Academe *Journal of Education & Psychology* *ISSN: 2249-040X*
Volume-16, Issue-1, (Special Issue: June 2026) *Peer Reviewed & Refereed Journal (IF: 6.25)*
PP: 114 - 122 *Journal Website: www.academejournal.in*

APPLICATION AND IMPACT OF MACHINE LEARNING IN HEALTH SCIENCES

Asha N

Associate Professor
Department of Computer Science, Government Science College, Bangalore, India

Sowmya S B

Associate Professor
Department of Mathematics, Government Science College, Bangalore, India

Sharada Devi J N

Associate Professor
Department of Zoology and Genetics, Government Science College, Bangalore, India

ABSTRACT

In health sciences, machine learning is obviously very crucial because it facilitates automated diagnosis, personalised treatment, disease prediction, and intelligent clinical decision support. Support Vector Machines (SVM), Random Forests (RF), K-nearest Neighbor, and Deep Learning are some of the applications of machine learning (ML) in the medical domain.Detecting the disease prior, drug discovery, and analysis of medical imaging all stand to benefit greatly from machine learning techniques. Discusses issues such the need for big labelled datasets, model bias, and data privacy. The paper emphasizes on results that show the development of digital health in the future will be greatly aided by machine learning.

Keywords: Predictive analytics, diagnosis, medical imaging, deep learning, and machine learning.

INTRODUCTION

Machine learning (ML), a subfield of artificial intelligence (AI), is one of the most innovative technologies in healthcare today. The need for automated systems that can extract valuable insights has arisen due to the exponential growth of clinical, genetic, and imaging data. Medical decision-making is supported by machine learning (ML), which offers methods for analyzing complicated information and improving diagnostic precision.

ML applications in healthcare include:

- Disease prediction and early diagnosis
- Analysis of X-ray, MRI, CT, and ultrasound images
- Patient monitoring using wearable devices
- Personalized treatment recommendations
- Predictive modelling for epidemics and hospital readmissions

Issues with data accessibility, privacy, and interpretability still exist in spite of these developments. This study focuses on the analysis of machine learning applications in the health sciences and identifies potential areas for further investigation.

LITERATURE REVIEW

Several studies have demonstrated the benefits of ML in medical applications. Deep learning models, particularly Convolutional Neural Networks (CNNs), have demonstrated exceptional accuracy in identifying anomalies, tumors, and fractures in medical pictures. Random Forest models have shown strong results in heart disease risk prediction by analysing clinical parameters. Recurrent Neural Networks (RNNs) have been used to model disease progression using time-series patient data.The author has underlined the necessity of incorporating a number of machine learning principles into medical curricula in order for medical practitioners to successfully direct and analyse medical research (Abdullah Alanazi,2022).

The literature that is currently available highlights that deep learning models perform well for unstructured medical data, such as images, signals, and text, but classic machine learning techniques are still useful. (Samuel-Soma M et.al, 2024).

The article discusses the retrospective application of machine learning and often outlines proof-of-concept techniques to enhance medical care. We predict that by identifying and eliminating significant translational barriers, such as real-time access to clinical data, data security, physician approval of "black box" generated results, and performance evaluation, a fundamental shift in medical practice will be possible, where specialized tools will assist the healthcare team in providing better patient care. (David et.al, 2020).

The paper emphasises the use of machine learning in radiology, genetics, electronic health records, and neuroimaging, among other areas of healthcare. We also offer recommendations for future uses while briefly discussing the dangers and difficulties of applying machine learning to healthcare, including issues with system privacy and ethics (Habehh H et.al, 2021).

Data-focused issues in ML deployment, highlighting the necessity of effectively supplying data to ML models for prompt clinical predictions and taking into consideration natural data shifts that may worsen model performance (Zhang, A et.al 2022).

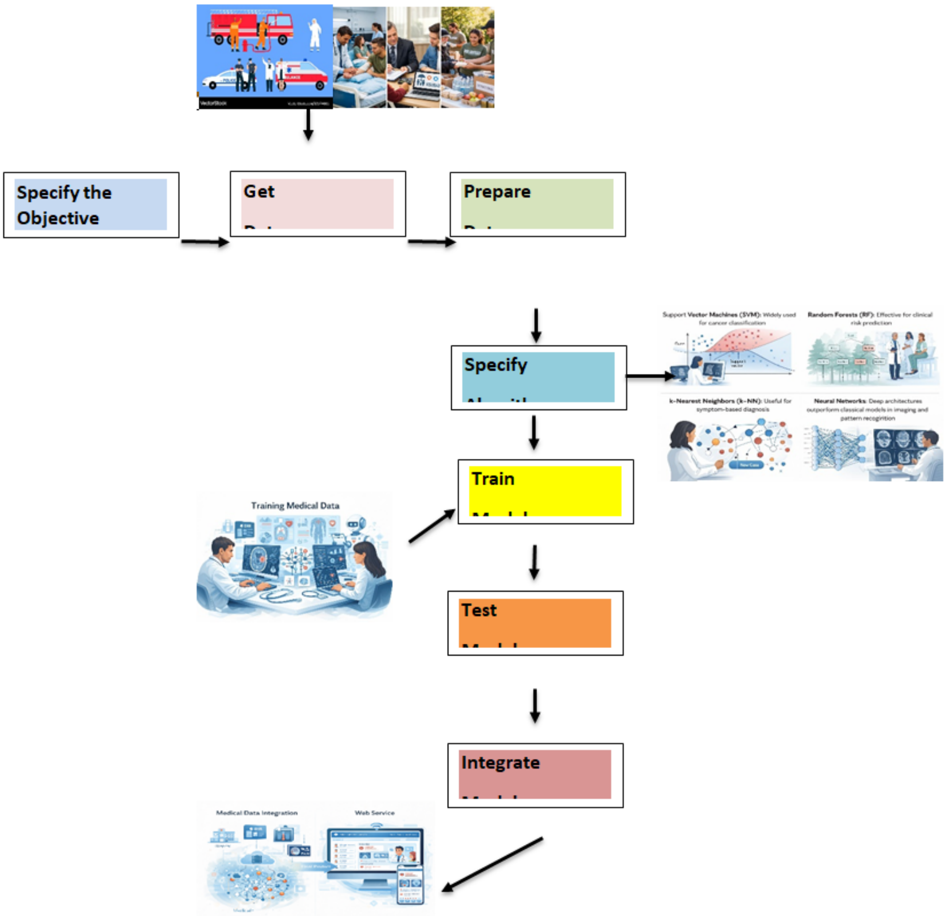
METHODOLOGY

A web service would be the end result of training the medical data in the following subsequent phases.

1. **Specify the Objective:** Determine the issue that needs to be resolved and establish a clear goal.
2. **Get data:** Gather information from medical facilities, insurance providers, social service organisations, police, and fire departments.
3. **Prepare data:** Creating routines to clean, match, and blend data is an essential part of the preparation process.
4. **Specify Algorithm:** The kind of data and the problem to be solved are taken into consideration while choosing an algorithm.
5. **Train Model:** The algorithm is set up with the necessary parameters when data is provided as input. A portion of the data can be used to train the model.
6. **Testing the Model:** The correctness of the model is tested using the remaining data. Improvements can be made iteratively in the "Train model" and "Specify Algorithm" phases based on the findings.

7. **Integrate Model:** To enable applications to utilize the model, publish the prepared experiment as a web service.

Fig 1.1 : Framework of Medical Data Analysis



DATA COLLECTION AND PREPROCESSING

Healthcare datasets consist of:

1. Electronic Health Records (EHR)
2. Data from medical imaging
3. Proteomic and genomic information
4. Signals from wearable sensors

1. COLLECTING ELECTRONIC HEALTH RECORDS (EHRS)

The EHRs data can be collected from the following sources:

A. **Hospitals & Clinics**

Patient records (diagnosis, prescriptions, lab results)

B. **EHR Systems**

Systems like Epic, Cerner, or hospital databases

C. **Public Datasets**

MIMIC-III / MIMIC-IV (critical care data)

WHO or government health portals

D. **Wearable Devices & Apps**

Heart rate, activity, sleep data

2. DATA FROM MEDICAL IMAGING

The sources of medical imaging data are as follows

A. **Radiology departments generate images like:**

- X-rays report
- CT scans report
- MRI scans report
- Ultrasound report

B. **Public Medical Imaging Datasets**

- **NIH Chest X-ray dataset**
- **TCIA (The Cancer Imaging Archive)**
- **Kaggle medical datasets**

3. PROTEOMIC AND GENOMIC INFORMATION

The sources of Proteomic and genomic information data are as follows

A. Experimental Techniques

- **Mass Spectrometry (MS)**
  - Identifies and quantifies proteins
- **2D Gel Electrophoresis**
  - Separates proteins

B. Public Proteomic Databases

- **UniProt** (protein sequences & functions)
- **Protein Data Bank (PDB)**
- **PRIDE Archive** (proteomics datasets)

4. SIGNALS FROM WEARABLE DEVICES

A. Consumer Wearables

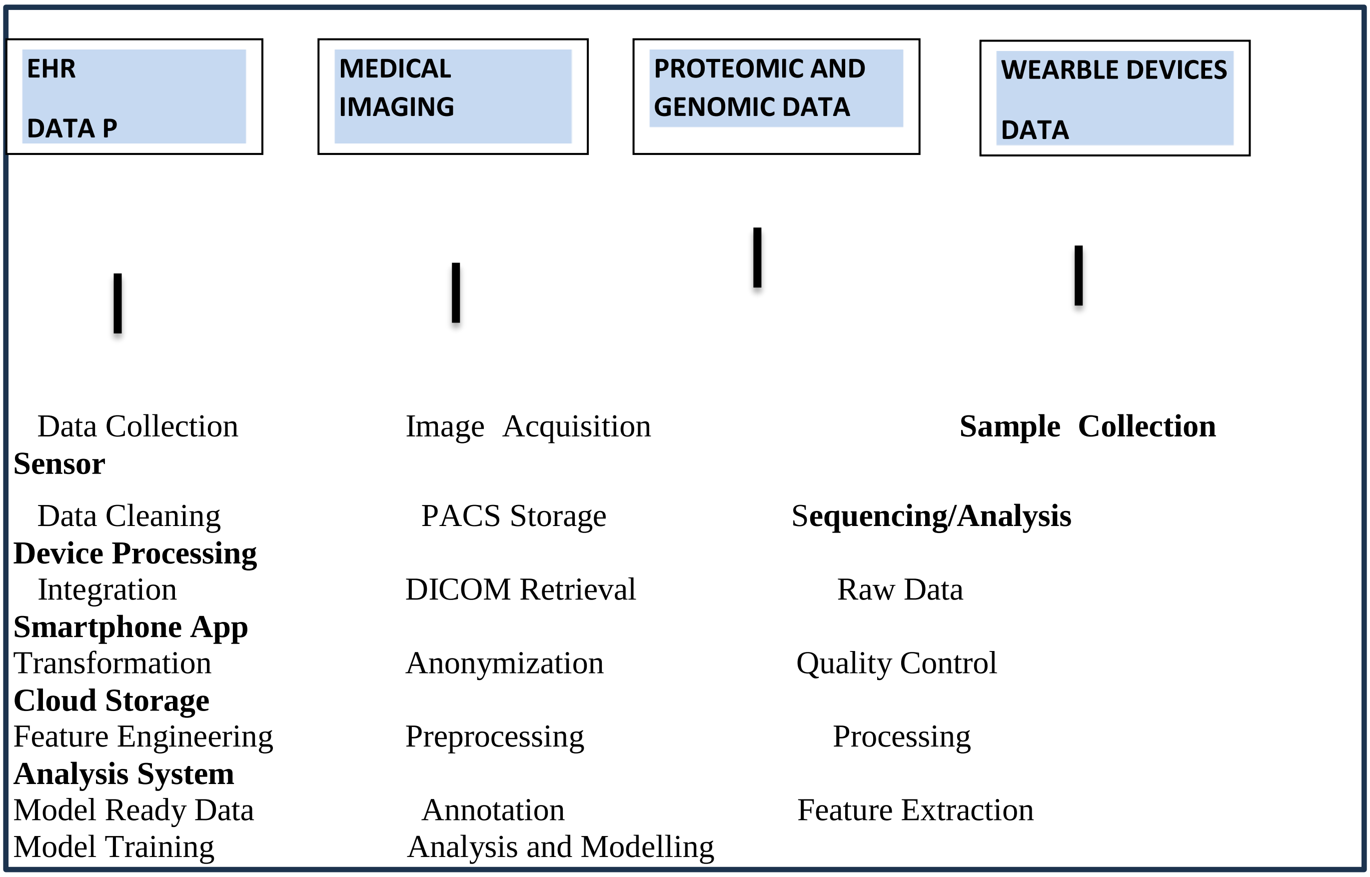
- Smartwatches, fitness trackers
- Track:
  - Heart rate
  - Steps
  - Sleep patterns
  - Calories burned

B. Medical Wearables

- ECG monitors
- CGMs, or continuous glucose monitors
- Monitors for blood pressure
- Oximeters for pulses

Data Preprocessing

Fig 1.2: Data Preprocessing Phases for each Medical data source



IMPLEMENTATION OF MACHINE LEARNING ALGORITHMS

1. **Random Forest Method:**

   A case study to investigate and identify a patient who is sensitive to high diabetes is made possible by the random forest method's ability to handle large datasets and facilitate feature extraction, which is used for clinical risk prediction. The following Fig 1.3 depicts tree representation of Random Forest Method.

Example:

Fig 1.3: Random Forest Method for Clinical Data

Glucose <140?

/ \

No Yes

/ \

Age > 45?BMI > 30?

/ \ / \

Yes No Yes No

/ \ / \

Diabetes No Diabetes No

2. **Support Vector Machines**

   Support Vector Machines easily classify two classes by using a decision boundary. A patient record with a tumour can be identified as Benign(non-cancerous) or Malignant (Cancerous). The target variable is set to Malignant = 1 and Benign = 0, and the sample attributes include radius, texture, perimeter, area, and smoothness.

The Algorithm is as follows

Algorithm for Tumor Data Set

Step 1: Load Tumor data set

Step 2: Normalize the data Set

Step 3: Divide the data set into training data (80%) and test data (20%)

Step 4: Call SVM Model

Step 5: Use training data to train the model

Step 6: Estimate Test Data Labels

Step 7: Assess Performance

Step 8: Model ready for tumor Prediction

3. **K-Nearest Neighbor Classifier**

   KNN is basically used for symptom-based diagnosis, like predicting whether a patient is diagnosed with typhoid or not. The parameters used are fever, fatigue, abdominal pain, weight loss, headache, diarrhoea. The following sample data set is used for training the algorithm.

Patient	**Body Temperature in Fahrenheit**	**Headache (0-10)**	**Abdominal Pain(0-10)**	**Fatigue (0-10)**	**Diarrhoea Yes=1,No=0**	**Clas Typ 1 No Typ 0**
PA1	**101**	**4**	**5**	**3**	**0**	**0**
PA2	**8**	**3**	**7**	**7**	**1**	**1**
PA3	**100**	**3**	**2**	**4**	**0**	**0**
PA4	**103**	**7**	**8**	**9**	**1**	**1**

Table 1.1: Sample Typhoid Data set

The Algorithm is as follows

Algorithm for Typhoid Data set

Step 1: Load the Typhoid data set

Step 2: Normalize the data

Step 3: Choose K Value

Step 4: Estimate the Euclidean Distance for all the data in the data set

Step 5: Choose the K nearest neighbors

Step 6: Count the class Labels

Step 7: Assign Majority Classes

Step 8: Predict the output

Step 9: Evaluate Model Performance

The Euclidean Distance formula is used to select the K value, where k is the number of neighbors chosen. When the input is given to the model, the training dataset is trained, so when we select the number of neighbors as k=3 or 5, based on the k value, the model will predict whether the patient is diagnosed with typhoid.

Therefore, from the above sample in Table 1.1 the patients PA2 and PA4 are diagnosed with Typhoid for k=3

4. Neural Networks

The Convolutional Neural Network in the paper is primarily built to predict medical images like x-rays, MRI scans, and CT scans. The MRI scans of Brain tumor data is taken as an input to predict the presence of tumor.

The Algorithm is as follows

Step 1: Load the MRI data set

Step 2: Normalize images (resize,remove noice etc)

Step 3: Label the data Tumor/ No Tumor

Step 4: Divide the Training data set to 80% and Test data set to 20%

Step 5: Build CNN Model

Step 6: Train the Model using Training data

Step 7: Validate the Model on Test data

Step 8: Input new MRI image

Step 9: Predict the presence of Tumor

Step 10: Output result

The model of a convolutional neural network is trained so that ,if the number of hidden layers is increased more accurate result will be obtained with appropriate selection of bias value and sigmoid function.

CHALLENGES

Machine Learning Algorithms' main drawbacks in the medical field are:

- A lack of labelled, high-quality medical data
- Security and privacy issues
- Algorithmic bias impacting the fairness of diagnoses
- Clinical validation is required.
- Deep learning models' high computational cost
- The inability to comprehend "black-box" models

CONCLUSION

By facilitating accurate diagnosis, predictive analytics, and individualised care, machine learning has transformed the health sciences. The main ML approaches, uses, difficulties, and new developments in healthcare were examined in this research. Future studies ought to concentrate on: Explainable AI (XAI)

• Federated learning, or privacy-preserving machine learning

• Extensive multimodal medical databases • Combining ML with IoT, robotics, and telemedicine

ML will play a significant part in next-generation medical technology and continue to bolster the global healthcare ecosystem.

REFERENCES

1. Abdullah Alanazi, Using machine learning for healthcare challenges and opportunities, Informatics in Medicine Unlocked, Volume 30, 2022, 100924, ISSN 2352-9148, https://doi.org/10.1016/j.imu.2022.100924.
2. Samuel-Soma M. Ajibade, Gloria Nnadwa Alhassan, Abdelhamid Zaidi, Olukayode Ayodele Oki, Joseph Bamidele Awotunde, Emeka Ogbuju, Kayode A. Akintoye, "Evolution of machine learning applications in medical and healthcare analytics research: A bibliometric analysis, Intelligent Systems

with Applications", Volume 24, 2024, 200441, ISSN 2667-3053, https://doi.org/10.1016/j.iswa.2024.200441.

3. DavidBen-Israel D, Jacobs WB, Casha S, Lang S, Ryu WHA, de Lotbiniere-Bassett M, Cadotte DW. The impact of machine learning on patient care: A systematic review. Artif Intell Med. 2020 Mar;103:101785. doi: 10.1016/j.artmed.2019.101785.

4. Epub 2019 Dec 31. PMID: 32143792.

5. Habehh H, Gohel S. Machine Learning in Healthcare. Curr Genomics. 2021 Dec 16;22(4):291-300. doi: 10.2174/1389202922666210705124359. PMID: 35273459; PMCID: PMC8822225.

6. Zhang, A., Xing, L., Zou, J. et al. Shifting machine learning for healthcare from development to deployment and from models to data. Nat. Biomed. Eng **6**, 1330–1345 (2022). https://doi.org/10.1038/s41551-022-00898-y.