BMS COLLEGE OF ENGINEERING

(Autonomous College under VTU)
Bull Temple Road, Basavanagudi, Bangalore - 560019

A report on

Seminar Based on Summer/Winter Internship

"Machine Learning in HealthServices"

Submitted in partial fulfillment of the requirements for the award of degree

BACHELOR OF ENGINEERING

IN

INFORMATION SCIENCE AND ENGINEERING

By

Anupam Singh(IBM19IS027)

Ayush Malik(IBM19IS036)

Chaitali Shekar(IBM19IS039)

Under the guidance of

Mrs. Sindhu K
Assistant Professor

Department of Information Science and Engineering
2020-21

BMS COLLEGE OF ENGINEERING

(Autonomous College under VTU)
Bull Temple Road, Basavanagudi, Bangalore – 560019

C E R T I F I C A T E

Certified that the Technical Seminar has been successfully presented at **BMS College of Engineering** by **Anupam Singh,Ayush Malik,Chaitali Shekar** bearing USN:**1BM19IS027,1BM19IS036,1BM19IS039** in partial fulfillment of the requirements for the IV Semester degree in **Bachelor of Engineering in Information Science & Engineering** of **Visvesvaraya Technological University, Belgaum** as a part of for the **course Seminar Based on Summer/Winter Internship Course Code - 19IS4SRSMI** during academic year 2020-2021

Faculty Name : Mrs. Sindhu K

Designation : Assistant Professor

Signature :

Introduction

Machine learning is a technique in which computers are programmed to learn the information without the intervention of humans.Machine learning algorithms are provided with data and then the computer learns from that data by finding patterns in them.Factors like growing volumes of data and varieties of available data,cheaper and powerful computational processing and cheap data storage.All these factors makes it possible to make machine learning models that can be trained on large large amount of data and produce accurate results.Machine learning in healthcare is used widely in helping patients and clinicians in many different ways.Artificial Neural Networks are the most widely used machine learning algorithms in the healthcare industry.The neurons are linked to each other and are trained on a the dataset to produce meaningful results.

Abstract

As a rapidly growing field there is a wide range of applications of machine learning in the healthcare industry such as personnel management,insurance policies,regulatory affairs and much more.In our report we have discussed about the application of machine learning in treating and identifying of diseases,production of drugs and vaccines and research done related to them.One of those applications include use of ML in medical images such as in CXR(chest x ray scan).The results obtained from medical imaging is used by doctors for interpretation of disease and diagnose them accordingly.Other application which we have mentioned in our report is using natural language processing to predict the diseases from the electronic media reports by extracting structured and unstructured data from it.We have discussed a wide variety of research papers about exciting and future looking applications of ML techniques and their applications in the healthcare space.

Literature Review:

Majid Nour,Kemal Polat[1]:When blood pressure is high it is termed as hypertension.There are two categories of blood pressure systolic blood pressure (SBP) and diastolic blood pressure (DBP).When SBP is greater than or equal to 140mm of Hg and DBP is greater than or equal to 90mm of Hg we call it as hypertension.We have used a machine learning algorithm for classification of the hypertension.To train the model PPG-BP(photo-plethysmorgraphy blood pressure) database has been used.The dataset comprises of the following features:sex age,height,weight,systolic blood pressure,diastolic blood pressure,heart rate and BMI.The output has been categorized into 4 classes normal,prehypertension,stage1 hypertension and stage2 hypertension on the basis of PPG signals.Also for all the features Pearson correlation coefficient has been calculated.C4.5 decision tree classifier,random forest,linear discriminant analysis and linear support vector machine have been used to determine the output.For the purpose of training and testing 5-fold cross validation method was used.In which all data are divided into 5 equal parts.Each part is further divided into 5 parts out of which 4 parts are used for training and 1 is used for testing.Confusion matrix and area under curve for ROC(receiver operating characteristic) curves are used to evaluate the performance of the model.And it has been observed that C4.5 decision tree gives the best accuracy.

Rajesh, T Maneesha et al.[2]: Heart disease is one of the most common disease. Different attributes are used and algorithms like Naive Bayes and Decision trees are implemented to predict heart diseases. The literature surveys from different papers are conducted focusing on different features of the medical assistance of ML required in the field of research of Cardiovascular medical sciences. Some papers dissect mining procedures which are required for medicinal information mining which is carried out for finding locally most visited illnesses. Other papers focus on remote health monitoring using First month and Baseline intervention data. Some state the use of the ID3 algorithm in TV and phones. This proposed method claims to reduce the death rate by the help of information mining. Some use expectation framework for heart infection utilizing Learning vector Quantization neural system calculation.As far as methodology

is concerned, the first step begins with elimination of unwanted data like noise. Then comes transformation, integration and reduction of complex data.

The ID3 algorithm is used to build the decision trees. It separates a dataset into fewer and fewer sub-sets while in the meantime a related decision tree is incrementally created. For making the tree, Yes and No nodes are constructed depending upon the certainty of the equation. They are classified accordingly and calculated recursively.

The Naive-Bayes classification is more accurate considering each element individually and is not hard to run. It goes by the formula Posterior Probability = product of likelihood and class prior priority divided by the predictor prior probability.What we can conclude is that Decision trees gives us a close solution but not an accurate one. On the other hand Naive Bayes theorem is more accurate if the input data is cleaned first and then sent into it. To get even more accuracy, we can combine Naive Bayes theorem with k-means.

Kenji Suzuki [3]: We can use machine learning for analysis of medical images and interpret the disease from those images.With improvement in computational power pixel based ML has emerged for the purpose of medical image processing which picks up the pixel values from the images instead of calculation features using segmentation method(process of dividing image into regions with similar properties such as color,texture,gray level).

Pixel based Machine Learning(PML) can be classified into three classes (1)Neural Filters(2)Convolution Neural Networks(CNN)(3)Massive Training Artificial Neural Networks(MTANNs)

1)Neural Filters are supervised neural networks where input are the image pixel values and its output is a single value.A backpropagation algorithm is used for training purpose.

2)An MTANN is an extension of neural filters.It consists of a neural network regression model and support vector machine regression model which can work directly on pixels.It uses a linear function in the output layers and sigmoid function in the hidden layers as an activation function.The input are the pixel values and output is a continuous scalar value.

O(x,y,z)=ML[I(x-i,y-j,z-k)] where O is the output,ML is the machine learning model into which image I is given an input.

3)Convolution Neural Network consists of one input layer,several hidden layers and one output layer.The hidden layers are interconnected by weights and sigmoid function is the

activation function used here.The output from a layer is computed with the weights and then sent as input to the next layer.

Applications:

1.Noise Reduction in x ray:Neural Filters are used to reduce noise in X Ray images.It is done by training the model with artificially created noisy x ray images.Noise is added to x ray image which is fed as input and the original x ray(without noise)acts as output.

2.Bone Separation from Soft Tissue in chest radiographs using MTANNs:Detection of lung cancer is a tough job because the affected tissues are covered by the overlying bones due to which they are not visible in CXRs.To handle this problem MTANN are used,they are trained with soft-tissue and bone images obtained from dual-energy radiography system.CXR along with the soft-tissue and bone images obtained from dual-energy radiography system acts as input and soft tissue image is obtained from the model.

Keyang Xu et al.[4]: The International Classification of Diseases(ICD) endorsed by the world health organization (WHO) is a medical classification of codes for diagnoses and procedures.We can make a machine learning model for to predict ICD from the Electronic Media Report(EMRs) by extracting unstructured text,semi-structured text and structured tabular data.(1)Unstructured data includes physicians notes.(2)Semi-Structured Text includes structured phrases and unstructured sentences that describe diagnosis written by physicians.(3)Structured Tabular Data includes prescription and clinical measurements.Dataset used for this study is MIMIC-III which contains details about 47000 patients from 58000 hospitals.

Model1-To make predictions on the basis of Unstructured Data we first perform tokenization on the text and then a convolution model is used to select a set of ICD codes which are most relevant to the text.

Model2-To deal with semi-structured data we use a deep learning model to find similarities between the diagnosis written by physicians and ICD code descriptions.The diagnosis description is mapped to a vector space and the ICD codes are given ranking on the basis of their distance from the diagnosis vector.

Model3-To make predictions using tabular data we convert the numerical features to binary features and then a decision tree is used to classify the ICD codes.

Finally we integrate the result of the three models to predict the output on the basis of the following formula:

$P=\sum_{k=1}\alpha_k P^k$

Where P is final probability of a particular ICD code and P^k is the probability predicted by the kth model for a particular ICD code and α_k is the weight parameter for the kth model.

The ICD code with maximum probability is the output.

Quan Zou et al.[5]: Diabetes is a very common chronic disease.When the blood glucose is higher than the normal levels because of defective insulin secretion it is defined as diabetes.We will use Decision Tree,Random Forest and Neural Networks as the classifiers.The dataset used for training of model is obtained from a hospital from Luzhou,China and another dataset is Pima Indians diabetes data.For the purpose of model validation k-fold cross and hold-out method have been used.In k-fold cross validation the dataset is divided into k parts called as folds and then from each fold k-1 data is used for training purpose and 1 is used for test purpose and this process is repeated k times.In hold-out method the dataset is divided into two parts,one part is used for training purpose and other part is used for testing purpose.Principal Component Analysis(PCA) and Minimum Redundancy Maximum Relevance(mRMR) are used to remove the redundant features from the dataset.We have used sensitivity(SN),specificity(SP),accuracy(ACC) and Matthews correlation coefficient(MCC) to measure the effectiveness of the algorithms.For Luzhou data set the maximum accuracy was achieved by Random forest of 73.95% and even for Pima Indians Random Forest achieved the maximum accuracy of 71.44%.

Konstantina Kourou et al.[6]: This paper is an introduction to what is cancer to implementations of different machine learning techniques that can be used to predict cancer in a person considering various techniques, algorithms and tools provided by ML.Some of the techniques used are Artificial Neural Networks, Bayesian Networks, Support Vector Machines and Decision trees which will be found everywhere where ML is involved.The main goals of using this technique are 3: susceptibility, recurrence and survival prediction and given data shows 15-20% of improvement of results. The ML technique involves two stages - estimation of unknown dependencies and using these dependencies to predict new outputs.The approaches used for breaking up and using the data are - dimensionality reduction, feature selection and extraction. A model should fit the training set well and its effectiveness is based mainly upon the sensitivity, accuracy

and area under the curve. The accuracy is measured by the Decision boundary i.e. the no. of accurate results which followed the predictions or not. Those which do not, are misclassified.The Decision Trees are the earliest methods of ML to be used whereas SVMs the most recent. These two are the most common techniques used currently. They use predictions whereas BN classifiers work on calculated estimated values. For predictions to be made, many features like family history, background, age, diet, weight, pre existing diseases and many other factors are seen.The drawbacks are mostly in the accuracy of collecting the samples and validation of the classified results. It is conclusive that ANN and SVMs are most used and most convenient but SVMs predict the results more accurately. Decision Trees being the easiest to use.

Zhongheng Zhan et al.[7]: When the amount of fluids increases it can be harmful and leads to acute kidney failure(AKI).So in this research paper a study has been conducted to prepare a model to differentiate between two types of AKI,(1)Volume Responsive AKI and (2)Volume Unresponsive AKI.Medical Information Mart for Intensive Care(MIMIC-III) dataset was used in which patients with urine output<0.5ml/kg/h for the first 6h after ICU admission were considered to have AKI.The model used for this study were XGBoost and logistic regression.The whole dataset was split in the ratio of 3:1 for training purpose.The hyperparameters used for XGBoost are as follows:learning rate=0.04,minimum loss reduction=10,maximum tree depth = 9, subsample = 0.6, and number of trees = 300.Model performance was calculated using area under curve of ROC and confidence interval(CI).XGBoost algorithm performed better than logistic regression(AU-ROC, 0.860; 95% CI, 0.842 to 0.878 vs. 0.728;95% CI 0.703 to 0.753, respectively).

Pravin Shinde,Prof.Sanjay Jadhav[8]: All the features like extracting words from Medline, Six algorithms in order to improve precision of output NB,CNB,Adaptive Boost,SVM,Decision Tree,BaseLine are used. Uses natural language processing to get information on prevention,cure,treatment about any diseases from database.There are two sets of extraction tasks that are capable of identifying and disseminating healthcare

information,first being, all information regarding diseases and treatments for which Natural Language Processing (NLP) and Machine Learning (ML) are used. After which Semantic relation, the information related to Symptoms, Causes and Treatment of certain disease and relation between diseases is extracted using Multinomial Naïve Bayes classification algorithm with a prior association rule mining. Weighted Bag-Of-Word representation technique along with word sequence pattern is used to analyze data and identify the pattern, for decision making, There are two features to represent BOW: binary feature value, uses 0 and 1, where 1 is for fact, and frequency feature values, is the number of times it appears in an instance, or 0 if it did not appear. Genia Tagger, specifically tuned for biomedical text such as Medline abstracts, is based on syntactic information which uses the Genia 11 Tagger tool that checks english sentences to provide an output. An algorithm is used for the diagnosis.

Alessio Rossi et al.[9]: We know that football players are prone to injuries, especially in a long season crowded with regular games. We use ML and GPS training data set to predict the time period where the footballers are most likely to get injured due to fatigue and muscle cramps. The major source of injury to the players is not on the pitch, it is mostly off it. Statistics show that extra work load in the training sessions like sprints and doing weights have been a major factor, hence it is important to assess each player's potential and ability and design their workload accordingly. Factors like height, weight do matter but what matters the most is the record of past injuries which may have a fear of recurring. An estimated 188 Million Euros are spent every season on injuries. Standard GPS sensing techniques are applied and this approach is based on automatic data collection which can help save money as well as personal setbacks. Players who are under the age of 18 show jump height, body size and presence of previously occurring injuries are correlated with thigh strain injury.Devices were placed on players' scapulae. Tests reveal that 19/23 injuries were those of people who were injured in the past at least once. Injury forecast is of two phases: Training dataset construction and construction and validation. The injury forecaster for low recall is risky and it detects less % of injuries. The throughput accuracy improves when the dataset is larger as the season goes on and results also improve in the 2nd half of the season. The forecast is based on ACWR and MSWR techniques. We also compare DT to Random Forecast classifier and a Logit classifier.The results show that 80% of injuries were detected by DTs and nearly

half of it were training injuries. LR has a lower performance than DTs, also DTs reduce false alarms. We see that coaches and managers need to take care of those player's total distance and sprint speed performed who are returning from injuries.

A. S. Albahri Rula A et al.[10]:The covid, or the coronavirus is commonly found in animal species like camels,cats,bats etc was always termed to rarely spread to humans.CoVs, a large family of viruses, including (MERS)-CoV, (SARS)-CoV,SARS-CoV-2 and the outbreak has affected severely. Three ML techniques were applied to the MERS-CoV dataset. Data mining and ML algorithms rely on classification algorithms, including decision tree, SVM and Naive Bayes classifier.The outcomes showed that the k-NN classifier(used 2 times) is the best model for the two-class problems, and the decision tree and Naïve Bayes(used 4 times) are the best models for multiclass problems. The decision tree classifier shows higher prediction proficiency than the other models, hence used more frequently(5 times), the others were each used once.Two types of ML are used i.e. supervised and unsupervised(clustering technique) for this prediction. An emotional recognition system which used ML was used to understand human reactions to this disease, such as MERS. A dataset collected in Korea in 2015 showed that elderly people are more likely to be infected. An SVM classifier based on sigmoid, normal and polynomial iterations was used, the results showed their behaviour similarities: the datasheet contains attributes like drug,patient etc .

Some of the aids to the system that uses ML, the nature and behaviour of MERS-CoV needs to be understood, larger datasets are needed, R language is recommended.

Jenni A. M. Sidey-Gibbons and Chris J. Sidey-Gibbons[11]: Machine learning techniques are based on algorithms . algorithm means sets of mathematical procedures which describe the relationships between variables. These algorithms include the General Linear Model regression (GLMs), Support Vector Machines (SVMs) with a radial basis function kernel, and single-layer Artificial Neural Networks. Here,breast mass samples (N=683) are used. ML methods are of two types: Supervised and Unsupervised. We take into account the person's characteristics like height,weight, lifestyle, smoking status to get a certain outcome like onset of diabetes etc. Supervised

ML algorithms are used for tasks like image recognition, language processing. A feature selector is used to process the variables. It selects the characteristics from the dataset which can be understood by the algorithm. There are two kinds of outcomes: discreet that uses classification algorithm or continuous that uses regression algorithm where regression is used to refer to both outcomes since it is binary. The dataset used is Breast Cancer Wisconsin Diagnostic Data Set. for cancer diagnosis, we use 3 models that uses the nuclei of breast masses Unsupervised ML algorithms use dimension reduction techniques and include processes like principal component analysis, latent dirichlet analysis and t-Distributed Stochastic Neighbour Embedding (t-SNE) y algorithms , that is input is taken from the user unlike supervised algorithm.

JohnHalamka et al.[12]: Hypertension, diabetes mellitus,obesity and cardiovascular disease ,zinc, stress and immune dysfunction,vitamin D deficiency, sedentary lifestyle were also considered as risk factors for COVID-19, 124 infected patients were tested. unsupervised ML tool (clustering) to analyze large data sets of COVID-19 helped detect suspected risk factors. Once these suspects are identified, random forest modeling proved useful in their relative contribution to the infection. Random forest analysis to look at 84 separate risk factors or subgroups, creating a series of 1000 decision trees from all the available data. All stored characteristics of the 5000 patients and divided the cohort into halves. The first half served as a training data set to generate hypotheses and to construct the decision tree, the second half of the data served as the testing data set for history of diabetes, muscle cramps in legs and feet, history of emphysema, kidney disease, amputation, dry skin, loud snoring, marital status, social functioning, hemoglobin A1c (HbA1c) level, self-reported health. These insights help to learn the behavior of the pandemic. An artificial intelligence (AI) technology used a deep learning approach called nferX augmented curation to analyze the electronic health records.

Yanxia Zhao et al.[13]: Subjects (N = 1,644) were introduced to the procedures to be followed in this study by an occupational physician. Feature variables related to noise-induced hearing impairment are collected from (1) questionnaires: like the following information: general personal information ,occupational history, personal life habits, and overall health.2) shift-long noise records: obtained for each noise-exposed subject using an ASV5910- R digital recorder(10 hours) and.(3) audiogram tests:Each subject took a general physical and Otoscopic and tympanometric screening to rule out

possible conductive hearing loss done in Audiometric booth using an audiometer .Some of the ML methods used are RF model, has a training module and a classification module, uses bootstrap resampling technique to generate N new training sample sets (N decision trees voting). The Adaboost classifier , highly accurate ,takes the single-layer decision tree as the base classification algorithm. The MLP classifier is a feed-forward neural network, uses backpropagation for training a number of weak learners based on the weight update. The SVM model is based on establishing a separating hyperplane, with maximum distance from the closest points of the training set. To identify risk factors, the selected data values were normalized by min-max normalization technique, then the p value was calculated using a t test. Finally,SVM, Adaboost, and MLP models were selected based on a significance level ($p < 0.01$). The RF algorithm handles high-dimensional data. Target variable defined as a binary classification variable ,and for the regression model, the target variable was a continuous variable where a root mean square error (RMSE) used to evaluate the performance that indicates the four regression models could be used to predict noise-induced hearing loss, the multilayer perceptron regression model had the best performance.

AlanKaplan et al.[14]: This paper mainly focuses on the broader aspect of technology in the medical field of Asthma and chronic obstructive pulmonary diseases(COPD). It talks about the general use of Artificial Intelligence and its sub parts which includes Machine Learning. Most part of the paper suggests that AI and ML involvement in this field is a lot less as compared to other fields of medicine but it has shown promise and has a great potential in treatment, on the other hand the work of general physicians and other respective doctors cant be underestimated given the limited amount of data they have on their hands.The major use of ML in this field is in the medical image recognition where digital input gives binary output. For example, skin diseases can either be benign or malignant. Moreover ML has super sensitivity which accurately gives the output. One of the advantages of AI is that the tests carried out at some place can be treated to the same machinery in geographically remote areas and still the output would be found in the same amount of time with the same accuracy. The AI usage correctly predicted around 88% of more than mild diabetic retinopathy for the FDA. It was also used in China to predict brain tumours and returned with almost the same success rate.We have come to find out that most of the poor predictions are in the early stages when the given data set is small and when extracting data is poor in the training

set. We also see there is an improved accuracy and speed to tests carried out by AI using ML to get better output predictions. Despite the increased use of AI in small scale studies, we don't have much success rate in larger scale studies and is being improved upon. Now the major task for us is to incorporate this in everyday clinical practices.

Clémence Réda et al.[15]: This paper focuses on the use of Machine learning technologies in the creation and development of drugs for medical purposes. Major role of ML here is in the speeding up of drug development process, drug discovery, and to provide a better understanding to the diseases so that the structure and composition of drugs can be made accordingly. The drug development process is majorly split up into 4 long processes called phases since it involves study of design, testing, analyzing the results etc. We have also seen that the research and development process involves very large capital with a chance of failure in later stages which results in loss of both money and time. This is where ML comes into picture which automates some more repetitive but important data, decreases drug development cost and time, transparent data quality control and sharing.As mentioned drug development goes through multiple processes and ML has its application in all the steps. The drug discovery stage is the first one. Here generative adversarial networks and bayesian optimisation techniques are used. Then comes the drug testing phase. Here the absorption, distribution, metabolism, excretion and toxicity levels are calculated. Next comes the process of drug repurposing. This uses the pre existing drugs and compounds for research work. ML is used here in dimension reduction.The dataset to be worked upon is mainly taken online from experts who have collected accurate data. Then comes feature selection which enables trimming and feeding only the required important data which can be performed either manually or automatically. Manual method is more preferable. This part is the most important step and this is where ML genetic algorithms are most heavily involved such as refined DL architectures and sequential algorithms.

Conclusion:

Machine Learning continues to grow in the healthcare industry with the ever-evolving technology advancements.But machine-learning models have not been implemented to a greater extent in healthcare as they have been in other verticals because machine learning is a new technology which is still evolving and is far away from the state of perfection.Thought the research done by the scientist show that the model developed by them are highly accurate but still we aren't confident enough to use them in the real life. Whether its FDA, ICMR or EMA approval, it is a long, arduous and expensive process to test, validate and approve the technology in a healthcare setting.In the healthcare industry, the technologies and systems must be developed in such a way that they follow respective data laws and rules of governing organizations.

References:

[1]:Majid Nour,Kemal Polat.”Automatic Classification of Hypertension Types Based on Personal Features by Machine Learning Algorithms”
Link:https://www.hindawi.com/journals/mpe/2020/2742781/

[2]:Rajesh,T Maneesha,Shaik Hafeez,Hari Krishna”Prediction of Heart Disease Using Machine Learning Algorithms” published in IJET(Singapore/Asia).
Link:https://www.researchgate.net/publication/326733163_Prediction_of_Heart_Disease_Using_Machine_Learning_Algorithms

[3]:Kenji Suzuki “Pixel-BasedMachine Learning in Medical Imaging”
Link:https://www.hindawi.com/journals/ijbi/2012/792079/

[4]:Keyang Xu ,Mike Lam Jingzhi Pang, Xin Gao, Charlotte Band ,Piyush Mathur, Frank Papay, Ashish K. Khanna, Jacek B. Cywiński, Kamal Maheshwari,Pengtao Xie ,Eric P. Xing “Multimodal Machine Learning for Automated ICD Coding”
Link:https://static1.squarespace.com/static/59d5ac1780bd5ef9c396eda6/t/5d472f4658329200011244fe/1564946248003/Xu_K.pdf

[5]:Quan Zou,Kaiyang Qu,Yamei Luo,Dehui Yin,Ying Ju and Hua Tang”Predicting Diabetes Mellitus With Machine Learning Techniques”
Link:https://www.frontiersin.org/articles/10.3389/fgene.2018.00515/full

[6]:Konstantina Kourou,Themis P. Exarchos,Konstantinos P. Exarchos,Michalis V. Karamouzis,Dimitrios I. Fotiadis”Machine learning applications in cancer prognosis and prediction”
Link:https://www.sciencedirect.com/science/article/pii/S2001037014000464

[7]:Zhongheng Zhang,Kwok M. Ho and Yucai Hong”Machine learning for the prediction of
volume responsiveness in patients with oliguric acute kidney injury in critical care”
Link:https://ccforum.biomedcentral.com/articles/10.1186/s13054-019-2411-z

[8]:Pravin Shinde,Prof.Sanjay Jadhav”Health analysis system using machine learning”
Link:http://www.ijcsit.com/docs/Volume%205/vol5issue03/ijcsit20140503280.pdf

[9]:Alessio Rossi ,Luca Pappalardo,Paolo Cintia,F. Marcello Iaia,Javier Fernàndez,Daniel Medina”Effective injury forecasting in soccer with GPS training data and machine learning”
Link:https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0201264

[10]:A. S. Albahri Rula A. Hamid Jwan k. Alwan Z.T. Al-qays A. A. Zaidan B. B. Zaidani A O. S. Albahri A. H. AlAmoodi Jamal Mawlood Khlaf E. M. Almahdi Eman Thabet Suha M. Hadi K I. Mohammed M. A. Alsalem Jameel R. Al-Obaidi H.T. Madhloom"Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review" Link:"https://link.springer.com/content/pdf/10.1007/s10916-020-01582-x.pdf"

[11]:Jenni A. M. Sidey-Gibbons and Chris J. Sidey-Gibbons"Machine learning in medicine: a practical introduction"
Link:https://bmcmedresmethodol.biomedcentral.com/track/pdf/10.1186/s12874-019-0681-4.pdf

[12]:JohnHalamkaMD, MSa, PaulCerratoMAb, Adam Perlman MD, MPH"Redesigning COVID-19 Care With Network Medicine and Machine Learning"
Link:https://www.sciencedirect.com/science/article/pii/S2542454820301910

[13]:Yanxia Zhao,Jingsong Li, Meibian Zhang,Yao Lu, Hongwei Xie,Yu Tian,Wei Qiu"Machine Learning Models for the Hearing Impairment Prediction in Workers Exposed to Complex Industrial Noise"
Link:https://journals.lww.com/ear-hearing/fulltext/2019/05000/machine_learning_models_for_the_hearing_impairment.24.aspx

[14]:AlanKaplan,HuiCao,J. MarkFitzGerald,NickIannotti,EricYang,Janwillem W.H.Kocks,KonstantinosKostikas,DavidPrice,Helen K.Reddel,IoannaTsiligianni, Claus F.Vogelmeier, PascalPfister,PaulMastoridis"Machine Learning in Respiratory Medicine and Asthma"
Link:https://www.sciencedirect.com/science/article/pii/S221321982100194X

[15]:Clémence Réda,Emilie Kaufmann,Andrée Delahaye-Duriez"Machine learning applications in drug development"
Link:https://www.sciencedirect.com/science/article/pii/S2001037019303988