Tuijin Jishu/Journal of Propulsion Technology
ISSN: 1001-4055
Vol. 44 No. 5 (2023)

Concept Extraction using Machine Learning Techniques In healthcare

Manya M.S [1] & Dr.Amuthabala[2]
[1]Research Scholar (CSE), Reva University, Bengaluru ,India
[2] Assistant Professor, Reva University, Bengaluru, India

ABSTRACT

Concept extraction is a natural language processing (NLP) task that involves identifying and extracting specific entities or concepts from unstructured text data. These entities could be names of people, organizations, locations, dates, or any other relevant information.

Keywords: Conditional Random Fields (CRF),Support Vector Machines (SVM) and Decision Trees, Named Entity Recognition s(NER) models

1. INTRODUCTION

Machine learning techniques are often employed for concept extraction .The use of Machine learning techniques for concept extraction is common because of their propensity to manage the complexity of natural language and produce precise predictions.

1. Data Collection and Preprocessing

Gather a dataset containing the textual content you wish to use to extract concepts is the first step. This data can be obtained from various sources such as websites, documents, or social media. Once collected, the data will be preprocessed, which involves tasks like tokenization, removing stop words, and handling special characters.

2. Annotated Data:

For supervised learning, you'll need annotated data, where each piece of text is labeled with the concepts to extract. The machine learning models need this data to be trained.

3. Feature Extraction:

Text data must be represented as numerical features in order to train a machine learning model. Common techniques include using bag-of-words representations, TF-IDF (Term Frequency-Inverse Document Frequency), word embeddings like Word2Vec or GloVe, and more advanced contextual embeddings like BERT (Bidirectional Encoder Representations from Transformers).

4. Model Selection:

Many machine learning techniques are appropriate for jobs involving concept extraction. Typical options include.

Named Entity Recognition (NER) models: These models were designed specifically for extracting entities like names of people, organizations, locations, etc.

Conditional Random Fields (CRF): A sequence labeling algorithm that is commonly used for NER tasks.
Support Vector Machines (SVM) and Decision Trees: These classifiers can be useful for concept extraction when combined with appropriate feature representations.

Deep Learning models: Approaches like recurrent neural networks (RNNs), LSTM (Long Short-Term Memory), and transformer-based models (e.g., BERT) e**xhibit positive divergence** results for NLP tasks, including concept extraction.

5. Model Training:

Create divisions within your dataset for both training and validation purposes. Train the chosen model using the annotated information and the features that have been extracted. The model will acquire the ability to identify patterns within the text that align with the concepts intended for extraction.

6. Evaluation:

Evaluate your model's performance using suitable metrics such as precision, recall, F1-score, or accuracy. This stage aids in comprehending the effectiveness of your model and whether it achieves the desired level of accuracy

7. Deployment:

Once you have a satisfactory model, you can deploy it to process new, unseen text data and extract concepts from it in real-time.

It's crucial to keep in mind that the performance of concept extraction is heavily reliant on the quality and quantity of annotated training data, along with the careful selection of suitable features and algorithms. Furthermore, in certain scenarios, rule-based systems can be utilized in conjunction with machine learning methods to enhance concept extraction endeavors. Machine learning (ML) is an evolving domain with significant potential to enhance healthcare services where handling the information is crucial task, the integration of this machine learning modalities to provide a different service from patient treatment to healing, protective, rehabilitation, and palliative care has revisualized the health care system. Currently, smart health care services with machine learning techniques have been integrated into multiple domains of clinical practice, biomedical research, and healthcare administration. Integration of machine learning into clinical medicine shows potential for significantly enhancing the delivery of healthcare services to patients. This research comprehensively covers recent progress, studies, applications, as well as various classifications and algorithms of machine learning within the healthcare sector.

Challenges :

Employing machine learning methods for concept extraction presents a set of distinct challenges. Among these, some significant hurdles encompass:

1. Insufficient and Biased Training Data: Machine learning models require a significant amount of annotated training data to learn effectively. Acquiring high-quality labeled data for concept extraction can be expensive and time-consuming. Moreover, the training data might be biased towards specific domains or sources, making the model less effective in handling diverse data.

2. Ambiguity in Language:Authentic text frequently carries ambiguity, as words possess diverse interpretations depending on the context. This ambiguity can result in misunderstandings and inaccurate concept identification

3.Rare and Out-of-Vocabulary Concepts:Machine learning models might struggle with extracting rare or previously unseen concepts, particularly when they are absent from the training dataset. Handling out-of-vocabulary words becomes a challenge as they might frequent in some scenarios, leading to information loss.

4. Named Entity Variation:Named entities like people's names, locations, or company names can have various forms and spellings. Variations in the representation of these entities make it challenging for the model to generalize and accurately extract them.

5.Contextual Understanding:Concepts often depend on the context in which they appear. Machine learning models, especially traditional ones, might struggle to capture long-range dependencies and context-specific information, limiting their ability to correctly extract concepts.

6. Noisy and Inconsistent Data: Textual information gathered from diverse origins can exhibit noise and incongruities, containing errors, misspellings, and irregularities. These noise factors can negatively impact the model's performance.

7. Scalability: Certain machine learning models, especially those in the realm of deep learning, might demand considerable computational resources for both training and inference. Enabling these models to efficiently process extensive data volumes in real-time poses a formidable scaling challenge.

8. Model Interpretability: Deep learning models, especially transformer-based architectures like BERT, are often considered black boxes. Understanding why a model made a particular concept extraction decision can be difficult, This issue becomes particularly problematic in critical applications where interpretability is necessary.

9. Domain Adaptation:Models trained on information from a specific domain might exhibit subpar performance when applied to textual data originating from an different domain. Adapting the concept extraction model to new domains or updating it with new data requires careful consideration.

10. Data Privacy and Security:Concept extraction models trained on sensitive text data usage can potentially raise privacy concerns and security concerns if not handled carefully. Proper anonymization and data protection measures need to be implemented.

To overcome these challenges, researchers and practitioners are continuously working on improving machine learning techniques, developing novel approaches, and exploring hybrid approaches that integrate machine learning techniques with rule-based systems or domain-specific knowledge. Additionally, advances in transfer learning and pre-training models on large-scale datasets have shown promising results in addressing some of these challenges.

1.General Healthcare

Every day, the healthcare industry manages a substantial volume of data from traditional software or hardware. Clinical and lab data, medical notes, machine generated data from medical equipment or at home monitoring sensors, health services financial data, hospital bills, literature data from medical journals, social media posts blogs in health subjects, and so on are all instances of health data. These data may be available internally in health services (e.g., EHR, LIMS) or may come from external sources (e.g., insurance companies, pharmacies, government) and may be in a structured (e.g., tables with laboratory results) or unstructured (e.g., text of medical notes in EHR) format and this leveraging big data analytics in the healthcare sector involve the methods of analysing the wide amount of electronic data related to patient healthcare and well-being [1].

One industry sector that stands to gain the most from the availability and growth of data is healthcare [10]. Massive volumes of information sourced from research, clinical studies public health, and insurance programmes are now being compiled by organisations including healthcare providers, pharmaceutical firms, academic institutions, and governmental organisations. There's considerable potential in amalgamating data from diverse origins [11]. However, medical practitioners are now adopting evidence-based medicine as a means of diagnosing and treating patients. Previously, doctors relied on symptoms for treatment decisions. Making informed choices by utilizing the most reliable data involves analyzing substantial datasets derived from clinical trials and other therapeutic approaches routes on a wide scale [12].

Healthcare deserves special attention because it is directly related to people's lives. There are specific issues and trends associated with information systems and their management in healthcare, advancements in research and technology as introduced in various ways such as ML, Artificial intelligence that's contributing significantly to revolutionising the healthcare sector [2]. Machine learning algorithms, employed for diagnostic purposes, hold significant significance in identifying treatments and analysing public health threats analyse medical records and disease forecasts using innovative technologies and tools to advance patient care and services [3].

In several disciplines, machine learning has made an indisputable contribution to medical research. Even so, a substantial amount of algorithms have been created via machine learning. Worldwide, many repressive possibilities have been enabled by machine learning algorithms. Frequently, machine learning algorithms utilize logic and mathematics to make simple predictions from a dataset. Employing machine learning algorithms enables be a highly accessible way to diagnosis complicated illnesses. Since there has proven to be a lot of study done on machine learning algorithms within the healthcare sector for disease prediction. Numerous scholars globally have achieved notable progress in predictive diagnostics through the application of machine learning algorithms [9].

2.Machine Learning:

Computer programmes (algorithms) acquire connections of predictive ability from examples in data via a procedure referred to as machine learning. The simplest definition of machine learning involves the usage of computers to apply statistical models to data. A wider range of statistical methods than those generally employed in medicine are applied in machine learning. In order to manage increasingly complicated data, more recent techniques, like Deep Learning, are built on models that make less assumptions about the underlying data [4].Machine Learning is a specialized method of data analytics that automates the creation of models, particularly in relation to model development. It's important to emphasize that in machine learning, we're not explicitly telling computers where to look. Instead, machines acquire the ability to employ specific techniques to unearth concealed insights from data.

Machine learning stands as an iterative process that enables the computer to modify its strategies and results in response to fresh events and data [5].In accordance with their specific objectives, machine learning algorithms may be broadly divided into three types. It comprises reinforcement learning, unsupervised learning, and supervised learning

Supervised learning: In the realm of supervised learning, a model undergoes training using labelled data, and then it is applied to new data to produce predictions. Data needs to be partitioned into two distinct sets: a training set and a testing set. Initially, the model is trained using the training set, and subsequently, its performance is assessed using the testing set. Performance metrics can be employed to evaluate the model's efficacy, and supervised learning can be applied to address classification or regression issues [6].

Unsupervised learning: Unsupervised learning is the same as utilizing supervised learning within that it includes data training, but the annotated value or target value is unknown. By identifying the underlying pattern, the machine attempts to group similar types of data into clusters. Unsupervised learning's primary goal is to identify patterns, not for the purpose of making predictions [7].

Semi-supervised learning: While supervised learning involves usage of labelled data where as unsupervised learning uses unlabelled data, a significant amount of data from a labeled so data is lost. This knowledge can still be gained from unlabelled data. Semi-supervised learning therefore comes to mind in this situation. It uses both unlabelled and labelled data, which represents a fusion of both supervised and unsupervised learning paradigms.
Reinforcement learning: Learning through reinforcement develops a system whose performance is improved by incorporating input from the surrounding environment and choosing appropriate action to enhance them. Without human assistance, it involves acquiring knowledge from the world by actively engaging with it, constituting an ongoing and continuous progression[8].

3.Machine learning in Health Care:

The healthcare sector is one such sector where machine learning may have significant social effects. The utilization of machine learning to analyze this data is increasingly prevalent within an expanding sector driven by devices like smartwatches, Fitbits, and similar gadgets that consistently gather a plethora of health-related information [13]. Machine learning has the potential to address the healthcare industry's rising expenses and enhance the patient-doctor relationship. Numerous health-oriented applications of machine learning and big data encompass aiding doctors in formulating more personalized prescriptions and treatments for patients, while also supporting patients in managing their own health in deciding whether or not to arrange follow-up appointments.

Advancements in science have considerably facilitated human existence through technology. Machine learning algorithms, being readily available, play a significant role in diagnosing intricate diseases, demonstrating their efficacy in medical applications..

1.Clinical Data Analysis using Machine Learning

Wengert and colleagues [14] proposed the use of machine learning (ML) algorithms to predict the early occurrence of pathological complete response (pcr) in response to neoadjuvant chemotherapy and to forecast the survival outcomes of breast cancer patients. They applied these algorithms to Multiparametric Magnetic Resonance Imaging (mpMRI) data. Eight different classifiers were employed to assess features related to pcr, such as residual cancer burden (RCB), recurrence-free survival (RFS), and disease-specific survival (DSS). These classifiers included methods like linear support vector machine, linear discriminant analysis, logistic regression, random forests, stochastic gradient descent, adaptive boosting, and Extreme Gradient Boosting (XGBoost). The area under the curve (AUC) was used to quantify the performance of each classifier for each PCR feature. The experimental findings revealed that XGBoost exhibited superior accuracy compared to other classifiers for RCB and DSS, while logistic regression excelled for RSS [14].

For the aim of predicting non-small cell lung cancer patients' two-year survival, Dagli et al. [15] constructed a multilayer perception model. The RelifF a feature selection methodology, the attributes of 559 patient samples were ranked. Through this process, a Multilayer Neural Network demonstrated the most effective prediction performance, achieving an area under the curve value of 0.75. In the context of forecasting patient survival for individuals with hepatocellular carcinoma (HCC), Kayal and colleagues [16] conducted their study suggested a new, better categorization technique. Authors studied 165 patient samples and determined that 15 risk factors—out of 49 risk factors—were accountable for HCC.
The experimental outcomes revealed a notable discrepancy in accuracy, with Deep Neural Networks outperforming Cox models (SVM) and Unsupervised models (KNN).

In order to predict the risk of delirium based on data from HER (Electronic Health Records), Five distinct machine learning techniques, including penalized, gradient boosting, linear support vector, logistic regression, an artificial neural network with a single hidden layer, and random forest, were examined by Andrew et al. [17]. In all, 18223 patient samples were gathered, and an investigation was conducted. The research results demonstrated that the gradient boosting strategy, with an AUC value of 0.855, produced the greatest results. Fatemeh et al. [18] proposed machine learning algorithms for initial emergency admission prediction based on EHR data. The authors used the Cox model, followed by the random forest and gradient boosting algorithms, to calculate the likelihood of the first emergency admission for a sample of 4.6 million patient samples. With an AUC score of 0.779, the authors claim that the gbm model performed the best [18].

Maryam et al. [19] investigated the Seattle heart failure model for the prediction of heart failure using EHR data. Samples from 5044 patients were acquired, and attributes were then retrieved in order to calculate the survival score. The Cox proportional regression model's survival score for cardiac patients who lived for one, two, or five years was initially estimated by the authors. The remaining patients were exposed to a variety of machine learning models, including random forest, logistic regression, support vector regression, decision tree, and ada boost, after patients who died after five years were excluded. The results of the experiment showed that logistic regression increased the value of the AUV curve by 11% [19]. Zheng et al. [20] published a method to identify Type-2 Diabetes Mellitus (T2DM) patients using Electronic Health Record (EHR) data. 300 patient samples in all were gathered, and 114 characteristics were taken out. Then, different machine learning methods, including k-Nearest Neighbor (kNN), Random Forest (RF), Decision Tree (DT), naive bayes, Support Vector Machine (SVM), and logistic regression, were applied to these characteristics. The results show that SVM produces the best output, with an accuracy of 96%. Zheng et al. [21] developed a system to identify Type-2 Diabetes Mellitus (T2DM) patients using Electronic Health Record (EHR) data. After 114 features had been retrieved from a total of 300 patient samples, a number of machine learning techniques, including k-Nearest Neighbor (kNN), Random Forest (RF), Decision Tree (DT), naive bayes, Support Vector Machine (SVM), and logistic regression, had been used. The results show that SVM produces the best output, with an accuracy rate of 96%.

2. Machine Learning in Medical Imaging

Accurate illness diagnosis using extensive medical data processing is becoming essential in the medical community. Machine learning algorithms represent being used for a variety of tasks in the fields within the realm of biology and medicine. Among the applications are the distribution of data determined by their characteristics, the examination of medical data, the planning of disease diagnosis and treatment, the collection and inspection of data, the correction of diagnostic of various diseases by medical imaging, and the extraction of features from medical images on diseases. As the uses of medical imaging are given more thought, it becomes clear that medical imaging is widely utilised to enhance the planning of surgical treatments with reference to a variety of disorders. Before examining the role of machine learning in medical imaging, the following applications will help illustrate how medical imaging finds application in surgical planning to achieve successful outcomes while minimising risks. The intriguing aspect is how machine learning may be used in conjunction with these medical imaging methods to improve surgery planning. It should be emphasised that when analysing several data sets, machine learning might reveal hidden links that may not be immediately obvious to humans. Because of this, machine learning techniques have even made it feasible for healthcare professionals to forecast diseases. As a result, by processing these medical image data using unsupervised machine learning techniques like clustering, an analysis could be done on the dataset that could later be used by the surgeon to determine if any important information has been lost while planning the surgery or even to even confirm that the decisions made on the approach of performing the surgery were correct. The analysis of medical images is constantly changing as a result of technological advancement. The development of 3D virtual models also contributes to this by enhancing comprehension of complicated anatomy and by offering effective tools for surgical planning and intraoperative guiding. Currently, clinical practise is increasingly utilising foetal MRI and 3D ultrasound [22]. Many machine learning techniques are used to analyse medical pictures, including linear discriminant analysis (LDA), SVM, and DT. Machine learning techniques are being utilised to build low binary pattern descriptors that might be used to biological imagery. In order to analyse the specifics of an illness, medical pictures are examined using the neural network approach. Expert systems focused on medicine can also use machine learning in medical imaging [23]. One of the finest models for image analysis is the convolution neural network (CNN). It features a number of layers that might use convolution filters to change the input. Regarding categorization, there are two categories for medical photos. These are classifications of objects and images, respectively. Deep learning is used in image categorization to look into clinically relevant concerns so that the patient can receive therapy as soon as feasible. The primary goal of object categorization is to analyse more closely selected portions of the medical picture. Deep learning algorithms assist in identifying, classifying, and counting illness patterns utilising image processing in medical image analysis [23]. However, when it comes to medical imaging, this can include real pixel values, edge strengths, regional variations in pixel values, etc. The chosen subset of features must be able to provide the best and most accurate forecasts when features are chosen [24].Nonlinear classification issues include image recognition and the classification of biological time series. Very complex nonlinear functions cannot be implemented using the feature extraction and classification methods currently in use. Yet, by adding more layers and neurons to the network, DNN may create nonlinear functions. Several classifiers might be coupled with ensemble learning to perform complex decision-making processes. Several kernel functions are used to create nonlinear functions in both SVM and ensemble learning. The generation approach in the current situation is to employ the same learning algorithm afterwards. As a result, several settings are required, such as learning parameters and training samples. This concept has led to the development of several approaches, which may be categorised into four groups. Manipulating the training samples comes first. For this, wagging, arcing, cross-validated committees, and bagging-type techniques are required. The second method involves modifying the input characteristics. This may be accomplished using techniques like similarity-based feature space, input decimation, and random subspace. The third method involves changing the class labels. Examples of this include output coding and class switching. The fourth method is to add randomization to the algorithm for learning. A few methods utilised for this are the RF, randomised first-order inductive learner, and the backpropagation algorithm. If various starting weights are given to the same training data in a neural network, the subsequent classifiers utilising backpropagation will be highly distinct [25].

3.Machine Learning in Disease Prediction

Machine learning is increasingly being used in the diagnosis of medical conditions. This may be attributed mostly to advancements in illness categorization and identification systems, which are able to give information that supports medical professionals in the early detection of lethal diseases and so considerably raises patient survival rates. Basically, based on predictive modelling, a disease prediction system determines the user's condition from the symptoms they submit as input to the system. The technology evaluates the user's symptoms as input and returns a likelihood of the disease in the form of an Asian output. Disease Implementing the Classifier allows for prediction. Classifier estimates the likelihood of the illness. Accurate medical data analysis facilitates early illness identification as big data usage increases in the biomedical and healthcare industries [26].Cancer: The proper number of each type of cell exists in the human body. Cancer starts with sudden alterations in cell structure. The control and cell division of a given cell are determined by signals produced by that cell. Cells overproduce when these signals go bad, creating a mass known as a tumour. As thermography is non-invasive and nonionizing now, it is more trustworthy. The new technology has been providing effective and advantageous outcomes, which have elevated it above competing technologies. The presence of cancer cells may be determined from thermographic pictures using feature extraction and machine learning approaches. To extract features from pictures, approaches like scale invariant feature transform (SIFT) and accelerated robust feature (SURF) can be utilised. The characteristics might be further filtered using principal component analysis (PCA) to get better interpretations [27].Breast Cancer By putting the tumour into a category, breast cancer is identified. There are two types of tumours: benign and malignant. Malignant tumours are more dangerous than benign tumours, it should be highlighted. Yet, it is difficult for doctors to differentiate between these malignancies. As a result, machine learning algorithms are crucial because they can automatically learn from experiences and get better without explicit programming [28] and the feature extraction stage is crucial because it aids in separating benign from malignant tumours. Following that, segmentation is used to extract picture attributes including smoothness, coarseness, depth, and regularity [29].The pictures can be translated from the time domain to the frequency domain using discrete wavelet transformation (DWT). The approximation coefficient matrix, the horizontal detailed coefficient matrix, the vertical detailed coefficient matrix, and the diagonal detailed coefficient matrix are the four matrices that make up this wavelet decomposition. These are the values that the machine learning algorithms will utilise [30].Lung Cancer: Compared to MRI and X-ray data, computerised tomography (CT) results are less noisy. To produce the picture in the desired form with less noise and distortion, procedures including grayscale conversion, noise reduction, binarization, and segmentation are crucial. The average of RGB is used when converting to greyscale. Noise reduction is accomplished using the median filter. Segmentation finds the objects and the boundaries in the photos and strips away extraneous features. Features like area, perimeter, and eccentricity are taken into account during the feature extraction step [31] and it also recommended to use the machine learning algorithms such as convolution neural network- (CNN-) based deep learning methods could be used in detecting lung cancer cell with CT image [32] .Acute Lymphoblastic Leukaemia: Acute lymphoblastic leukaemia (ALL) is a kind of cancer in which a significant number of immature lymphocyte blood cells proliferate and negatively impact the development of other blood cells. A variety of machine learning methods have been employed to treat ALL. Leukaemia detection has been carried out using a variety of machine learning techniques, including KNN, SVM, NB, radial basis function network (RBFN), and multilayer perceptron (MLP). However, there are generally four steps in each of these methods: preprocessing, feature extraction, developing a classification model, and classifier assessment. The image will be cropped during the preprocessing step so that the region of interest (ROI) is clearly displayed, and the extraneous data is removed. The photos may be further processed to improve the picture by decreasing noise using the Gaussian blur smoothing approach. Concern is given to colour-based features, geometrical features, statistical features, the Haralick texture feature, picture moments, local binary pattern, and the existence of neighbouring cells throughout the feature extraction process [33].

Diabetes: Diabetes is a chronic condition that requires early diagnosis in order to receive the proper treatment. As the blood's sugar ratio rises, diabetes develops. Discriminant analysis (DA) is a process in which a set of equations derived from input characteristics determines the class label of an input. Prediction of type 2 diabetics was made possible by combining machine learning algorithms like Gaussian Naive Bayes (GNB), LR, KNN, CART, RFA, and SVM with elements from electronic medical records (EMRs), such as serum-glucose levels, body mass index (BMI), age, race, gender, creatinine level, and others [34]. As compared to other machine learning methods, predictions generated using neural networks were shown to be more accurate [35], It is important to note that both of the aforementioned neural network-based techniques have demonstrated diabetes prediction accuracy of close to 97% [36].Heart diseases: Blood tests are a common investigative technique in cardiology among the several precision medicine inquiry techniques. AGES is an unique precision medicine test that prevents ischemic heart disease by using other parameters in addition to blood testing. In precision medicine,

the focus is mostly on genetics, and several research projects are being conducted to identify the genetic origins of various diseases. Precision cardiology has particular interests in cardiac genetics, cardiac oncology, and ischemic heart disease. The genetics of many cardiovascular disorders is deeply ingrained. Hence, particularly for certain kinds of disorders, remedies utilising precision medicine are thought to be more effective. Long short-term memory (LSTM), NLP, SVM, CNN, and recurrent neural network (RNN) are a few machine learning approaches that might be applied effectively to construct exact clinical decision support systems (CDSSs) utilising deep learning [37] and the execution time is decreased and the classification accuracy is raised by implementing feature selection prior to the classification. Several feature selection algorithms exist. Popular feature selection techniques include Relief, Minimum Redundancy Maximal Relevance (mRMR), and Least Absolute Shrinkage and Selection Operator (LASSO) [38]Chronic Kidney Disease (CKD): kidney disease of the type known as CKD progressively impairs kidney function and results in renal failure. The primary machine learning classifiers employed in this domain were LR (Logistic Regression), DT (Decision Trees), and ANN (Artificial Neural Network). When compared to DT and LR on CKD diagnosis, the findings indicated that ANN performed far better [39].

4.Machine Learning Applications in Molecular Tests:

Gene expression levels and protein abundance may be measured in a range of materials, including blood, saliva, and tissue, using molecular tests to detect genetic alterations. By identifying complicated sets of biomarkers linked diverse disease states, machine learning has the ability to expand the usefulness of these data by identifying treatment options and patient outcomes. Using DNA methylation is one example from cancer biology [40].By measurement of mRNA in the blood, machine learning has been utilised to identify people who are sleep deprived, revealing how sleep insufficiencies adversely influence health [41].

5. Machine learning to Predict a Patient's No. in healthcare sector

Patients who skip their appointments cause a number of issues for the healthcare industry. Understanding the patient's profile and predicting possible absences are the key challenges, Salazar, L e. at.al.2021 as proposed a machine learning technique based model that contribute to a patient's no-show and develop a prediction model able to identify whether the patient will attend their scheduled appointment or not, where the most popular methods for predicting absent attendance, a binary classification problem, are DTs and LR. There are numerous alternative algorithms that target classification difficulties, such as Naive Bayes and Support Vector Machines (SVMs), however they are not appropriate for this study due to the features of the dataset. For instance, SVMs are efficient in high-dimensional spaces but have the disadvantage of not immediately providing probability estimates, necessitating the use of pricey k-fold cross-validation to calculate results. The following classification techniques are taken into account in this work: Random Forest classifier, Decision tree classifier, and Logistic Regression classifier, The supervised learning models in this research were created using the open-source scikit-learn (SKLearn) package (https://scikit-learn.org/stable/, viewed on November 25, 2021). The Python programming language is used by the SKLearn package to create several ML algorithms and model performance analysis tools. In the context of this study, a classification model is defined as an algorithm that is implemented by a predefined function from the SKLearn library that accepts a certain set of potential parameters. this work can help the improvement of solutions to the public health system with the help of machine learning techniques and the availability of additional information on patients and medical appointments may help to improve the ability of the model to learn new behaviours [42].6.Financial Fraud Detection in Healthcare Using Machine Learning One of the well-known industries where a lot of data may be gathered not only in terms of health but also in terms of finances. Due to the widespread use of credit cards in the healthcare industry and the ongoing development of electronic payments, credit card fraud monitoring has been difficult financially for the various service providers. Hence, continuous enhancement is necessary for the system for detecting frauds.The technique of examining cardholder transaction activity to determine if a transaction is legitimate is known as fraud detection. Credit card fraud is the unauthorised use of credit card information to complete a purchase. When doing a transaction in person, a credit card is required; however, when conducting one online or over the phone, other methods are used to gather the card's number, verification number, and expiration date. Many deep learning and machine learning classifiers, including Logistic Regression, Naive Bayes, Decision Tree, KNN, and the sequential model, were used in the experiment along with a dataset. Before creating the classifier, all of these algorithms go through many steps such as data collecting, data preprocessing, data analysis, data training with various classifiers, and later data testing. Preprocessing involves converting all of the data into a format that can be used. Using two separate data distribution sets, the hybrid under sampling (negative class) and oversampling (positive class) procedures were applied. Pre-processed data are input into the classifier algorithm during the training phase. The accuracy of identifying credit card fraud is afterwards determined by evaluating the test data [43].

6.Machine Learning Solution for the Smart Healthcare Sector's Bed Occupancy Problem

Asian nations and less developed nations with large populations are dealing with problems linked to health care services. One of these nations is India, where the health care system is having trouble accommodating patients due to the nation's high population density. To overcome this issue S. Gochhait e. at.al.2021 as proposed framework helps hospitals to enhance the decision process for bed occupancy for patients concerning their departments and their diseases based on machine learning technique.

a calculus-based solution to the patient confirmation booking problem. Something will be done to address the programming problem. This study provided a few extra concepts that aid in the scientific articulation of the problem. Model will then go on to explain the decision-making criteria, the intended work, and the imperatives. Notably, the numerical model only takes into account the most well-known scenario, in which the gender of the room is determined by the first persisting’s sexual orientation, with a SAP A collaborative enterprise modelling tool called Power Designer was created by Sybase, which is now owned by SAP & They developed the application for the administration to deal with the bed problem and manage bed allocation using the Apache, SAP, SQL, and JS programming languages [44].

7. Machine Learning Algorithm to Identify Patients at Risk for Cardiac Syndrome

According to a World Health Organization (WHO) report, heart attacks are regarded as a serious illness everywhere. Mental stress, a heavy workload, and other factors are among the causes of heart disease. Healthcare data is quite helpful for finding the underlying pattern for the decision-making process for such cardiac conditions [45]. 13 risk factors and useful variables are used in the traditional approaches for classifying heart disorders. The newly developed methodology for detecting cardiac problems uses a unique hybrid computational modelling technique. The artificial neural networks used in this study represent a variety of current methodologies for determining cardiovascular risks (ANN). This ANN-based methodology commonly assumed that heart failure diagnoses would be associated with the same risk factors. In this research, a hybridised methodology of K-Nearest Neighbour clustering and Spiral optimization in the classification of the cardiovascular hazards is used to assess the strategy of an effective recognition method for analysing the failure associated to heart illnesses. Support vector machine (SVM), convolutional neural network, and other data mining approaches are paired with the hybridised KNN methodology [46].

8.Machine Learning Methods for Mental Illness Diagnosis

The exponential growth of data has resulted from the quick development of computer and internet technologies in every industry, including the educational and industrial sectors. Examples include a sizable amount of student data from the OECD and customer-related data from Walmart and Facebook [47-49] and it's not only limited to the prediction of disease in health Domaine but also till diagnosis the of mental ill health.

Post-Traumatic Stress Disorder

post-traumatic stress disorder was diagnosed using Stochastic GBM (Gradient Boosting Machine) along with questionnaires in semi-structured interviews data [50]. A ML method was used by Augsburger and Elbert [32] to forecast risk-taking behaviour in refugees who have undergone traumatic events. 56 cases were interviewed, and data sets included the Balloon Analog Risk Task (BART) and questionnaires. Stochastic GBM was the ML method utilised in this experiment, while R was used for analysis. This approach was superior to traditional approaches because it allowed for the simultaneous testing of several variables in a small number of samples. The use of a short dataset to build a model in this experiment was a drawback. [51] and SVM, Target information equivalence algorithm utilising Event characteristics, emergency department records, early symptoms [52].

Schizophrenia

Magnetic resonance imaging (MRI) scans were used by Schnack et al. [53] to categorise schizophrenia patients, bipolar disorder patients, and healthy controls. SVM was used to create three models using 66 MRI scans of schizophrenia patients, 66 MRI scans of people with bipolar illness, and 66 MRI scans of healthy individuals.

The model's average discrimination accuracy between people with schizophrenia and healthy controls was 90%, while its average accuracy in differentiating between those with schizophrenia and bipolar illness was 88%. Compared to previous models, the model used to discriminate between patients with bipolar illness and healthy controls performed less accurately, correctly categorising 53% of the bipolar disorder patients and 67% of the healthy participants.

A hybrid ML technique was utilised by Yang et al. [54] to distinguish between schizophrenia patients and healthy controls. Twenty people with schizophrenia and twenty healthy controls provided the data. In this study, the single nucleotide polymorphism (SNP) and functional magnetic resonance imaging (fMRI) data were used to develop the SVM algorithm. The accuracy for the strategy that utilised SNPs to create an SVM ensemble (SVME) was 0.74, and it was 0.82 for the method that created another SVME using the voxels from the fMRI map. Moreover, the accuracy of the approach used to develop an SVM classifier utilising fMRI activation components discovered by independent component analysis was 0.83, and the accuracy of the method used to integrate the aforementioned three models was 0.87.

Depression

A ML algorithm was created by Chekroud et al. [55] to forecast clinical remission following a 12-week citalopram therapy. 1949 depressed patients from level 1 of the Sequenced Therapy Options to Relieve Depression made up the data set. Out of 164 patient-reportable variables, 25 were chosen to produce the best accurate result. GBM was used for prediction because, when formed, it integrates a number of imperfectly predictive models. 64.6% accuracy was reached using GBM. In a cross validation test, the escitalopram treatment group (n=151) of Combining Medicines to Improve Depression Outcomes (COMED) shown accuracy of 59.6%, while the combined escitalopram-bupropion treatment group (n=134) of COMED demonstrated accuracy of 59.7%.

Autism Spectrum Disorders

In order to categorise newborns into greater risk groups for autism spectrum disorder (ASD) or control, Bosl et al. [56] employed ML algorithms. Data set comprised of 79 babies, 46 of whom were at high risk for ASD and 33 of whom were healthy controls. Modified multiscale entropy (mMSE) was estimated based on resting state EEG data. To find the best classifiers for their data, the k-nearest neighbours (k-NN), SVM, and Naive Bayes algorithms were utilised for classification. Both the k-NN and SVM offered 0.77 accuracy for babies that were 9 months old, and the outcome was statistically significant. The maximum statistically significant result Naive Bayes produced was accuracy of 0.80, which was attained at the age of 18 months.

By using ML algorithms, Maenner et al. [57] identified autism spectrum disorder (ASD). To pick and classify features, Random Forest was used. The words and phrases from the 5,396 assessments of 1,162 children from the 2008 Georgia Autism and Developmental Disabilities Monitoring (ADDM) site served as the project's data set. The classifier was evaluated using a data set consisting of 9,811 evaluations of 1,450 children from the Georgia ADDM surveillance in 2010. For prediction, Random Forest achieved an accuracy of 86.5%.

5.Conclusion:

The healthcare sector is a significant and vital industry because it provides assistance to numerous individuals. Moreover, it supports the regional economy. To enhance the quality of life for patients worldwide, numerous business and research endeavors are dedicated to integrating machine learning technology into the healthcare sector. Notably, recent advancements underscore this progress contributing a lot in revolutionizing the healthcare sector. The benefits offered by these technologies in health care sector are endless starting from Screening and daily fitness monitoring, diagnostic services in gastroenterology, pathology, and radiology, as well as support for clinical decision-making and palliative care. Nevertheless, the widespread use of ML in healthcare confronts significant obstacles include higher installation and maintenance costs, potentially harmful medical mistakes, holes in the ethical standards governing AI, unemployment, and a reduced ability to train human workers. In short, AI/ML may be extremely important in addressing the difficulties with complexity and the explosion of data in the healthcare system. All in all, ML is a component of contemporary healthcare, and its further use is contingent on thoroughly resolving pertinent issues.

CONCLUSION

Concept extraction using machine learning techniques is a powerful and versatile approach to automatically identify and extract specific entities or concepts from unstructured text data. It has numerous applications in various fields, including information retrieval, sentiment analysis, question-answering systems, chatbots, and more.

In conclusion, concept extraction using machine learning techniques offers several key advantages:

1. Automation: Machine learning models enable the automation of concept extraction tasks, saving time and effort compared to manual annotation or rule-based methods.

2. Scalability: Once trained, machine learning models can efficiently process large volumes of text data in real-time, making them suitable for handling big data applications.

3. Generalization: With appropriate training and feature representations, machine learning models can generalize well to new, unseen data, allowing them to work effectively across various domains and sources.

4. Flexibility: Machine learning models can be adapted and fine-tuned for specific contexts or domains, providing flexibility to cater to different requirements.

5.Incremental Improvement: Ongoing research and advancements in machine learning continually enhance the performance of concept extraction models, leading to better accuracy and robustness.

However, it's essential to be aware of the challenges involved in concept extraction using machine learning techniques, such as the need for high-quality labeled data, handling language ambiguity, and addressing out-of-vocabulary concepts. Additionally, the interpretability of deep learning models can be a concern in critical applications.

To make the most of concept extraction with machine learning, it is crucial to choose appropriate algorithms, feature representations, and model architectures based on the specific use case and data characteristics. Regular evaluation and refinement of the models are necessary to ensure their effectiveness and relevance over time.

As research in NLP and machine learning continues to progress, concept extraction is likely to see further improvements, opening up new possibilities for understanding and utilizing unstructured text data to extract valuable insights and knowledge.
Natural language processing (NLP) task known as idea extraction is locating and extracting particular entities or concepts from unstructured text input. These entities could be names of people, businesses, or locations, or they could be dates or any other pertinent data. The use of machine learning techniques for idea extraction is common because of their propensity to manage the complexity of natural language and produce precise predictions.

Here is a general summary of how concept extraction can be accomplished using machine learning techniques:

1. Data Collection and Preprocessing: The first stage is to compile a collection of text data from which concepts are to be extracted. Websites, documents, and social media platforms are just a few of the places where you can find this information. Data must be preprocessed after collection, which includes operations like tokenization, getting rid of stop words, and handling special characters.

2. Annotated Data: You'll need annotated data for supervised learning, where each textual segment is labelled with the concepts you intend to extract. The machine learning models need this data to be trained.

3. Feature Extraction: You must describe the text data as numerical features in order to train a machine learning model. The use of Word2Vec or GloVe word embeddings, as well as more sophisticated contextual embeddings like BERT (Bidirectional Encoder Representations from Transformers), are examples of common methodologies.

4. Model selection: A number of machine learning algorithms can be used to extract concepts from data. Typical options include:

- Named Entity Recognition (NER) models: These models are made especially for extracting entities, such as names of people, businesses, places, etc.

NER projects frequently use the sequence labelling algorithm known as Conditional Random Fields (CRF).
Support Vector Machines (SVM) and Decision Trees are two classifiers that, when used with the right feature representations, might be helpful for concept extraction.

- Deep Learning models: For NLP tasks, such as idea extraction, techniques like recurrent neural networks (RNNs), LSTM (Long Short-Term Memory), and transformer-based models (e.g., BERT) have shown promising results.

Divide your dataset into training and validation sets for your model. Utilise the annotated data and the extracted features to train the chosen model. The model will develop the ability to spot textual patterns that relate to the concepts you want to extract.

6. Evaluation: Evaluate your model's performance using the proper measures, such as precision, recall, F1-score, or accuracy. This stage enables you to assess the performance of your model and determine whether it achieves the appropriate degree of accuracy.

7. Deployment: Once your model is effective, you can use it to process fresh, previously unexplored text input and instantly extract concepts from it.

It's crucial to remember that concept extraction performance significantly depends on the quality and amount of annotated training data as well as the selection of appropriate features and algorithms. Additionally, in some circumstances, rule-based systems can be used in conjunction with machine learning methods to complete idea extraction tasks.

REFERENCES:

[1] Raghupathi, V. & Raghupathi, W. (2014). An Unstructured Information Management Architecture Approach to Text Analytics of Cancer Blogs. International Journal of Healthcare Information Systems and Informatics (IJHISI), 9(2), 16-33. http://doi.org/10.4018/ijhisi.2014040102.

[2] Panagiota Galetsi and Korina Katsaliaki and Sameer Kumar (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. International Journal of Information Management,50, 206-216

[3] Vivek P. Chavda,Kaushika Patel,Sachin Patel,Vasso Apostolopoulos(2023). Artificial Intelligence and Machine Learning in Healthcare Sector.

[4] Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health. 2018 Dec;8(2):020303. doi: 10.7189/jogh.08.020303. PMID: 30405904; PMCID: PMC6199467.

[5] Srinivasan, K., Raman, K., Chen, J., Bendersky, M., & Najork, M. (2021, July). Wit: Wikipedia-based image text dataset for multimodal multilingual machine learning. In Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval(pp. 2443-2449).

[6] Linthicum KP, Schafer KM, Ribeiro JD. Machine learning in suicide science: Applications and ethics. Behav Sci Law. 2019;37(3):214-222.

[7] Alpaydin E. Introduction to Machine Learning. MIT press; 2009.

[8] Arwinder Dhillon and Ashima Singh (2019), “Machine Learning in Healthcare Data Analysis: A Survey”. Journal of Biology and Today's World.

[9] M. Ferdous, J. Debnath and N. R. Chakraborty, "Machine Learning Algorithms in Healthcare: A Literature Survey," 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), Kharagpur, India, 2020, pp. 1-6, doi: 10.1109/ICCCNT49239.2020.9225642.

[10] R. Bhardwaj, Sethi A. Adhiraaj and R. Nambiar, "Big data in genomics: An overview", IEE Big Data Conference, 2014.

[11] B. Kayyali, D. Knott and S. Kuiken, The Big Data Revolution in US Health Care: Accelerating value and innovation, McKinsey & Company.

[12] K. Applod, five big industry changes to watch in 2016.

[13] S. Dolley, Big Data Solution to Harnessing Unstructured Data in Healthcare.

[14] Tahmassebi A, Wengert GJ, Helbich TH, Bago-Horvath Z, Alaei S, Bartsch R, et al. Impact of machine learning with multiparametric magnetic resonance imaging of the breast for early prediction of response to neoadjuvant chemotheraphy and survival outcomes in breast cancer patients. Invest Radiol. 2019;54(2):110-117.

[15] Dagli Y, Choksi S, Roy S. Prediction of two year survival among patients of non-small cell lung cancer. ICCMIA. 2019; 31:169-177.

[16] Kayal CK, Bagchi S, Dhar D, Maitra T, Chatterjee S. Hepatocellular carcinoma survival prediction using deep neural network. Proceedings of International Ethical Hacking Conference. 2019;349-358.

[17] Wong A, Young AT, Liang AS, Gonzales R, Douglas VC, Hadley D. Development and validation of an electronic health record–based machine learning model to estimate delirium risk in newly hospitalized patients without known cognitive impairment. JAMA Netw Open. 2018;1(4):e181018.

[18] Rahimian F, Salimi-Khorshidi G, Payberah AH, Tran J, Solares RA, Raimondi F, et al. Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records. PLoS Med. 2018;15(11): e1002695.

[19] Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs and machine learning for heart failure survival analysis. Stud Health Technol Inform. 2015; 216:240.

[20] Zheng T, Xie W, Xu L, He X, Zhang Y, You M, et al. A machine learning-based framework to identify type 2 diabetes through electronic health records. Int J Med Inform. 2017; 97:120-127.

[21] Corey KM, Kashyap S, Lorenzi E, Lagoo-Deenadayalan SA, Heller K, Whalen K, et al. Development and validation of machine learning models to identify high-risk surgical patients using automatically curated electronic health reecord data (Pythia): A retrospective, single-site study. PLoS Med. 2018;15(11):e1002701.

[22] R. Pratt, J. Deprest, T. Vercauteren, S. Ourselin, and A. L. David, "Computer-assisted surgical planning and intraoperative guidance in fetal surgery: a systematic review," Prenatal Diagnosis, vol. 35, no. 12, pp. 1159–1166, 2015.

[23] J. Latif, C. Xiao, A. Imran, and S. Tu, "Medical imaging using machine learning and deep learning algorithms: a review," in Proceedings of the 2nd International Conference on Computing, Mathematics and Engineering Technologies, Sukkur, Pakistan, March 2019.

[24] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, "Machine learning for medical imaging," RadioGraphics, vol. 37, no. 2, pp. 505–515, 2017

[25] L.-p. Jin and J. Dong, "Ensemble deep learning for biomedical time series classification," Computational Intelligence and Neuroscience, vol. 2016, Article ID 6212684, 2016.

[26] Pingale, K., Surwase, S., Kulkarni, V., Sarage, S., & Karve, A. (2019). Disease prediction using machine learning. International Research Journal of Engineering and Technology (IRJET), 6, 831-833.

[27] V. Mishra, Y. Singh, and S. Kumar Rath, "Breast cancer detection from thermograms using feature extraction and machine learning techniques," in Proceedings of the IEEE 5th International Conference for Convergence in Technology, Bombay, India, March 2019.

[28] A. Bharat, N. Pooja, and R. A. Reddy, "Using machine learning algorithms for breast cancer risk prediction and diagnosis," in Proceedings of the 3rd International Conference on Circuits, Control, Communication and Computing, Bangalore, India, July 2018.

[29] H. Dhahri, E. Al Maghayreh, A. Mahmood, W. Elkilani, and M. Faisal Nagi, "Automated breast cancer diagnosis based on machine learning algorithms," Journal of Healthcare Engineering, vol. 2019, Article ID 4253641, 2019.

[30] M. R. Al-Hadidi, A. Alarabeyyat, and M. Alhanahnah, "Breast cancer detection using K-nearest neighbor machine learning algorithm," in Proceedings of the 9th International Conference on Developments in eSystems Engineering, Liverpool, UK, September 2016.

[31] W. Rahane, H. Dalvi, Y. Magar, A. Kalane, and S. Jondhale, "Lung cancer detection using image processing and machine learning healthcare," in Proceedings of the International Conference on Current Trends towards Converging Technologies, Coimbatore, India, March 2018.

[32] Q. Wu and W. Zhao, "Small-cell lung cancer detection using a supervised machine learning algorithm," in Proceedings of the International Symposium on Computer Science and Intelligent Controls, Budapest, Hungary, October 2017.

[33] S. Mandal and V. Daivajna, "Machine learning based system for automatic detection of leukemia cancer cell," in Proceedings of the IEEE 16th India Council International Conference, Rajkot, India, December 2019.

[34] S. Mani, Y. Chen, T. Elasy, W. Clayton, and J. Denny, "Type 2 diabetes risk forecasting from EMR data using machine learning," pp. 606–615.

[35] S. Priya and R. Rajalaxmi, "An improved data mining model to predict the occurrence of type-2 diabetes using neural network," in Proceedings of the International Conference on Recent Trends in Computational Methods, Communication and Controls, Kochi, India, August 2012.

[36] S. Ayon and M. Islam, "Diabetes prediction: a deep learning approach," International Journal of Information Engineering and Electronic Business, vol. 7, no. 6, pp. 21–27, 2019.

[37] S. Niazi, H. A. Khattak, Z. Ameer, M. Afzal, and W. A. Khan, "Cardiovascular care in the era of machine learning enabled personalized medicine," in Proceedings of the International Conference on Information Networking, Barcelona, Spain, April 2020.

[38] A. U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," Mobile Information Systems, vol. 2018, Article ID 3860146, 2018.

[39] Y. Amirgaliyev, S. Shamiluulu, and A. Serek, "Analysis of chronic kidney disease dataset by applying machine learning methods," in Proceedings of the IEEE 12th International Conference on Application of Information and Communication Technologies, Almaty, Kazakhstan, October 2018.

[40] Kang, S., Li, Q., Chen, Q., Zhou, Y., Park, S., Lee, G., ... & Zhou, X. J. (2017). CancerLocator: non-invasive cancer diagnosis and tissue-of-origin prediction using methylation profiles of cell-free DNA. Genome biology,18(1), 1-12.

[41] Laing, E. E., Möller-Levet, C. S., Dijk, D. J., & Archer, S. N. (2019). Identifying and validating blood mRNA biomarkers for acute and chronic insufficient sleep in humans: a machine learning approach.Sleep,42(1), zsy186.

[42] Salazar, L. H. A., Leithardt, V. R. Q., Parreira, W. D., da Rocha Fernandes, A. M., Barbosa, J. L. V., & Correia, S. D. (2021). Application of Machine Learning Techniques to Predict a Patient's No-Show in the Healthcare Sector. Future Internet,14(1), 3. MDPI AG. Retrieved from http://dx.doi.org/10.3390/fi14010003.

[43] bolfazl Mehbodniya, Izhar Alam, Sagar Pande, Rahul Neware, Kantilal Pitambar Rane, Mohammad Shabaz, Mangena Venu Madhavan, "Financial Fraud Detection in Healthcare Using Machine Learning and Deep Learning Techniques",Security and Communication Networks, vol. 2021, Article ID 9293877, 8 pages, 2021. https://doi.org/10.1155/2021/9293877.

[44] Gochhait, S., Butt, S.A., De-La-Hoz-Franco, E.et al.A Machine Learning Solution for Bed Occupancy Issue for Smart Healthcare Sector.Aut. Control Comp. Sci. 55, 546–556 (2021). https://doi.org/10.3103/S0146411621060043.

[45] Dwivedi, A. K. (2018). Performance evaluation of different machine learning techniques for prediction of heart disease.Neural Computing and Applications, 29, 685-693.

[46] Zubar, A. H., & Balamurugan, R. (2020). Green computing process and its optimization using machine learning algorithm in healthcare sector. Mobile Networks and Applications, 25, 1307-1318.

[47] Ozgur C, Kleckner M, Li Y. Selection of Statistical Software for Solving Big Data Problems: A Guide for Businesses, Students, and Universities.SAGE Open; 2015. pp. 1–12.

[48] OECD What is PISA? Available at: http://www.oecd.org/pisa/aboutpisa/. Accessed Aug 21, 2018.

[49] McAfee A, Brynjolfsson E. Big data: the management revolution.Harv Bus Rev.2012; 90:60–66.

[50] Augsburger M, Elbert T. When do traumatic experiences alter risktaking behavior? A machine learning analysis of reports from refugees.PLoS ONE.2017;12:e0177617.

[51] Galatzer-Levy IR, Ma S, Statnikov A, Yehuda R, Shalev AY. Utilization of machine learning for prediction of post-traumatic stress: a re-examination of cortisol in the prediction and pathways to non-remitting PTSD.Transl

[52] Psychiatry.2017;7:e0.[52]. Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia.Front Hum Neurosci.2010;4:192.

[53] Schnack HG, Nieuwenhuis M, van Haren NE, Abramovic L, Scheewe TW, Brouwer RM, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects.Neuroimage.2014;84:299–306.

[54] Yang H, Liu J, Sui J, Pearlson G, Calhoun VD. A hybrid machine learning method for fusing fMRI and genetic data: combining both improves classification of schizophrenia. Front Hum Neurosci.2010;4:192.

[55] Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach.Lancet Psychiatry.2016;3:243–250.

[56]. Bosl W, Tierney A, Tager-Flusberg H, Nelson C. EEG complexity as a biomarker for autism spectrum disorder risk.BMC Med.2011;9:18.

.