

A Comparative Study on Health Care using Machine Learning and Deep Learning

Ajinkya Padule
Student, CSE

Department of Computer Science and Engineering
MIT School of Engineering,
MIT Arts Design and Technology University
Pune, India

Aman Shaikh
Student, CSE

Department of Computer Science and Engineering
MIT School of Engineering,
MIT Arts Design and Technology University
Pune, India

Aman Patel
Student, CSE

Department of Computer Science and Engineering
MIT School of Engineering,
MIT Arts Design and Technology University
Pune, India

Arsalan Patel
Student, CSE

Department of Computer Science and Engineering
MIT School of Engineering,
MIT Arts Design and Technology University
Pune, India

Jyoti Gavhane

Assistant Professor

Department of Computer Science and Engineering
MIT School of Engineering,
MIT Arts Design and Technology University
Pune, India

Abstract—Health care is very fundamental for having a good life. However, seeking a doctor's consultation in the event of a health issue is challenging. Health care is a growing industry that affects every aspect of our lives. The healthcare industry is progressing alongside technological advancements. Technology is rapidly integrating itself with medical sciences for the welfare of humanity, with better preventative, diagnosis, and treatment choices. Machine learning is reshaping the world through transforming a variety of industries, including healthcare, education, transportation, food, entertainment, and various manufacturing lines, among others. It will have an impact on practically every element of people's lives. Deep learning is a type of machine learning in which computers are taught to do things that humans do naturally. In this research, we look at some of the machine learning and deep learning techniques that have been used to build effective healthcare applications. This paper contributes to narrowing the research gap in the development of an effective decision support system for medical applications.

Keywords— *Machine Learning, Deep Learning, Health Care, Diseases*

I. INTRODUCTION

AI has lately been adopted by medical physicians to attempt to overcome challenges that the medical community is facing. Based on what we already know, it looks that we will be able to solve some of the most important problems in medicine today. Machine learning is considered a major field in artificial intelligence. Health care is a growing industry that affects all aspects of our life. The health care industry is progressing in tandem with technological advancements. Technology is rapidly integrating itself with medical sciences

for the welfare of humanity, with better preventative, diagnosis, and treatment choices. Healthcare management is utilizing this method to forecast wait times for patients in exigency department waiting for places [1]. According to doctors, time is a crucial factor in diagnoses, and arriving at an appropriate decision in a timely fashion can aid patients greatly [2-3]. The majority of medical records are now handled electronically. With so much data coming in, a proper way is required to organize, analyze, secure, and store data electronically while being quick and efficient. As a result, the integration of machine learning and medical sciences would indeed contribute to the future. Precision medicine aims to "ensure that the appropriate treatment is given to the right patient at the right time" by considering a variety of factors in a patient's data, such as molecular features, environment, electronic health records (EHRs), and lifestyle [4-6].

II. MACHINE LEARNING

Machine Learning: the classic definition is - A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E [7]. Machine Learning is a subset of artificial intelligence that involves several probabilistic, statistical, and optimization techniques to enable computers to learn from previous examples and uncover difficult-to-find patterns from large, noisy, or complex data sets. Machine learning algorithms are classified into four categories: Supervised learning, Unsupervised learning, Semi-supervised learning, and Reinforcement Learning.

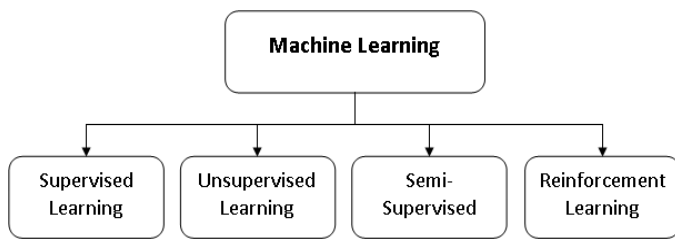


Fig. 1 ML Types

A. Supervised learning

Supervised machine learning, as the term suggests, is based on supervision. It means that in the supervised learning technique, we train the machines using a "labeled" dataset, and the machine predicts the output based on the training. We can observe that we first train the machine with the input and output, and then we ask it to predict the output using the test dataset.

B. Unsupervised learning

Unsupervised machine learning uses an unlabeled dataset to train the machine. Models are trained with data that is neither classified nor labeled in unsupervised learning, and the model operates on the data without any supervision. The unsupervised learning algorithm's main goal is to classify or categorize the unsorted dataset into classes or clusters based on similarities, patterns, and differences.

C. Semi-supervised learning

This technique is a mix of supervised and unsupervised learning techniques. This means that the system will train on both labeled and unlabeled data.

D. Reinforcement Learning

Reinforcement Learning refers to a group of strategies in which the system tries to learn through direct interaction with the environment to maximize some metric of cumulative reward. It's important to note that the system has no prior knowledge of the behavior of the environment, and the only way of learning is through trial and error.

III. CLASSIFICATION AND PREDICTION

Prediction and classification Classification or prediction is one of the most important goals for using machine learning algorithms. 'Classification' and 'prediction' are two terms that can be used interchangeably [8]. The ability to generalize a dataset, which means the ability to recognize new outcomes from previous data, is known as classification or prediction [9].

IV. ML ALGOS

A. Support Vector Machines

The Support Vector Machine, or SVM, is a widely used Supervised Learning technique for solving classification and regression problems. It is, however, mostly used in Machine Learning to solve classification problems. The SVM algorithm's purpose is to find the best line or decision boundary for partitioning n-dimensional space into classes so that new data points can be properly placed in the correct category. A hyperplane is called the best decision boundary.

B. Naïve Bayes Clustering

A classifier that classifies labeled data is a Naive Bayes classifier. To determine the probabilistic classification value, they require a large number of parameters.

C. Decision Trees

A Decision Tree is a supervised learning technique that can be used to solve both classification and regression problems, however, it is most commonly used to solve classification problems. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the outcome in this tree-structured classifier.

D. Logistic regression

In statistics, a logistic regression model is one in which the dependent variable is categorical, i.e., the dependent variable can only take two values, "0" and "1," which reflect outcomes such as pass/fail, win/lose, alive/dead, or healthy/sick. Machine learning, most medical professions, and social sciences are all examples of where logistic regression is applied [10].

E. Random forest

Random forest is a supervised learning method that can be used to predict and classify data. It can be used in both classification and regression. However, it is mostly used to solve classification problems. The Random Forest method builds decision trees on data samples, then gets predictions from each of them, and then selects the best solution through voting. It's an ensemble method that's preferable than a single decision tree because it averages the results to reduce overfitting.

V. DEEP LEARNING

Deep learning is a subset of machine learning, which is essentially a neural network that tries to mimic human brain behavior. Given the increasing complexity of healthcare data in recent years, the use of Machine Learning methodologies such as Deep Neural Network (DNN) models has grown more appealing in the healthcare system. Because we're dealing with an image dataset, machine learning could be quite useful in diagnosing. Medical pictures supplied by medical imaging techniques can be analyzed using machine learning algorithms such as neural networks. A weighted and bias-corrected input value is processed through a non-linear activation function such as ReLu and softmax to generate an output in a traditional Deep Neural Network (DNN) [11]. As a result, the goal of DNN training is to optimize the network's weights so that the loss function is minimized [12].

CONVOLUTIONAL NEURAL NETWORK:

The three primary layers of a CNN are the convolution layer, pooling layer, and fully connected layer.

- Each of these levels is responsible for specific spatial operations.
- CNN creates feature maps by convolving the input image with different kernels in convolution layers.
- The pooling layer is usually applied after a convolution layer, and it helps to reduce the size of feature maps and network parameters.
- A flatten layer follows the pooling layer, followed by some fully connected layers.

- The flatten layer converts the 2D feature maps created in the previous layer into 1D feature maps suitable for the fully connected layers that follow.
- The flattened vector can be used to classify the images later on.

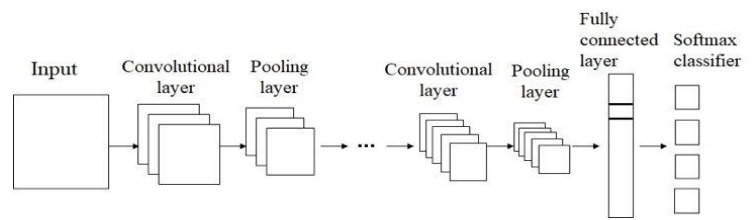


Fig. 2 CNN Architecture

VI. LITERATURE SURVEY

Table I. Different ML and DL Techniques for Diagnosis of Various Diseases

Disease	Author	Technique used	Dataset	Accuracy	Reference
Heart Disease	Apurb Rajdhan (2020)	Random Forest	UCI devaland dataset	90.16%	[13]
	Archana Singh(2020)	K-nearest neighbour	UCI repository dataset	87%	[14]
	Abhijeet Jagtap(2019)	SVM	UCI dataset	64.4%	[15]
	Harshit Jindal (2021)	KNN	UCI repository dataset	87.5	[16]
Diabetes Disease	Nesreen Samer El_Jerjawi (2018)	ANN	Documentation of the Association of diabetic's city of Urmia	87.3%	[17]
	Safial Islam Ayon (2019)	Deep Neural Network	Pima Indian diabetes dataset	98.35%	[18]
	Aishwarya Mujumdae (2019)	Logistic Regression		96%	[19]
	Amani Yahyaoui (2020)	Random Forest	Pima Indian Diabetes dataset	83.67%	[20]
Liver Disease	A. K. M. Sazzadur Rahman (2019)	Logistic Regression	dataset from the UCI Machine Learning Repository. In addition, the original dataset was collected from the northeast of Andhra Pradesh, India	75%	[21]
	M. Banu Priya (2018)	J.48 and Bayesnet	Indian Liver Patient Dataset	95.04 % and 90.33 %	[22]
	Dr.S.Vijayarani	SVM	Indian Liver Patient Dataset(ILPD)	79.66%	[23]
	Md.Irfan	K-nearest neighbour	Indian Liver Patient Dataset(ILPD)	73.97%	[24]
Malaria Disease	Gautham Shekar(2020)	Basic CNN, VGG-19 Frozen CNN, and VGG-19 Fine Tuned CNN		94%,92% and 96%	[25]
	Krit Sriporn(2020)	CNN	The data were collected from a thin blood smear on a slide containing malaria from the hospital by using a microscope. The total sample comprised 201 patients, of which 151 were infected and 50 patients were not.	96.85%	[26]
	Soner Can Kalkan	CNN	U.S National Library of Medicine and used consists of 27,558 cell images	95%	[27]

	Vijayalakshmi(2019)	VGG19-SVM		91%	[28]
Pneumonia Disease	Tilve et al, 2020	VGG16 (CNN)		93.6%	[29]
	Racic et al, 2021	CNN	Chest X-Ray Images (Pneumonia)	88.90%	[30]
	Ayan E, Unver H. M., 2019	Transfer Learning 1. VGG16 2. Xception	Frontal chest X-ray images	VGG - 87% Xception - 82%	[31]
	Ayush Pant et al, 2020	ResNet-34 based U-Net and EfficientNet-B4 based U-Net	Dataset from Kaggle provided by Guangzhou Women and Children's Medical Centre	82%	[32]

A. Heart disease

UCI Machine Learning Repository Dataset was used for the detection of the disease. Vikas Chaurasia and Saurabh Pal in their paper have used different machine learning algorithms for predicting heart disease. Out of which J48 gives an accuracy of 84.35% and bagging gives an accuracy of 85.03% [33].

Rahma Atallah and Amjed Al-Mousa applied SGD classifier, Random forest classifier and logistic regression and got an accuracy of 88%, 87%, and 87% respectively [34].

B. Diabetes disease

Pahulpreet Singh Kohli and Shriya Arora proposed a Support Vector Machine (Linear) algorithm. With a test size of 40%, they achieved an accuracy of 77.92%, and that of 35% they achieved an accuracy of 76.75% [35].

V. Anuja Kumari and R.Chitra in their paper used the Support Vector Machine algorithm on the UCI dataset for detecting diabetes disease and got an accuracy of 78%, sensitivity of 80%, and specificity of 76.5% [36].

C. Liver disease

Nazmun Nahar¹ and Ferdous Ara² compared various decision tree techniques. Out of which Decision Stump gave the highest accuracy with 70.67%. Using Random Forest Classifier, they achieved an accuracy of 69.47% [37].

A. S.Aneeshkumar and C.Jothi Venkateswaran in their study used two classification methods which are Naive Bayesian and C4.5 decision tree. Naive Bayesian gave an accuracy of 89.6% and C4.5 gave an accuracy of 99.2% [38].

D. Malaria disease

Godson Kalipe used XGboost, ANN, Random Forest, and SVM algorithms. Dataset used was 6 years old data collected from various health centers in the district of Visakhapatnam. The results showed that XGboost got the highest accuracy of 96.26% [39].

Aimon Rahman proposed two network architectures that are Custom network architecture, Fine-tuning on pre-trained models, and extracting features from a convolutional network (CNN) followed by a support vector machine classifier (SVM). From their proposed models, TLVGG16 achieves an accuracy of 97.77% [40].

E. Pneumonia disease

Nada M. Elshennawy and Dina M. Ibrahim developed four different models by changing the used deep learning method; two pre-trained models, ResNet152V2 and MobileNetV2, a Convolutional Neural Network (CNN), and a Long Short-Term Memory (LSTM). ResNet152V2 model outperformed

other models in terms of accuracy. They got an accuracy of 99.22% [41].

Karan Jakhar and Nishtha HoodaIt observed that the accuracy of the DCNN is the best among all other models with 84%. Dataset used was Chest X-ray Images (pneumonia) for classification from the medical database [42].

CONCLUSION

When applied to the healthcare industry, machine learning and deep learning algorithms assist us in a variety of ways. They not only help in better prevention, diagnosis, and treatment of disease, but they are also more efficient and faster than technologies present today. They have been demonstrated not just in theory but also in practice. The healthcare industry is facing more problems and becoming more expensive. Several machine learning algorithms are used to correct these issues. As a result, such integration should be favored for the betterment of mankind. This paper presents various ML techniques for the prediction of various diseases like heart disease, diabetes disease, liver disease as well as DL techniques for detection of malaria and pneumonia disease. To summarise, we conclude that deep learning models and their applications in medicine and healthcare systems have huge potential, especially given the amount and complexity of health data.

REFERENCES

- [1] Shailaja, K., Seetharamulu, B., & Jabbar, M. A. (2018). Machine learning in healthcare: A Review. *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. <https://doi.org/10.1109/iceca.2018.8474918>.
- [2] Beca, J., Cox, P. N., Taylor, M. J., Bohn, D., Butt, W., Logan, W. J., Rutka, J. T., & Barker, G. (1995). Somatosensory evoked potentials for prediction of outcome in acute severe brain injury. *The Journal of Pediatrics*, 126(1), 44-49. [https://doi.org/10.1016/s0022-3476\(95\)70498-1](https://doi.org/10.1016/s0022-3476(95)70498-1).
- [3] Alanazi, Hamdan & Jumah, Mohammed. (2013). A Critical Review for an Accurate and Dynamic Prediction for the Outcomes of Traumatic Brain Injury based on Glasgow Outcome Scale. *Journal of Medical Sciences(Faisalabad)*. 13. 244-252. 10.3923/jms.2013.244.252.
- [4] Precision Medicine Initiative (NIH). <https://www.nih.gov/precision-medicine-initiative-cohort-program> (12 November 2016, date last accessed)..
- [5] Lyman GH, Moses HL. Biomarker tests for molecularly targeted therapies — the key to unlocking precision medicine. *N Engl J Med* 2016;375:4-6.
- [6] Collins FS, Varmus H. A new initiative on precision medicine. *N Engl J Med* 2015;372:793-5.
- [7] Machine Learning, Tom Mitchell, McGraw Hill, 1997.

- [8] Shekhar, S., Schrater, P.R., Vatsavai, R.R., Wu, W., and Chawla, S., Spatial contextual classification and prediction models for mining geospatial data. *IEEE Trans. Multimedia*. 4:174–188, 2002.
- [9] Juhola, M., and Laurikkala, J., Missing values: how many can they be to preserve classification reliability? *Artif. Intell. Rev.* 40:231– 245, 2013.
- [10] Joshi, Tejas & Pramila, M & Chawan, Pramila. (2018). Diabetes Prediction Using Machine Learning Techniques. 2248-9622. 10.9790/9622-0801020913.
- [11] J. Schmidhuber, Deep learning in neural networks: an overview, *Neural networks*. 61 (2015) 85–117.
- [12] W.Y.B. Lim, N.C. Luong, D.T. Hoang, Y. Jiao, Y.-C. Liang, Q. Yang, et al., Federated learning in mobile edge networks: a comprehensive survey, *IEEE Commun. Surv. Tutorials* (2020).
- [13] Rajdhan, Apurb & Agarwal, Avi & Sai, Milan & Ghuli, Poonam. (2020). Heart Disease Prediction using Machine Learning. *International Journal of Engineering Research and*. V9. 10.17577/IJERTV9IS040614.
- [14] Singh, Archana & Kumar, Rakesh. (2020). Heart Disease Prediction Using Machine Learning Algorithms. 452-457. 10.1109/ICE348803.2020.9122958.
- [15] Jagtap, A., Malewadkar, P., Baswat, O. and Rambade, H., 2019. Heart disease prediction using machine learning. *International Journal of Research in Engineering, Science and Management*, 2(2), pp.352-355.
- [16] Jindal, Harshit & Agrawal, Sarthak & Khera, Rishabh & Jain, Rachna & Nagrath, Preeti. (2021). Heart disease prediction using machine learning algorithms. *IOP Conference Series: Materials Science and Engineering*. 1022. 012072. 10.1088/1757-899X/1022/1/012072.
- [17] EL_Jerjawi, Nesreen & Abu-Naser, Samy. (2018). Diabetes Prediction Using Artificial Neural Network. *Journal of Advanced Science*. 124. 1-10.
- [18] Ayon, Safial & Islam, Md. (2019). Diabetes Prediction: A Deep Learning Approach. *International Journal of Information Engineering and Electronic Business*. 11. 21-27. 10.5815/ijieeb.2019.02.03.
- [19] Mujumdar, Aishwarya & Vaidehi, V.. (2019). Diabetes Prediction using Machine Learning Algorithms. *Procedia Computer Science*. 165. 292-299. 10.1016/j.procs.2020.01.047.
- [20] Yahyaoui, Amani & Rasheed, Jawad & Jamil, Akhtar & Yesiltepe, Mirsat. (2019). A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. 10.1109/UBMYK48245.2019.8965556.
- [21] Rahman, A. K. M. & Shamrat, F.M. & Tasnim, Zarrin & Roy, Joy & Hossain, Syed. (2019). A Comparative Study On Liver Disease Prediction Using Supervised Machine Learning Algorithms. 8. 419-422.
- [22] Priya, M.B., Juliet, P.L. and Tamilselvi, P.R., 2018. Performance analysis of liver disease prediction using machine learning algorithms. *International Research Journal of Engineering and Technology (IRJET)*, 5(1), pp.206-211.
- [23] Mohan, Vijayarani. (2015). Liver Disease Prediction using SVM and Naïve Bayes Algorithms.
- [24] Kannapiran, Thirunavukkarasu & Singh, Ajay & Irfan, Md & Chowdhury, Abhishek. (2018). Prediction of Liver Disease using Classification Algorithms. 1-3. 10.1109/CCAA.2018.8777655.
- [25] Shekar, Gautham; Revathy, S.; Goud, Ediga Karthick (2020). [IEEE 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI) - Tirunelveli, India (2020.6.15-2020.6.17)] 2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184) - Malaria Detection using Deep Learning. , (), 746–750. doi:10.1109/ICOEI48184.2020.9143023
- [26] Sriporn, Krit & Tsai, Cheng-Fa & Tsai, Chia-En & Wang, Paohsi. (2020). Analyzing Malaria Disease Using Effective Deep Learning Approach. *Diagnostics* (Basel, Switzerland). 10. 10.3390/diagnostics10100744.
- [27] Kalkan, Soner Can; Sahingoz, Ozgur Koray (2019). [IEEE 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) - Istanbul, Turkey (2019.4.24-2019.4.26)] 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT) - Deep Learning Based Classification of Malaria from Slide Images. , (), 1–4. doi:10.1109/EBBT.2019.8741702
- [28] Arunagiri, Vijayalakshmi & B, Rajesh. (2020). Deep learning approach to detect malaria from microscopic images. *Multimedia Tools and Applications*. 79. 10.1007/s11042-019-7162-y.
- [29] Tilve, A., Nayak, S., Vernekar, S., Turi, D., Shetgaonkar, P. R., & Aswale, S. (2020). Pneumonia detection using deep learning approaches. 2020 International Conference on Emerging Trends in Information Technology and Engineering (Ic-ETITE). <https://doi.org/10.1109/ic-etite47903.2020.152>
- [30] Racic, L., Popovic, T., cacic, S., & Sandi, S. (2021). Pneumonia detection using deep learning based on Convolutional Neural Network. 2021 25th International Conference on Information Technology (IT). <https://doi.org/10.1109/it51528.2021.9390137>
- [31] Ayan, E., & Unver, H. M. (2019). Diagnosis of pneumonia from chest X-ray images using Deep Learning. 2019 Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT). <https://doi.org/10.1109/ebbt.2019.8741582>
- [32] Pant, A., Jain, A., Nayak, K. C., Gandhi, D., & Prasad, B. G. (2020). Pneumonia detection: An efficient approach using Deep Learning. 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). <https://doi.org/10.1109/icccnt49239.2020.9225543>
- [33] Chaurasia, Vikas & Pal, Saurabh. (2013). Data Mining Approach to Detect Heart Diseases. *International Journal of Advanced Computer Science and Information Technology (IJACSIT)*. 2. 56-66.
- [34] R. Atallah and A. Al-Mousa, "Heart Disease Detection Using Machine Learning Majority Voting Ensemble Method," 2019 2nd International Conference on new Trends in Computing Sciences (ICTCS), 2019, pp. 1-6, doi: 10.1109/ICTCS.2019.8923053.
- [35] P. S. Kohli and S. Arora, "Application of Machine Learning in Disease Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-4, doi: 10.1109/CCAA.2018.8777449.
- [36] Kumari, V.A. and R. Chitra, "Classification of Diabetes Disease Using Support Vector Machine", *International Journal of Engineering Research and Applications*, vol.3, pp. 1797-1801, 2013.
- [37] Nahar, Nazmun & Ara, Ferdous. (2018). Liver Disease Prediction by Using Different Decision Tree Techniques. *International Journal of Data Mining & Knowledge Management Process*. 8. 01-09. 10.5121/ijdkp.2018.8201.
- [38] A S Aneeshkumar and Jothi C Venkateswaran. Article: Estimating the Surveillance of Liver Disorder using Classification Algorithms. *International Journal of Computer Applications* 57(6):39-42, November 2012.
- [39] G. Kalipe, V. Gautham and R. K. Behera, "Predicting Malarial Outbreak using Machine Learning and Deep Learning Approach: A Review and Analysis," 2018 International Conference on Information Technology (ICIT), 2018, pp. 33-38, doi: 10.1109/ICIT.2018.00019.
- [40] (arXiv:1907.10418) Aimon Rahman et al 2019 'Improving Malaria Parasite Detection from Red Blood Cell using Deep Convolutional Neural Networks' <https://doi.org/10.48550/arXiv.1907.10418>.
- [41] Elshennawy, N.M.; Ibrahim, D.M. Deep-Pneumonia Framework Using Deep Learning Models Based on Chest X-Ray Images. *Diagnostics* 2020, 10, 649. <https://doi.org/10.3390/diagnostics10090649>.
- [42] K. Jakhar and N. Hooda, "Big Data Deep Learning Framework using Keras: A Case Study of Pneumonia Prediction," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-5, doi: 10.1109/CCAA.2018.8777571.