, online shopping, and many other activities contribute to the development and accumulation of tens of terabytes of data per second. We can now consider alternative viewpoints because of the growth of structured and unstructured data in the big data paradigm. This was made possible thanks to the abundance of information[10]. To better understand an individual's behaviour and motives, these new sources of information enable for the detection of immediate signals and triggers that indicate a person's interest in a certain offer or product. Large and diverse volumes of data can be mined for valuable information that can be used to improve the quality of the user experience in a meaningful way after it has been discovered and exploited[11]. Free, open-source, and paid EHR systems all have a substantial impact on the progress of any medical institution's medical records system. Researchers are being compelled by the advent of big data to think more broadly about the future and to focus on projects that do so. Medical practitioners[12] can benefit from the use of a clinical decision support system, or CDSS, which employs data analysis to assist them in making better decisions for patients' health. A clinical decision support system (CDSS) aims to provide clinical recommendations based on a variety of patient-related information. It is now able to integrate processes[13], provide assistance to patients in their care, and give recommendations for treatment plans thanks to clinical decision support systems. Patients are better served by a CDSS, which helps clinicians establish diagnoses and enhance patient care by minimising the need for unnecessary testing, boosting patient safety, preventing repercussions that could be both dangerous[14] and costly. A variety of medical applications for big data exist, including the reduction of medical treatment costs, the elimination of disease-related risk factors, the forecasting of disease, the enhancement of preventive care, and the analysis of drug efficacy[15]. With regard to the healthcare industry, the following issues need to be addressed: How can we tell which treatment is going to be the most effective for a certain illness? What effect do specific policies have on spending and behaviour? How much do you expect medical care prices to rise over the next few years? If the claim was made fraudulently, how can it be determined?, Is there a regional variation in the cost of healthcare? The best way to overcome these issues is to use a variety of methodologies and technology for big data analysis. High-quality medical care is supported by four fundamental pillars. For instance, real-time patient monitoring, patient-centered care, improved treatment techniques, and disease-prediction analytics are all examples. All four pillars of high-quality healthcare can be effectively managed through the use of descriptive, predictive, and prescriptive approaches of big data analysis[16]. Currently, a wide range of applications are making use of Deep Neural Networks, which are widely regarded as the leading edge of machine learning and big data analytics. Human-computer interaction

and question-answering systems, as well as defence and surveillance, are all examples of these uses. There are three main categories of DNN architecture[17], each of which can be subdivided into subcategories. Each of the three types of neural networks is referred to as a feed-forward neural network, a convolutional neural network, or an RNN. As a result of deep learning in healthcare, clinicians are better able to diagnose and treat illnesses, and this leads to better medical decision-making overall. Hospital management information systems can benefit from the use of deep learning technology in order to reduce expenses and lengthen hospital stays while also preventing fraud, identifying shifts in illness trends, and improving the quality of medical care. Examples of applications based on different types of biomedical data, such as pictures and time signals derived from lab findings and genomics or from wearable devices, are given in the following paragraphs, which are organised by type of biomedical data.

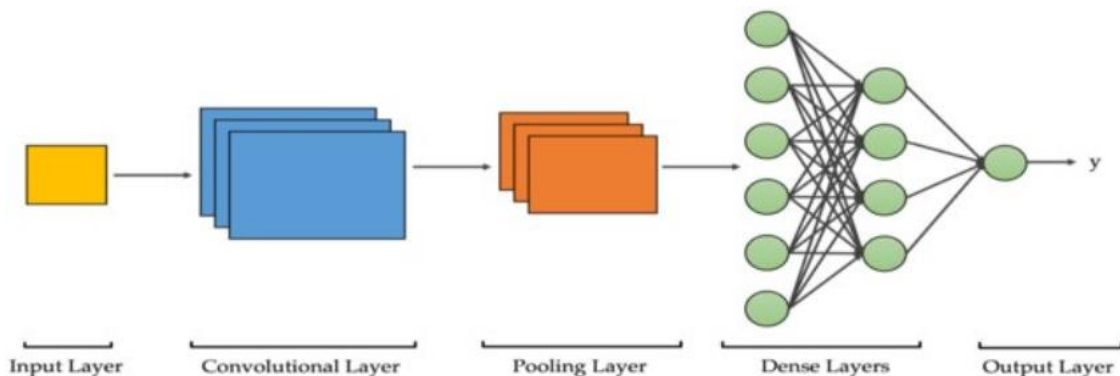


Figure 1: Convolutional neural networks can be used to categorise images in a simple way. An ANN is a type of artificial neural network (CNNs) ConvNet (LeCun, 1989) and CNN (LeCun, 1989) both have at least one convolution layer that prepares the system for a three-dimensional input tensor. A convolutional layer employs several cascaded convolution kernels to extract fine details from the input.

The CNN's primary architecture[18] is based on the deep neural network (NN), which is widely used in a variety of academic subjects and research domains. After that, we'll go over the advantages of neural networks (NNs) in comparison to more traditional statistical and machine learning approaches in health informatics, which are now in extensive use. The following steps should be followed when dealing with an input space known as R^{np} , which has n samples and p features: Neuronal networks (NNs) are built using multiple layers of perceptrons, each of which combines input data using a weighted and biased total. There are two parameters that can be trained to transform each input feature vector into an activation result within the first layer of the perceptron: weight and bias. The weight and bias parameters are shown in the following equation: where the activation function is used to transform each feature vector into an activation result within the first layer. These parameters can be trained in order to transform any given input feature vector X_j , for example. This layer's perceptrons receive its input from an activation vector represented by a l . As the name suggests, a deep neural network (NN) is an architecture that allows for deeper embeddings of the original data space and for non-linear transformations between the input and output layers. Between the two layers, there are a number of preliminary embedding layers.

$$a_k^{l=1} = \sigma \left(\sum_{j=1}^p w_j^l a_j^{l-1} + b_j^l \right),$$

$$a_j^0 = X_j$$

There are a total of 50 and 30, respectively, perceptrons present in each of the ECNN's two hidden layers inside the architecture of the network as a whole. In order to discover the optimum number of perceptrons to include in the model, an approach known as simple grid search hyper-parameter optimization[19] was utilised. The cross-entropy cost function is minimised by using the $f(w)$ weighted function, where y is the model probability output and y is the classification target. This method is based on equation (1), forward and back-propagation methods to update model

parameters, and minimises the cross-entropy cost function. In order to accomplish this regularisation, we will make use of the entropy weight regularisation function that we covered earlier.

$$\min_{\hat{y}}[-y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})] + \lambda f(w)$$

Results Analysis

SPSS was used to conduct the analysis on a PC with a 64-bit Core i5 processor that was running Windows 10 professional and had 12 gigabytes of free RAM. We are able to have a better understanding of the proposed strategy now that we have the data from the Inspection Survey. The context data includes measurements of fitness as well as certain essential traits. There were a total of 10,806 residents that participated in the real-time health assessments[20]. It is possible to develop health-related variables that reduce numerical complexity and successfully reduce the number of features by making use of the multivariate regression feature that is available in SPSS. This can be done without compromising critical values. Researchers examined numerous characteristics of obesity, high blood pressure, and diabetes in a model through a series of experiments that focused on efficiency. Combining features and calculating the coefficient of correspondence for each one allows for the possibility of achieving a correlation coefficient for the health-related indicators that is equal to or greater than 0.5. As a consequence of this, the factors that are the most essential when estimating obesity, high blood pressure, and diabetes are selected first. Both positive and negative correlations have been seen between the various parameters that have been chosen. Keeping track of the good effects that are brought on by the factors that are positively connected is a fantastic method to make your everyday life better. Maintaining a watchful check on parameter values that are negative can lead to the discovery of some anomalies. The CNN approach is utilised in order to swiftly differentiate between the two. In addition, the activity of the regular factors is analysed in order to categorise any external variables that may be present within the related factors. Diabetes, increased blood pressure, and obesity are three of the factors that are associated with the parameter. This determines which qualities, when compared to one another, are considered normal or aberrant. A significant improvement in health would result from an increase in awareness of the factors that contribute to conditions such as obesity, high blood pressure, and diabetes. CNN was developed with the help of an algorithm that was based on identification, reference, and regular behaviour patterns. This algorithm was developed using three different disorders. We supplemented the comparative research utility of our model by using regular correlation algorithms on a total of medical literature. In addition to this, a comparison of the model's performance is included. According to the results of the MCNNmodel testing, the diagnostic precision and reference values of our model are, in descending order, 85.32 percent, 85.61 percent, 92.32 percent, 81.34 percent, and 96.24 percent, respectively. The simulation will be executed with the usage of Python. These findings demonstrated that it was useful to delete only irrelevant information from the data that was received after the pre-processing step was completed.

Accuracy

According to this part, the overall accuracy of the MCNN classifier is evaluated. Coefficient correlation analysis is used to evaluate the proposed intrusion detection system's performance. The precision of the system can be estimated using

$$\text{Accuracy} = \frac{\text{Total Correctly Classified Samples}}{\text{Total Samples Available}} \times 100$$

the following formula.

The effectiveness of both the recently constructed system and the underlying strategy is broken down in great depth, as can be seen in Table 1. For this experiment, we will be making use of the same collection of data as we did for the one before it. The accuracy of the system as it is now implemented demonstrates data classification and normalisation. According to the findings, the proposed classifier performs better than the base CNN when it comes to learning for proposed data quality enhancement and improvement classification of the system. The proposed classifier in this graph is able to attain a higher accuracy rate while dealing with a significant amount of data points that have already been successfully categorised.

Figure 1: Tabular form of Accuracy

Number of Experiments	Proposed method	Base method
1	95.22	88.35
2	96.32	93.65
3	98.66	94.45
4	97.53	88.32
5	95.11	85.78
6	96.61	89.32

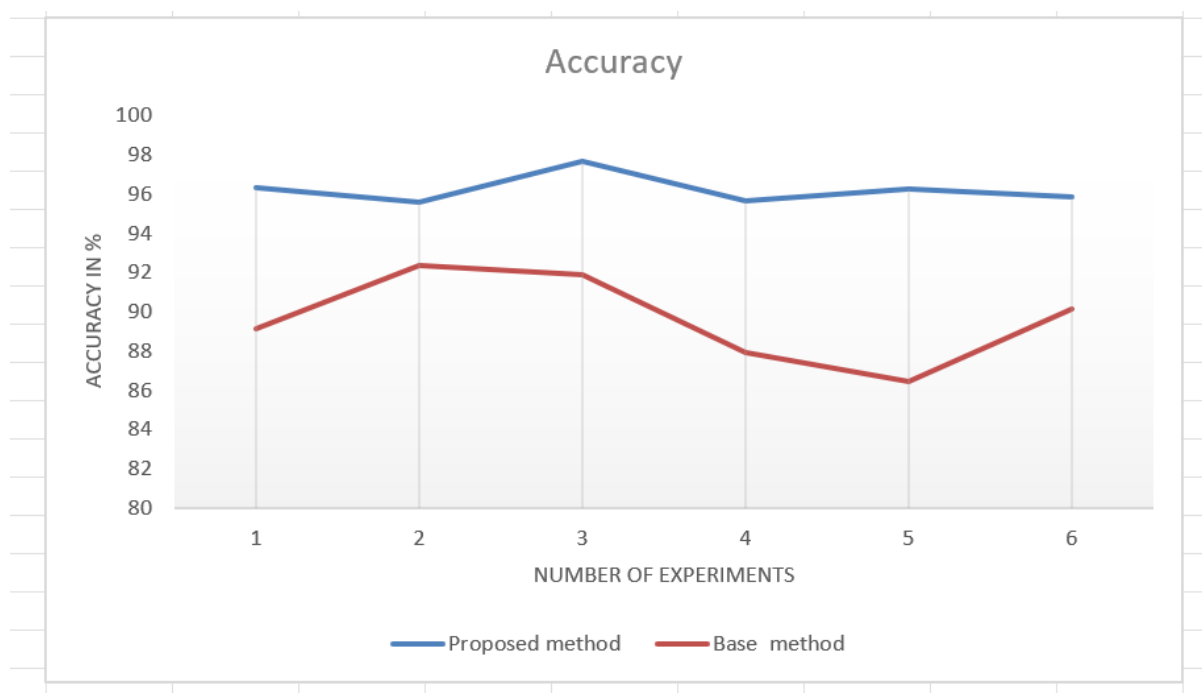


Figure 2 comparative Analysis in term of accuracy

The majority of the time, the model represents the potential accuracy of more traditional approaches to instructing students. Data mining and the evaluation of medical information are essential to the process of conducting research and making diagnoses in the medical field . Using the information that we acquired, we developed a model that was capable of demonstrating findings for obesity, hypertension, and diabetes in an easy and dependable manner.

Conclusion and Future work

It is becoming increasingly difficult to analyse medical big data because of its complexity and variation in format, which makes it more difficult to process. In terms of data analysis, deep learning is unquestionably the most powerful method currently accessible. For the two basic categories of medical data — medical picture data and medical text data

— this study builds the corresponding deep learning model independently. Afterwards, accurate early diagnosis and risk prediction for particular diseases are intelligently identified and achieved.. To begin, a deep learning model for AutoEncoder has been developed. Because the network can be trained ahead of time, this model can save time and resources by reducing computations. There are no technical challenges in applying the method to other kinds of medical image data, which is critical for improving the accuracy with which diseases may be diagnosed. Using 3D convolutional neural networks with spatial pyramid pooling, a deep learning model is created. Model's spatial pyramid pooling structure can process input data of any length, and model's 3D convolution structure can preserve the timing characteristics of various data while extracting the internal properties of the data. This allows for smart identification and control of disease risk through data analysis for the benefit of the patient's future.

Reference

1. K. R. Swetha, N. M. A. M. P and M. Y. M, "Prediction of Pneumonia Using Big Data, Deep Learning and Machine Learning Techniques," 2021 6th International Conference on Communication and Electronics Systems (ICCES), 2021, pp. 1697-1700, doi: 10.1109/ICCES51350.2021.9489188.
2. Hariprasath Manoharan; Ganesan Sivarajan; Subramanian Srikrishna, "Deep Learning Analysis of Location Sensor Data for Human-Activity Recognition," in Sensor Data Analysis and Management: The Role of Deep Learning , IEEE, 2021, pp.103-116, doi: 10.1002/9781119682806.ch6.
3. Y. Attiga et al., "Applying Deep Learning to Public Health: Using Unbalanced Demographic Data to Predict Thyroid Disorder," 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 851-858, doi: 10.1109/IEMCON.2018.8614888.
4. S. Chi, Y. Tian, F. Wang, Y. Wang, M. Chen and J. Li, "Deep Semisupervised Multitask Learning Model and Its Interpretability for Survival Analysis," in IEEE Journal of Biomedical and Health Informatics, vol. 25, no. 8, pp. 3185-3196, Aug. 2021, doi: 10.1109/JBHI.2021.3064696.
5. K. Oyama and K. Sakatani, "Deep Learning-based Prediction of Cognitive Function Using Basic Blood Test Data and NIRS-measured Cerebral Hemodynamics," 2021 IEEE International Conference on Digital Health (ICDH), 2021, pp. 218-219, doi: 10.1109/ICDH52753.2021.00040.
6. T. Karatekin et al., "Interpretable Machine Learning in Healthcare through Generalized Additive Model with Pairwise Interactions (GA2M): Predicting Severe Retinopathy of Prematurity," 2019 International Conference on Deep Learning and Machine Learning in Emerging Applications (Deep-ML), 2019, pp. 61-66, doi: 10.1109/Deep-ML.2019.00020.
7. R. D. Sah, S. P. Patro, N. Padhy and N. Salimath, "Diabetics Patients Analysis Using Deep Learning and Gradient Boosted Trees," 2021 8th International Conference on Computing for Sustainable Global Development (INDIACom), 2021, pp. 937-941.
8. B. Shickel, P. J. Tighe, A. Bihorac and P. Rashidi, "Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR) Analysis," in IEEE Journal of Biomedical and Health Informatics, vol. 22, no. 5, pp. 1589-1604, Sept. 2018, doi: 10.1109/JBHI.2017.2767063.
9. W. A. Zgallai, T. Brown, E. Z. Dalah, A. K. Obaideen and M. E. Faris, "Automation of the Assessment of Subcutaneous Adipose Tissue and Visceral Adipose Tissue in Human Abdominal Area Employing Deep Learning," 2021 IEEE Western New York Image and Signal Processing Workshop (WNYISPW), 2021, pp. 1-5, doi: 10.1109/WNYISPW53194.2021.9661279.
10. D. Shah, A. Malhotra and M. Patidar, "Sensor Based Human Activity Recognition by Multi-Headed ConvLSTM," 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), 2021, pp. 1127-1129, doi: 10.1109/ICACCS51430.2021.9441948.
11. S. C. Cetindag, K. Guran and G. Bilgin, "Transfer Learning Methods for Using Textural Features in Histopathological Image Classification," 2020 Medical Technologies Congress (TIPTEKNO), 2020, pp. 1-4, doi: 10.1109/TIPTEKNO50054.2020.9299220.
12. A. S. Rajawat, O. Mohammed and P. Bedi, "FDLM: Fusion Deep Learning Model for Classifying Obstructive Sleep Apnea and Type 2 Diabetes," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 835-839, doi: 10.1109/I-SMAC49090.2020.9243553.
13. D. I. Dogaru and I. Dumitrache, "Big Data and Machine Learning Framework in Healthcare," 2019 E-Health and Bioengineering Conference (EHB), 2019, pp. 1-4, doi: 10.1109/EHB47216.2019.8969944.

14. G. Gogi, A. Gegov, M. Bader-El-Den and B. Vatchova, "Application of Deep Learning in Cardiovascular Medicine using Multiple Biomedical Data," 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 200-204, doi: 10.1109/IS48319.2020.9199844.
15. M. A. R. Refat, M. A. Amin, C. Kaushal, M. N. Yeasmin and M. K. Islam, "A Comparative Analysis of Early Stage Diabetes Prediction using Machine Learning and Deep Learning Approach," 2021 6th International Conference on Signal Processing, Computing and Control (ISPCC), 2021, pp. 654-659, doi: 10.1109/ISPCC53510.2021.9609364.
16. A. K. Mahapatra, R. Satyam Oza and S. Shankar, "Deploying Deep Transfer Learning Models to a Web-App for Sars-Cov-2 Detection Using Chest Radiography Images," 2021 5th International Conference on Electrical, Electronics, Communication, Computer Technologies and Optimization Techniques (ICEECCOT), 2021, pp. 137-144, doi: 10.1109/ICEECCOT52851.2021.9707991.
17. A. Mashyal, B. Chougula, S. Kobal, H. G. Bajantri and Veeresh, "Facial Mask Detection and Alert System," 2021 International Conference on Intelligent Technologies (CONIT), 2021, pp. 1-5, doi: 10.1109/CONIT51480.2021.9498278.
18. Antima Jain, Anand Singh Rajawat, Rabindra Nath Shaw and Vishesh Kumar Mishra Conference: 2021 IEEE 4th International Conference on Computing, Power and Communication Technologies (GUCON), Year: 2021, Page 1 DOI: 10.1109/GUCON50781.2021.9573624
19. N. Sadman, S. Tasneem, A. Haque, M. M. Islam, M. M. Ahsan and K. D. Gupta, "'Can NLP techniques be utilized as a reliable tool for medical science?'" - Building a NLP Framework to Classify Medical Reports," 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2020, pp. 0159-0166, doi: 10.1109/IEMCON51383.2020.9284834.
20. Y. Wu, J. Ma, X. Huang, S. H. Ling and S. Weidong Su, "DeepMMSA: A Novel Multimodal Deep Learning Method for Non-small Cell Lung Cancer Survival Analysis," 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2021, pp. 1468-1472, doi: 10.1109/SMC52423.2021.9658891.