A SURVEY ON DEEP, MACHINE LEARNING METHODS USED IN HEALTHCARE

V. Naren Thiruvalar
Research Scholar, Department of EEE, Anna University Regional Campus, Coimbatore

Dr.M. Yuvaraju,
Assistant Professor, Department of EEE, Anna University Regional Campus, Coimbatore

1. INTRODUCTION:

The technology that allows computers and other machines to imitate human intelligence and problem-solving abilities is referred to as artificial intelligence, or AI. AI has the potential of accomplishing tasks that typically require human intelligence or guidance, either on its own or in collaboration with other technologies (like sensors, geolocation, and robotics). A few cases of AI in the news and in our daily lives are digital assistants, GPS navigation, autonomous vehicles, and generative AI tools. Artificial intelligence is a discipline of computer science which comprises machine learning and deep learning, and is frequently discussed in conjunction with them. In these fields, artificial intelligence (AI) algorithms that replicate human decision-making processes evolve with the ability to learn from available data and produce progressively more precise predictions and categorizations over time.

1.1 MACHINE LEARNING:

Within the field of artificial intelligence (AI), machine learning (ML) gives machines the capacity to self-learn from data and historical experiences in order to detect patterns and anticipate outcomes with the least amount of human involvement. Without explicit programming, computers can function independently due to machine learning techniques. They have the ability to learn, grow, evolve, and adapt on their own. Large data sets can yield useful information via machine learning, which uses algorithms to find patterns and learn through iterations.

A training dataset is employed to shape machine learning algorithms to generate a model. The trained machine learning algorithm utilizes its built model to anticipate outcomes when it receives new input data. ML algorithms do not rely on any preset equation that may be used as a model; instead, they use computation techniques to learn directly from data. As the number of samples available increases during the learning processes, the performance of machine learning algorithms improves appropriately.

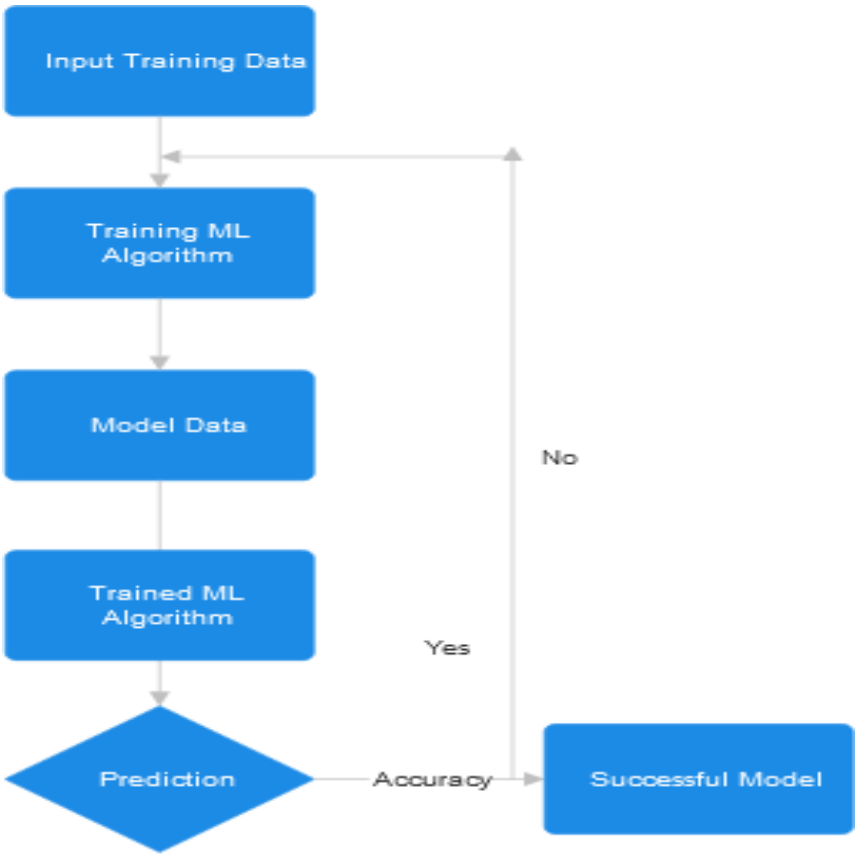


Fig.1. Machine Learning Workflow

1.1 Types

1.1.1 Supervised ML

Under supervision, machines are trained on labeled datasets and granted the ability to anticipate outputs based on the training they have received. According to the labeled dataset, some input and output parameters have already been mapped. The input and matching output therefore serve to teach the machine. In subsequent phases, a device is created to forecast the outcome based on the test dataset. For example, a dataset of photos of tigers and lions is taken into account. The computer is initially trained to recognize the colors, shapes, sizes, and eye colors of the lion and tiger in the images. Following training, the computer is given an input image of a lion, and its task is to recognize the item and predict the result.

In order to arrive at a final forecast, the trained machine analyzes the inputpicture for the object's numerous properties, including color, eyes, shape, and so forth. Thus it does the object identification. Mapping the input variable to the output variable is the main goal of this technique.

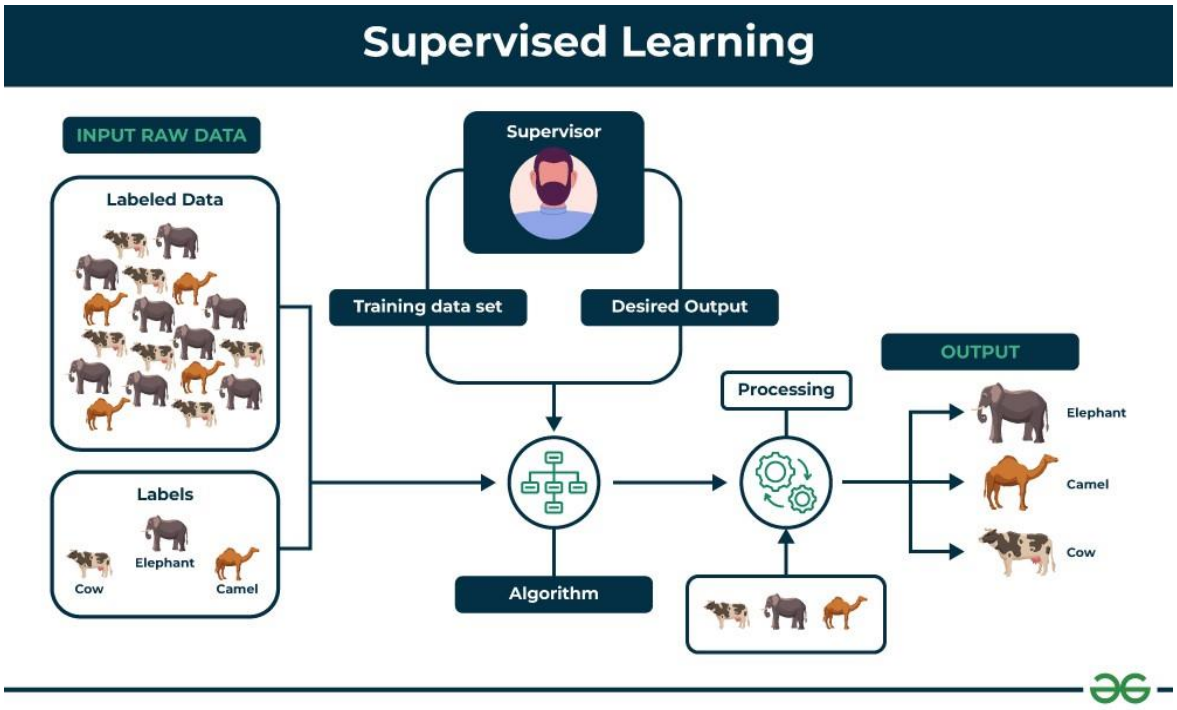


Fig.2. Supervised Learning

1.1.2 Un-supervised ML

Unsupervised learning is the process of examining and grouping unlabeled datasets using machine learning algorithms. These algorithms identify hidden relationships or patterns in the data without necessitating human involvement. It is the best approach for exploratory data analysis, cross-selling tactics, consumer segmentation, and image recognition due to its capacity to identify patterns in data.

The process of instructing a machine with neither labeled nor classed data and letting the algorithm make choices depending only on that data without human supervision is known as unsupervised learning. The machine's job is to classify unsorted data based on similarities, patterns, and differences without requiring any prior data training. In contrast to supervised learning, the absence of a teacher implies that the machine will not receive any sort of guidance. Consequently, the machine's ability to independently identify the hidden structure in unlabeled data is limited.

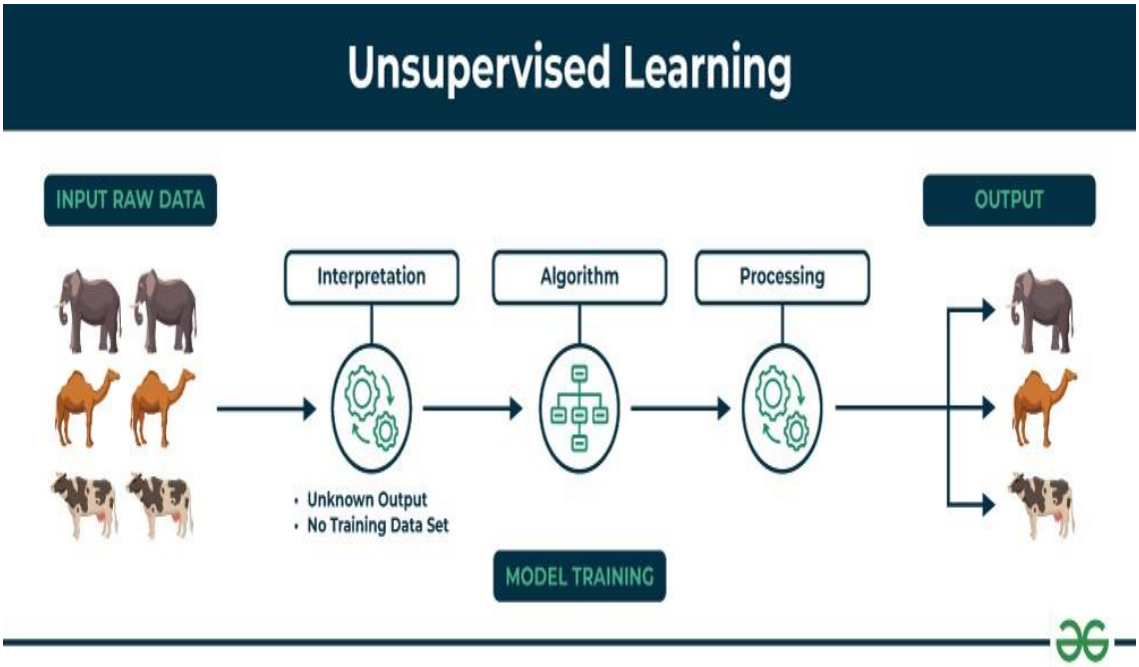


Fig.3. Un-supervised Learning

1.1.3 Reinforcement ML

Within machine learning lies the field of reinforcement learning. It involves acting appropriately to maximize reward in a certain scenario. It is utilized by a variety of applications and devices to determine the optimal course of action or behaviour for an instance. Reinforcement learning and supervised learning are different in that in supervised learning, the model is trained with the correct answer already included in the training data, while in reinforcement learning, the model is trained in an absence of an answer and is directed by the reinforcement agent's decision on the task completion. It is certain to learn from its experience even in the lack of a training dataset. It involves figuring out how to behave in a situation to maximize rewards. Trial- and-error machine learning systems provide the data used in reinforcement learning (RL). Whether it is supervised or unsupervised machine learning, data is not an input source. Algorithms that use reinforcement learning examine results and determine the best course of action are used. The algorithm gets feedback after every action to help it decide if the decision it made was neutral,incorrect, or correct. It's a useful method for automated systems that need to make numerous minor judgements on their own without human assistance. A self-governing, self-teaching system, reinforcement learning basically learns bymaking mistakes.

Fig.4. Reinforcement Learning

SURVEY TABLE

S. No.	AUTHOR & YEAR	METHOD USED	REMARKS
1.	MIN CHEN et.al June 2017	CNN based Multimodal Disease Risk Prediction Algorithm (CNN-MDRP)	The proposed CNN-MDRP algorithm achieves a prediction accuracy of **94.8%** and converges faster than existing CNN-based unimodal disease risk prediction algorithms. In this paper, machine learning algorithms are used for effective prediction of chronic disease outbreak in disease-frequent communities.
2.	Xianlong Zeng et.al Jan 2021	Multi-View Deep Learning Framework	This context describes a study that proposed a multi-view deep learning framework to predict future healthcare expenditures. The approach models demographic, medical code, drug and facility utilization data separately and then combines them. Experimental results on a pediatric dataset showed the proposed method out-performed baselines in predicting medical expenditures Applying the model to a high utilizer selection task showed promise in helping allocate healthcare resources better.
3.	Jayroop Ramesh Et.al April 2021	Support Vector Machine Algorithm	A remote monitoring framework has been proposed for automated diabetes risk prediction and management using personal health devices, wearables, and smartphones. The framework utilizes a support vector machine algorithm to predict diabetes risk based on the Pima Indian Diabetes Database. The algorithm achieved an accuracy of **83.20%**, sensitivity of 87.20%, and specificity of 79%. The framework allows patients to use multiple healthcare devices to measure vital parameters and communicate with medical professionals. The system automates diabetes detection and alerts medical professionals for timely intervention.
4.	Ankush D. Jamthikar Et.al 2022	Ensemble Machine Learning	This study aims to show the effectiveness of ensemble-learning-driven machine learningalgorithms in predicting cardiovascular events such as coronary artery disease (CAD) and acute coronary syndrome (ACS) compared to conventional ML algorithms. The study collected and analyzed 24 risk predictors and carotid

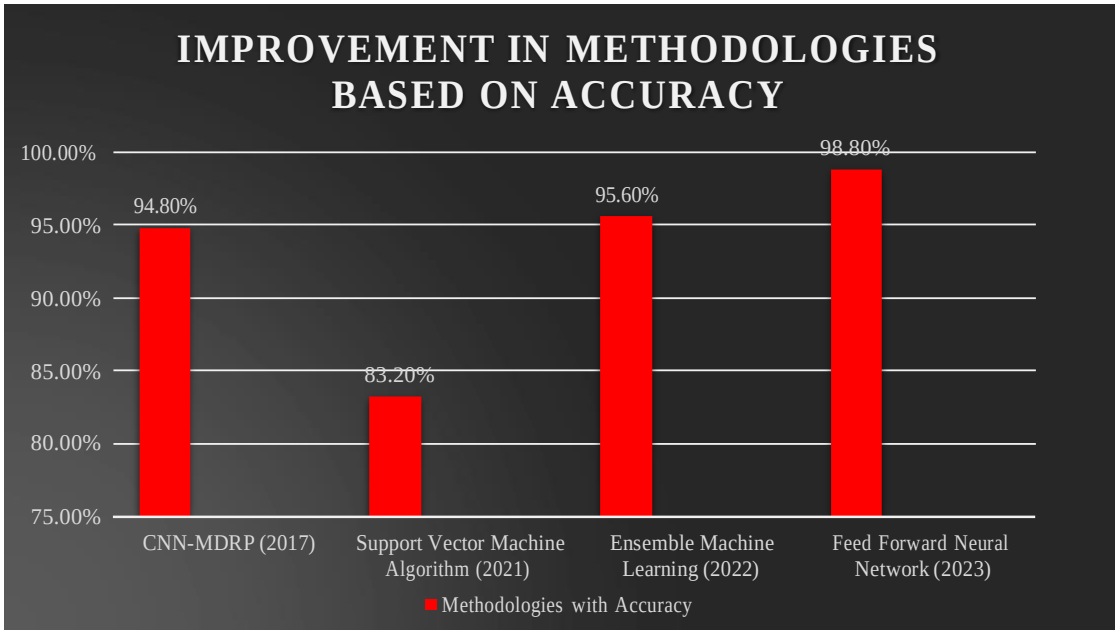
			ultrasound image phenotypes as features. The ensemble ML (EML) systems were found to perform better than conventional ML (CML) systems in predicting CAD and ACS, with improvedarea under the curve (AUC) values. The proposed EML-based system was validated against two databases and proved to be more accurate in predicting CAD and ACS. Among the ensemble algorithms, XGBoost and RF had better prediction abilities for CAD and ACS, respectively.
5.	Usman Naseem Et.Al July 2022	Ensemble Machine Learning	Breast cancer (BC) is the most deadliest cancer among women, making it crucial to detect it early for better prognosis and recovery. To achieve this, a system using machine learning (ML) has been proposed. ML algorithms, including Artificial Neural Networks (ANN), have proven effective in classifying BC patterns. In this study, an ensemble of classifiers is utilized for automatic BC detection and prognosis. Various ML algorithms and ensemble models, with and without up-sampling techniques, are evaluated on benchmark datasets. Results show that the ensemble method outperformed other state-of-the-art methods, yielding an accuracy of **98.83%**. The proposed system is of great importance to the medical industry and research community. SVM outperforms other ML classifiers and deep learning (DL) classifiers individually. Performance is further improved with balanced class weights and the up-sampling technique.
6.	Xin Wang et.al Sep 2022	Ensemble Machine Learning	A machine learning model was proposed to identify and predict risk factors for hospitalization outcomes in geriatric patients with dementia. The study analyzed over 150 clinical and demographic factors of 15,678 encounters and narrowed them down to twenty factors. The model achieved an accuracy of **95.6%** and outperformed other risk assessment methods for different types of dementia. The top identified risk factors included encephalopathy, number of medical problems at admission, pressure ulcers, urinary tract infections, falls, admission source, age, race, and anaemia. These factors can be modified or intervened to prevent adverse effects and improve outcomes. The study provides valuable insights for healthcare systems and clinicians to implement early interventions and improve the quality of life for geriatric dementia patients

7.	Giovanni Annuzzi Et.Al Feb 2023	Feed Forward Neural Network	A critical issue in T1D patients is the managing of Postprandial Glucose Response (PGR), through the dosing of the insulin bolus to inject before meals. Carbohydrates, proteins, lipids, fibers, and energy intake) in the short and middle term on Blood Glucose Levels (BGLs) prediction was conducted by Machine Learning (ML) methods. The aim of this research work was to experimentally investigate the impact of nutritional factors on the capability to obtain postprandial BGL predictions, in the short and middle term after a meal, by machine learning methods. A series of experiments to predict BGLs was carried out on the self-produced AI4PG dataset containing both CGM measurements and meal nutritional factors of a set of patients. In particular, the impact of nutritional factors such as carbohydrates, proteins, lipids, and fibres, as well as meal energy intake, on postprandial blood glucose response was investigated.
8.	Yeeun Yoo Et.Al Aug 2023	Graph Neural Network and Machine Learning	This study demonstrates that Medicare fraud detection can be significantly enhanced by introducing graph analysis with considering the relationships among medical providers, beneficiaries, and physicians. In this study, fraud- detection models were developed using two approaches to reflect graph information: a graph GNN and a conventional machine-learning model with input features of graph centralities. This study is the first to apply various GNN and machine learning algorithms to detect Medicare fraud. Recently, various fraud detection studies have been conducted with the emergence of GNN algorithms that can learn feature information from a graph- structured dataset. The machine learning model withgraph centrality features showed improved precision of 4 percent point (%p), recall of 24 %p, and F1- score of 14 %p compared to the best GNN model. This outcome suggests that successful and efficient detection of Medicare fraud can be achieved if graph centrality measures are used to capture the relationships among medical providers, physicians, and beneficiaries.

9.	Mustufa Haider Abidi Et.Al Oct 2023	Hybrid Dingo Coyote Optimization Deep Ensemble Learning (HDCO-DEL)	Human life has become smarter by utilizing big data, telecommunication technologies, and wearable sensors over pervasive computing to give better healthcare services. Big data makes the interconnection between patients, wearable sensors, healthcare caregivers, and providers through the utilization of Information and Communication Technology (ICT) and software. The combiner phase classifies the physical activities using the developed Deep Ensemble Learning (DEL) consisting of classifiers such as Extreme Learning Machine (ELM), deep Convolutional Neural Network (CNN), Long short-term memory (LSTM),Deep Belief Network (DBN), and Deep Neural Network (DNN). The developed HDCO-DEL has secured better accuracy than ELM, CNN, LSTM, DBN, DNN, and Health Fog, respectively on second dataset. The comparison with existing methods shows its better performance and also predicts physical activities with overall high accuracy. Therefore, it can be inferred that the proposed Map Reduce framework for the health monitoring model with HDCO-DEL based provided a better prediction of health monitoring performance than the existing techniques based on two datasets.
10.	Krishnakant V. Saboo et.al Oct 2023	Relative Entropy Based Machine Learning Models using EEG	Predicting the onset of seizure clusters is crucial to enable patients to receive preventative treatments. This paper presents machine learning models that use bivariate intracranial EEG (iEEG) features to predict seizure clustering. Specifically, they utilized relative entropy (REN) as a bivariate feature to capture potential differences in brain region interactions underlying isolated and cluster seizures. They observed that REN was significantly different between isolated and cluster seizures in majority of the patients. 5% Area under the ROC Curve (AUC), predicted if a seizure is the first one in a cluster withup to 55. Overall, the findings could be beneficial in addressing the clinical burden associated with seizure clusters, enabling patients to receive timely treatments and improving their quality of life.
11.	Muhammad Hamza Rafique Bhatti et.al	hybrid Deep Learning (DL) models	Accurate prediction of healthcare costs is challenging due to various factors that affect the overall prediction. These factors include the increasing cost of healthcare providers, which is

	Dec 2023		influenced by inflation and population growth. To address this, the study proposes three hybrid Deep Learning (DL) models: VGG-SAE, VGG-DNN, and SAE-DNN. These models optimize the learning of hidden patterns from the data more efficiently than individual models. The pre- processing techniques used include mode imputation for handling missing values, Z-score for removing outliers, and standard scaling for data standardization. The Random Search technique is employed to find the best hyper- parameters for each hybrid model. The performance of the hybrid models is compared to baseline DL models. Additionally, the robustness of the proposedapproach is assessed using two different healthcare datasets. The results show that the hybrid models outperform other deep models and Machine Learning (ML) techniques in terms of healthcare providers cost, demonstrating their high efficiency.

Summary



REFERENCES

[1] M. Chen, Y. Hao, K. Hwang, L. Wang and L. Wang, "Disease Prediction by Machine Learning Over Big Data from Healthcare Communities," IEEE Access, vol. 5, pp. 8869-8879, 2017, doi: 10.1109/ACCESS.2017.2694446.

[2]X. Zeng, S. Lin and C. Liu, "Multi-View Deep Learning Framework for Predicting Patient Expenditure in Healthcare," IEEE Open Journal of the Computer Society, vol. 2, pp. 62-71, 2021, doi: 10.1109/OJCS.2021.3052518.

[3]Ramesh J, Aburukba R, Sagahyroon A. A remote healthcare monitoring framework for diabetes prediction using machine learning. Healthc Technol Lett. 2021 May 2;8(3):45-57. doi: 10.1049/htl2.12010. PMID: 34035925; PMCID: PMC8136765.

[4] A. D. Jamthikar, D. Gupta, L. E. Mantella, L. Saba, A. M. Johri and J. S. Suri, "Ensemble Machine Learning and Its Validation for Prediction of Coronary Artery Disease and Acute Coronary Syndrome Using Focused Carotid Ultrasound," in *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-10, 2022, Art no. 2503810, doi: 10.1109/TIM.2021.3139693.

[5]U. Naseem *et al*., "An Automatic Detection of Breast Cancer Diagnosis and Prognosis Based on Machine Learning Using Ensemble of Classifiers," in *IEEE Access*, vol. 10, pp. 78242-78252, 2022, doi: 10.1109/ACCESS.2022.3174599.

[6]Wang X, Ezeana CF, Wang L, Puppala M, Huang YS, He Y, Yu X, Yin Z, Zhao H, Lai EC, Wong STC. "Risk factors and machine learning model for predicting hospitalization outcomes in geriatric patients with dementia", Alzheimer's Dement (N Y). 2022 Sep 29;8(1):e12351. doi: 10.1002/trc2.12351. PMID: 36204350; PMCID: PMC9520763.

[7] G. Annuzzi *et al*., "Impact of Nutritional Factors in Blood Glucose Prediction in Type 1 Diabetes Through Machine Learning," *IEEE Access*, vol. 11, pp. 17104-17115, 2023, doi: 10.1109/ACCESS.2023.3244712.

[8] Y. Yoo, J. Shin and S. Kyeong, "Medicare Fraud Detection Using Graph Analysis: A Comparative Study of Machine Learning and Graph Neural Networks," *IEEE Access*, vol. 11, pp. 88278-88294, 2023, doi: 10.1109/ACCESS.2023.3305962.

[9] M. H. Abidi, U. Umer, S. H. Mian and A. Al-Ahmari, "Big Data-Based Smart Health Monitoring System: Using Deep Ensemble Learning," in *IEEE Access*, vol. 11, pp. 114880-114903, 2023, doi: 10.1109/ACCESS.2023.3325323.

[10]Saboo KV, Cao Y, Kremen V, Sladky V, Gregg NM, Arnold PM, Karoly PJ, Freestone DR, Cook MJ, Worrell GA, Iyer RK. Individualized Seizure Cluster Prediction Using Machine Learning and Chronic Ambulatory Intracranial EEG. IEEE Trans Nanobioscience. 2023 Oct;22(4):818-827. doi: 10.1109/TNB.2023.3275037.

[11] M. H. R. Bhatti, N. Javaid, B. Mansoor, N. Alrajeh, M. Aslam and M. Asad, "New Hybrid Deep Learning Models to Predict Cost From Healthcare Providers in Smart Hospitals," in *IEEE Access*, vol. 11, pp. 136988-137010, 2023, doi: 10.1109/ACCESS.2023.3336424.