# A Comprehensive Survey on Artificial Intelligence and Machine Learning Techniques

Anisha. C. D
Dept of CSE
PSG College of Technology
Coimbatore, India

Dr. K. G. Saranya
Dept of CSE
PSG College of Technology
Coimbatore, India

**Abstract:- Artificial Intelligence (AI) is a broad term used to represent the intelligence of the system. Machine Learning is the sub category of Artificial Intelligence. The application of Machine Learning algorithms is prevalent in every fields from the recommendation systems of news, music and diagnosis decision systems and it is negligible because it makes life easy and comfortable. It has been helpful in predicting and forecasting the information even for the COVID 19 pandemic. This paper presents a comprehensive review on Artificial Intelligence (AI) and Machine Learning (ML) Techniques employed in various domain. The review covers the advantages, limitations and various applications of the AI and ML. The prevalent AI and machine learning applications which aids the situation of COVID 19 pandemic is also presented.**

*Keywords— Artificial Intelligence (AI), Machine Learning (ML), COVID 19.*

## I.    INTRODUCTION

COVID 19: The novel Corona Virus is transmitted from person to person and it has bought a huge global threat. The transmission is mainly through respiratory and its symptoms mostly include cough, fever and cold. The Machine learning techniques helps in the diagnosis of COVID 19 and also helps in forecasting the demographic information of the COVID 19.

Artificial Intelligence (AI) is also known as Intelligent Agent. It has an enormous growth covering all fields of applications from AI assistants in mobile phones, Intelligent Telemonitoring systems and Robotic surgery in Healthcare and driverless car in automobile. Automation is essential in fast moving world which is made possible through AI technology.[1]

Machine Learning (ML) is the branch of AI wherein it builds a model based on the data which is known as training data and the model is checked based on testing data or validation data and the accuracy of the model is improved by tuning its parameters and the process.

Deep Learning Algorithms is a wider part of the Artificial Neural Network, Machine Learning Algorithms. It plays a vital role in prediction, classification of large image data.

## II.    PAPER ORGANIZATION

The rest of the paper is organized as: Section III provides the classification of Machine Learning Algorithms, Section IV presents the Applications related to AI which covers the products available in market, applications related to ML, recent research along with its techniques, advantages and disadvantages. Section V provides the review of the state-of-the-art AI and Machine Learning techniques deployed in COVID 19 diagnosis.

## III.    CLASSIFICATION OF MACHINE LEARNING ALGORITHMS

The machine Learning Algorithms can be classified into four categories, they are Unsupervised Learning, Supervised Learning, Reinforcement Learning and Semi-Supervised Learning.

### 1.Unsupervised Learning

In Unsupervised Learning, the model forms the structure from the unlabeled data. There is a grouping process of the data. Examples of Unsupervised Learning includes clustering algorithms namely K means Clustering, hierarchical clustering, Agglomerative clustering.[15]

### 2.Supervised Learning

In Supervised Learning, the data is split into features and the corresponding class label, the model is built based on the mapping of features with the class label. Examples of Supervised Learning includes Regression algorithms namely Linear Regression, Lasso Regression, Ridge Regression and Classification algorithms namely K Nearest Neighbor (KNN), Support Vector Machine (SVM), Random Forest, Linear Discriminant Analysis, Artificial Neural Network (ANN).

### 3.Reinforcement Learning

In Reinforcement Learning, the model is exposed to trial-and-error process and learns through the past experience, Examples of Reinforcement Learning includes Markov Decision Process.

### 4.Semi-Supervised Learning

As the name itself indicates, it is combination of unsupervised learning and supervised learning. Examples of Semi- supervised Algorithms include Generative Models, Graph Based Model and Heuristic Approaches.

## IV.     APPLICATIONS

### A.  Artificial Intelligence Applications- Products in market

i.Marketing

Movies Recommendation system, for example: Netflix a popular online application to watch films uses a high predictive technology to recommend movies based on the past reviews, it parses across million records to provide an accurate prediction.

ii.Banking

Banks provides a user support tool which will be available throughout the day to answer queries from the user. Example: HDFC bank has launched an Electronic Virtual Assistant (EVA) which address user queries.

iii.Agriculture

a.See and Spray Robot: This robot has been launched by Blue River Technology which sprays herbicide accurately on plants.
b.Plantix: This application developed by PEAT, detects the fault plant which has less nutrient deficiency.

iv.Health care

A decision support system will assist doctors for patient monitoring. An organization named Cambio Health Care developed a stroke prevention decision support system which provides doctor with an alert when the patient is at risk.

v.Space

The rover sent to the MARS by NASA are AI based which has to perform independent targeting of cameras to obtain information of MARS.

vi.Automobile

Self – Driving Cars is the automated car which uses AI system that gathers data and produce signals which aids the operation of the vehicle. For Example: Tesla's Car which automatically detect objects without human involvement.

vii.Chatbots

Chatbots are virtual Assistant and it is available on mobile phones, laptop. The popular chatbots are Siri, Cortona and Amazon Echo. The technology used in these applications are Speech Recognition and Natural Language Processing (NLP) which has to translate the human language into set of instructions to be performed like playing music, a making a call etc.

### B.  Unsupervised Machine Learning Algorithms Applications

i.Medical Field – Neurological Disorder

The Clustering based algorithm is applied to Neurological dataset mainly Alzheimer Disorder. The different Clustering methods applied are K means, Hierarchical Agglomerative Clustering, Multi-Layer clustering. [16]

ii.Networking and Security

Intrusion in a network is a serious issue. The Clustering based and Expectation Maximization Algorithm were used to detect the Network Intrusion detection system.[17]

### C.  Supervised Machine Learning Applications

i.Covid 19 Prediction

The standard Models created for the Covid 19 prediction were Multi-Layer Perceptron and Adaptive Network based Fuzzy Inference System. [18].

ii.Optical Networks and Systems

Autoregressive Integrated Moving Average (ARIMA) is a supervised learning method used for Traffic Prediction in optical Networks.[19]

### D.  Reinforcement Learning Applications

i.Human Level Video Game Play

Multi-Layer Artificial Neural Network and Deep Q Network (DQN) has been used for functioning the Video Game Play.[20]

ii.Watson's Daily-Double Wagering

This is the system which uses Backpropagation Artificial Neural Network developed by the team of developers of IBM.[20]

### E.  Semi-Supervised Learning Applications

i.Customer Behavior Modeling

The Multilayer Perceptron with Back Propagation Algorithm is used to predict the Customer Behavior. The training data of the Algorithm consists of set of Labeled data and set of unlabeled data.[22]

ii.Video Recommendation System

A graph- based Algorithm Known as the Adsorption Algorithm has been used for Video Recommendation.[23]

Table 1 presents the Techniques of ML used in various domain and its advantages and limitations. The main domain considered in the review are Online Advertising, Healthcare, Agriculture, Education, Governance, Banking, Transportation and Environment.

TABLE 1- RECENT AI AND ML - APPLICATIONS, ITS TECHNIQUES, ADVANTAGES AND DISADVANTAGES.

| APPLICATION | TECHNIQUES | ADVANTAGES | DISAVANTAGES |
|---|---|---|---|
| Online Advertising | Click through Rate (CTR) Prediction [4]<br>1. Deep Belief Neural Network<br>2. Logistic Regression | There is flexibility in training the Deep Belief Neural Network, it can be trained without supervision and with supervision. | The Computational Cost of Deep Belief Neural Network is expensive. |
| Healthcare | Stroke Early Detection and Diagnosis [24]<br>1. Genetic Fuzzy Finite State Machine<br>2. Principal Component Analysis (PCA) | The use of Principal Component Analysis (PCA) improves the performance of the system by removing the correlated features. | There is possibility of information loss while reducing the features if the variance parameters of the PCA are not tuned properly. |
| Agriculture | Irrigation Optimization [6][7]<br>1. Artificial Neural Network (ANN) – Backpropagation and feedforward.<br>2. Partial Least Square Regression (PLSR) | The efficiency of the system is improved, since the handling of variables is proper. | The PLSR may be prone to overfitting. |
| Education | Academic Achievement Assessment [3] [8]<br>1. Decision Tree<br>2. Extremely Randomized Tree<br>3. Support Vector Machine<br>4. K Nearest Neighbor (KNN) | The Tuning process of Hyperparameters in Tree based Algorithms is not necessary. | Time Complexity is higher with respect to tree-based Algorithms namely decision tree, Extremely Randomized Tree. |
| Governance | Automated Decision Making in Democratic Settings [11]<br>1. Support Vector Machine | Support Vector Machine (SVM) works efficiently for high dimensional space. | SVM performance will be poor when there is noise present in the dataset. |
| Banking | Prediction of Loan Sanctioning Process [21]<br>1. Naïve Bayes Algorithm<br>2. Logistic Regression<br>3. Decision Tree | Naïve Bayes Algorithm is not reactive to missing data. | The input structure is vital the features are related then the classification is not good in naïve Bayes. |
| Transportation | Prediction of Road Freight Transportation [10]<br>1. Bayesian Network | The "Prior" specification in the Bayesian algorithm it provides clear insights. | The feasibility in terms of computation it is not flexible in Bayesian Algorithms. |
| Environment | Satellite Predictors to Forecast Global Solar Radiation [12][13]<br>1. Random Forest<br>2. Multivariate Adaptive Regression Spline (MARS) | MARS algorithm can handle categorical and continuous data. | The process is prolonged if there are many variables. |

## V. STATE OF THE ART AI/ML TECHNIQUES DEPLOYED IN COVID 19 DIAGNOSIS

In [5] Mei et al has presented an Artificial Intelligence enabled Covid 19 diagnosis tool wherein three AI models has been used. The first model is built on Chest CT scan data which is processed using Convolutional Neural Network, the second model is built on clinical information using Machine Learning Model Namely Random Forest, Support Vector Machine and Multi-Layer Perceptron (MLP). The Fusion data of Chest CT scan and Clinical information data is integrated by the Multi – Layer Perceptron.

In [25] Sina et al has provided an outbreak prediction using the Multi-Layer Perceptron (MLP) Machine Learning technique and Adaptive Network based Fuzzy Inference system from the statistical data of five different countries namely Italy, Germany, Iran, USA and China.

In [26] Zixin Hu et al, developed an AI based forecasting tool for COVID 19 wherein the data has been collected from World Health Organization (WHO) and the algorithms such as Autoencoders and Clustering algorithms has been used. The use of Clustering algorithms is to group the provinces and cities to know the transmission pattern.

Shuai Wang [27] et al, proposed a Screening method using the Convolutional Neural Network, a deep learning technique which is a branch of Artificial Intelligence. The data used in this system were CT scan Images of Positive patients.

Mohammad Pourhomayoun et al, developed a Mortality Risk Prediction Model [28] using the Machine Learning Techniques such as Neural Network, decision tree, Support Vector Machine (SVM), Logistic Regression and K Nearest Neighbor (KNN), Random Forest and Decision Tree. The data set has been retrieved from 76 countries which consists of

demographic and physiological information of 1,1,7,000 Positive Patients. The features are selected from the raw data using Filter and Wrapper Methods. Neural Network classifier performs better than the other classifier because the hyperparameters are tuned using Grid Search Method.

Yazeed Zoabi et al,[29] presents a Machine Learning Approach for the diagnosis of COVID 19 based on the symptoms. Gradient Boosting Machine (GBM) has been used for the prediction. The dataset is released by the Ministry of Israeli of 51,831 tested individuals. The features present in the dataset age and gender as baseline features and symptoms such as cough, sore throat, fever, shortness of breath and headache. The results indicate that the fever and cough are the important key features.

Davide Brinati et al [30], provides two machine learning models, one for distinguishing Covid patients from healthy patients and the other one is for interpreting the results of blood tests. The machine learning algorithms used were Decision Tree, Random Forest, K – Nearest Neighbor, Support Vector Machine, Naïve Bayes, Extremely Randomized Tree (Extra Tree Classifier) and Logistic Regression. The dataset has been retrieved from the 273 cases blood samples information released by the IRCCS Ospedale San Raffaele. It consists of demographic information attached.

Ali Naren et al [31] has developed an automatic detection method using Convolutional Neural Network (CNN) on the X- Ray Images data from the 50 Covid 19 Patients which has been available in Github Repository. The three Convolutional Network (CNN) pre trained models has been used and they are ResNet50, Inception V3 and Inception ResNetV2.

Random Forest Based Prediction from clinically available blood samples has been developed in [32]. The blood samples of 49 patients have been considered in this analysis. The validation methods used to assess the trained classifier were 10 Cross Validation and external validation set.

In [32], a Convolutional Neural Network based for Covid 19 Prediction. The three transfer learning methods are used in this system and they are AlexNet, GoogleNet and ResNet18. The dataset used consists of X ray images of 69 Covid19 Patients ,79 Normal subjects, 79 Pneumonia Bacteria Patients and 79 Pneumonia Virus Patients.

A. *INSIGHTS FROM THE AI/ML BASED COVID 19 RESEARCH REVIEW*
i. Data Considered in the Recent Research Works
a. Demographic data
b. Clinical Physiological Data
c. X Ray Images data
d. CT Scan Images
e. Blood Samples data
ii. Classifiers used based on the type of data
a. Convolutional Neural Network has been mainly used in the X ray and CT scan images based Covid Diagnosis.
b. Machine Learning Techniques Namely Random Forest, KNN, SVM, Naïve Bayes, Logistic Regression has been used for Blood Samples based Covid diagnosis.
c. Neural Network are mostly used for demographic and clinical physiological data.
iii. Hyperparameter Tuning, validation techniques such as cross validation using Kfold and usage of External validation set has played a key role in the Covid19 research reviewed for improving the accuracy.

Table 2 – Summary of ML techniques in COVID application with its advantages and Limitations

| PAPER | DATA USED | ML TECHNIQUES | ADVANTAGES | LIMITATIONS |
|---|---|---|---|---|
| [5] | CT Scan | Covid 19 Diagnosis [5] <br> Processing: <br> Convolutional Neural Network (CNN) <br> Prediction: <br> 1.    Support Vector Machine <br> 2.    Random Forest <br> Multi-Layer Perceptron | The hybrid approach of combining models makes the system robust. | Orientation and spatial information are given least importance in CNN. |
| [25] | Statistical data obtained from five countries Italy, Germany, Iran, USA and china | Covid 19 outbreak Prediction <br> 1.    Multi-Layer Perceptron <br> 2.    Adaptive Network based Fuzzy inference system (ANFIS) | The computation of the complex problem is accurate since it uses one or more hidden layers which optimizes the results. | The training time of Mutli Layer Perceptron is long. |
| [26] | Data of World Health Organization (WHO) from January 11 to February 27, 2020 | AI based Real Time Forecasting of COVID 19 <br> 1.    Autoencoders <br> Clustering Algorithms | The feature extraction performed by autoencoders are better even when the data is unclean. | The prerequisite of autoencoder training requires lots of data and more processing time with appropriate tuning of hyperparameters. |

| [27] | 1,065 Computed Tomography Images of Positive Patients | Deep Learning Based Screening of Corona Virus Convolutional Neural Network (CNN) | Convolutional Neural Network performs well for the image's dataset. | Convolutional Neural Network faces the problem with the Translation and Pooling layers which would be solved using proper augmentation methods. |
|---|---|---|---|---|
| [28] | 1,17,000 Covid 19 Positive cases demographic and physiological data from 76 countries. | AI based Mortality Risk Prediction in Patients 1. Feature Selection: Filter Method and wrapper methods. 2. Classifiers: Neural Network, decision tree, Support Vector Machine (SVM), Logistic Regression and K Nearest Neighbor (KNN), Random Forest and Decision Tree. | The hyperparameters were tuned using Grid Search Method for Neural Network which yield a good result than the other classifiers employed in the system. | The preprocessing stage neglected using standardization technique namely standard scaler method which will optimize the results by providing a higher accuracy than the prevalent accuracy of 93.75%. |
| [29] | 51,831 Records of Patients. | Machine Learning Approach based on COVID 19 Symptoms Gradient Boosting Machine (GBM) | The GBM constructs decision trees with features of residual errors of previous decision tree which makes GBM provide accurate results and the missing values are handled by the classifier itself. | The parameters of GBM have to be tuned to have a good accuracy. |
| [30] | IRCCS Ospedale San Raffaele dataset of 279 cases with blood samples information. | Covid 19 Detection from Blood Samples using ML Classifiers: 1. Decision Tree 2. Random Forest 3. K- Nearest Neighbors (KNN) 4. Naive Bayes 5. Extremely Randomized Tree 6. Support Vector Machine (SVM) 7. Logistic Regression | The bias in the data due to low samples has been controlled using cross validation technique. | There is less focus on appropriate preprocessing of data which is to be given as input to each classifier, if focused on this aspect there is possibility of improvement in accuracy. |
| [31] | Chest X-ray radiographs from 50 Covid 19 patients from Github Repository. | CNN Based Automatic Detection using X ray Images 1. Classifier: Convolutional Neural Network (CNN) 2. Cross Validation: K fold (1-5) | The X ray data used in the system can be obtained easily and quickly, it performs better without any feature extraction stage. | The performance of the different CNN models has not been tested by increasing the images in the dataset. |
| [32] | Blood sample data of 49 patients. | Classifier 1. Random Forest Validation Method 1. 10 Cross Validation. 2. External Validation set | The data used played a vital role as the virus infecting the body will change the chemical composition which led to identify the Covid 19 patients more efficiently using ML algorithms. | This application has not covered the Covid19 patients with atypical symptoms. |
| [33] | X ray Images from Covid, Pneumonia Bacteria, Pneumonia Virus and Normal subjects. | Classifier Convolutional Neural Network Models (AlexNet, GoogleNet and Restnet18 | The CNN models chosen were based on less training and computation time. | The model works accurate for low sample set, the model is not trained on large data set. |

## VI. CONCLUSION

This paper provides a comprehensive review on Artificial Intelligence, Machine Learning, types and its applications. The application of Machine for various prediction and forecasting is essential for fast and informed decision. The need for AI and ML is increasing as there is numerous data generating which has to be converted to meaningful information faster than the rate of generation. The recent research work in various domain with its techniques, advantages and disadvantages has been presented which will be useful for researchers. The paper also presents the state-of-the-art AI/ML techniques which is used in COVID 19 predictions along with the data used, techniques with its advantages and disadvantages which provides an insights and future ideas to the researches to pursue the research in the field of COVID diagnosis using AI/ML.

## REFERENCES

[1]. Sumit Das et al, "Applications of Artificial Intelligence in Machine Learning: Review and Prospect", International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 9, April 2015

[2]. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke and Vascular Neurology 2017

[3]. Tarik A. Rashid, "Student Academic Performance Using Artificial Intelligence", ZANCO Journal of Pure and Applied Sciences,2015

[4]. Z. Jiang, S. Gao, and W. Dai. "Research on CTR prediction for contextual advertising based on deep architecture model." Journal of Control Engineering and Applied Informatics 18, no. 1, pp. 11-19, 2016.

[5]. Mei, X., Lee, H., Diao, K. et al. "Artificial intelligence–enabled rapid diagnosis of patients with COVID-19". Nat Med (2020).

[6]. M. Pathan, N. Patel, H. Yagnik, et al., Artificial cognition for applications in smart agriculture: A comprehensive review, Artificial Intelligence in Agriculture (2020)

[7]. Cramer, R.D. Partial Least Squares (PLS): Its strengths and limitations. Perspectives in Drug Discovery and Design 1, 269–278 (1993)

[8]. Rashid, Tarik & Aziz, Nian "Student Academic Performance Using Artificial Intelligence.", 28. 56-69. 10.21271/zjpas. v28i2.544 (2016).

[9]. Frederico Cruz-Jesus, "Using Artificial Intelligence to assess academic achievement in Public Schools of a European Union Country", Heliyon (2020)

[10]. B. Mrowczynska, M. Ciesla, A. Krol, A. Sladkowski "Application of artificial intelligence in prediction of road freight transportation", Promet Traffic Transp., 29 (4) (2017), pp. 363-370

[11]. E. Nissan et al, "Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement." AI Soc., 32 (3) (Aug. 2017), pp. 441-464

[12]. R.C. Deo, M. Şahin, J.F. Adamowski, J. Mi, "Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: a new approach", Renew. Sustain. Energy Rev., 104 (2019), pp. 235-261

[13]. Maciej Przybyłek, et al "Application of Multivariate Adaptive Regression Splines (MARSplines) Methodology for Screening of Dicarboxylic Acid Cocrystal Using 1D and 2D Molecular Descriptors", Crystal Growth & Design 2019 19 (7), 3876-3887

[14]. Raffaele Cioffi et al, "Artificial Intelligence and Machine Learning Applications in Smart Production: Progress, Trends, and Directions", Sustainability 2020

[15]. Khanam, Memoona et al, "A Survey on Unsupervised Machine Learning Algorithms for Automation, Classification and Maintenance", International Journal of Computer Applications,2015

[16]. Alashwal Hany et al, "The Application of Unsupervised Clustering Methods to Alzheimer's Disease", Frontiers in Computational Neuroscience, 2019.

[17]. Uttam Kumar Dey et al, "Network Intrusion Detection with Unlabeled Data using Unsupervised Clustering Approach." International Journal of Engineering Science and Computing, February 2019.

[18]. Ardabili, Sina et al "COVID-19 Outbreak Prediction with Machine Learning.", Research Gate ,2020

[19]. Francesco Musumeci et al, "An Overview on Application of Machine Learning Techniques in Optical Networks" arXiv,2018.

[20]. A. G. Barto et al, "Some Recent Applications of Reinforcement Learning."

[21]. E. Chandra Blessie et al, "Exploring the Machine Learning Algorithm for Prediction the Loan Sanctioning Process, International Journal of Innovative Technology and Exploring Engineering (IJITEE), ISSN: 2278-3075, Volume-9 Issue-1, November, 2019

[22]. Siavash Emtiyaz et al, "Customers Behavior Modeling by Semi-Supervised Learning in Customer Relationship Management." Advances in information Sciences and Service Sciences (AISS) Volume3, Number9, October 2011.

[23]. Baluja, S. et al, "Video suggestion and discovery for youtube: Taking random walks through the view graph." In Proceedings of the 17th international conference on world wide web (pp. 895–904). ACM

[24]. Villar JR, González S, Sedano J, et al. "Improving human activity recognition and its application in early stroke diagnosis", Int J Neural Syst 2015;25:1450036.

[25]. Ardabili, Sina et al, "COVID-19 Outbreak Prediction with Machine Learning". Research Gate (2020).

[26]. Hu, Z., Ge, Q., Jin, L., and Xiong, M., "Artificial intelligence forecasting of Covid-19 in China", arXiv preprint arXiv:2002.07112 (2020).

[27]. Shuai Wang et al, "A deep learning algorithm using CT images to screen for Corona Virus Disease (COVID-19), MedRxiv, April 2020

[28]. Mohammad Pourhomayoun et al , "Predicting Mortality Risk in Patients with COVID-19 Using Artificial Intelligence to Help Medical Decision-Making", MedRxiv, April 2020.

[29]. Yazeed Zoabi et al ,"COVID-19 diagnosis prediction by symptoms of tested individuals: a machine learning approach.", MedRxiv, May 2020.

[30]. Davide Brinati et al, "Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study", Research Gate, April 2020.

[31]. Ali Narin et al, "Automatic Detection of Coronavirus Disease (COVID-19) Using X-ray Images and Deep Convolutional Neural Networks."

[32]. Jiangpeng Wu et al, "Rapid and accurate identification of COVID-19 infection through machine learning based on clinical available blood test results", medrxiv,2020

[33]. Mohamed Loey et al, "Within the Lack of Chest COVID-19 X-ray Dataset: A Novel Detection Model Based on GAN and Deep Transfer Learning.", symmetry, MDPI, April 2020.