Undergraduate Conference on Intelligent Computing and Systems (UCICS 2025)
26-27 February, 2025; Varendra University, Rajshahi, Bangladesh

Machine Learning in Healthcare: Key Applications and Insights from Recent Studies

Touhid Alam*, Md All Shahria*, Sanjeda Dewan Mithila*, and Mohammad Sakib Mahmood†
*Dept. of Computer Science, American International University-Bangladesh, Dhaka, Bangladesh
†Dept. of Computer Science, Missouri State University, Springfield, Missouri, USA
**22-46330-1@student.aiub.edu; *22-48639-3@student.aiub.edu; *22-48337-3@student.aiub.edu; †mm974s@missouristate.edu*

***Abstract*—This manuscript provides a systematic review of machine learning (ML) applications in healthcare, focusing on disease prediction, medical imaging, and sentiment analysis. Supervised ML is extensively used for disease diagnosis and prediction. These models can play a crucial role in disease diagnosis, identifying patterns, and decision-making. Deep learning (DL), has enabled advances in medical imaging, allowing the identification of complex patterns in diagnostic images for diseases such as cancer and infectious diseases. Pre-trained models and custom architectures have been fine-tuned to enhance their performance in clinical applications. In addition, sentiment analysis with Natural Language Processing (NLP) techniques has been used to analyze social media data, clinical texts, and audiovisual records to detect psychological and neurological conditions. Despite significant progress, challenges such as limited dataset size, lack of diversity, interpretability of complex models, and biases in data and algorithms persist. This paper highlights the applications of these ML techniques in healthcare and examines their potential to improve clinical decision-making while addressing existing limitations to enhance the reliability and applicability of these technologies in real-world settings.**

***Index Terms*—Healthcare, Machine Learning, Deep Learning, Natural Language Processing, Predictive Modeling, Sentimental Analysis.**

I. Introduction

ML is a branch of AI, that identifies patterns, predicts outcomes, and generates actionable insights from data across various domains [1]. ML encompasses sub-fields like supervised learning, which uses labeled data for predictions, and DL, which employs neural networks to learn complex features. NLP generates human language, powering advancements in disease diagnosis, speech recognition, and language translation [2]. In recent years, the global popularity of using ML to solve problems has significantly increased due to the vast availability of data [3].

ML is increasingly being utilized across diverse sectors, including agriculture, finance, and even healthcare, to solve complex challenges [4]. Data in healthcare, especially Electronic Health Records (EHR) faces various issues including timeliness of the data, handling large volumes of data, and biases in decision making [5]. ML can address these issues by offering efficient data processing and continuous improvement. Nevertheless, some issues still remain such as lack of diversity and interpretability. This paper aims to highlight the usage of ML in healthcare as explored in recent studies, while also addressing the challenges faced in implementing these technologies.

Section II of the study gives a brief introduction to types of machine learning, including supervised learning, DL, and NLP. Section III discusses their healthcare applications in recent studies and Section IV presents an overview of the studies and future directions

II. Literature Review

This section of the paper gives a brief discussion about the concepts of supervised machine learning, DL and NLP.

A. Supervised Machine Learning

Supervised learning is an ML paradigm where the model learns to predict outcomes based on labeled input data. It involves training a model using a dataset where each input is associated with a corresponding target output. The goal is to learn from inputs to outputs, enabling the model to make accurate predictions on new, unseen data. Supervised learning is commonly applied to tasks such as classification, where the output is a discrete label, and regression, where the output is a continuous value. The performance of these models is evaluated using evolution metrics depending on the task. A key strength of supervised learning is its ability to generalize from training data to make reliable predictions, provided the data is representative and well-structured [6].

B. Deep Learning

DL is a subset of ML that uses neural networks with multiple layers to model complex patterns in large datasets. It is particularly effective for tasks involving high-dimensional data, such as iimages Key techniques in DL include Convolutional Neural Networks (CNNs), used primarily for image processing, and Recurrent Neural Networks (RNNs), which excel in sequence data like time series and natural language. Applications of DL in medical include object localization, image registration, classifications and detections, and segmentations [7].

C. Natural Language Processing

Natural Language Processing (NLP) is a field of ML focused on enabling machines to understand and generate human language. It allows for tasks such as text analysis, language translation, and sentiment understanding. NLP encompasses a wide range of functionalities, including tokenization, which breaks text into smaller units like words or sentences; Named Entity Recognition (NER), which identifies and classifies entities such as names, dates, and locations; and Part-of-Speech (POS) tagging, which labels words according to their grammatical roles. Additionally, NLP techniques enable sentiment analysis to determine emotional tone, machine translation to convert text between languages, and text summarization to condense large documents. These functionalities make NLP a powerful tool for applications like chatbots, search engines, and automated content analysis [8].

III. Applications of Machine Learning in Healthcare and Limitations

This section of the paper aims to give highlight of applications of ML in healthcare through data-driven approaches, predictive models, and sentimental analysis along with the limitations of the highlighted studies. Fig. 1 shows the highlighted applications of ML in healthcare. Furthermore, Table I shows a summary of the application of ML in healthcare with their findings and limitations.

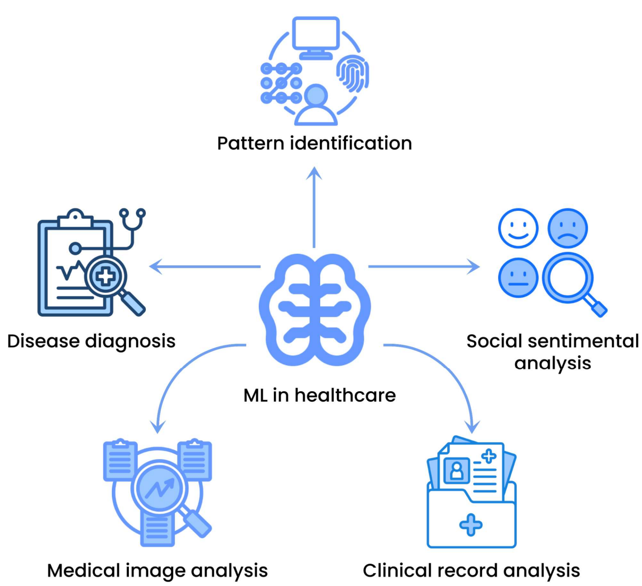

Fig. 1. Applications of Machine Learning in Healthcare.

A. Supervised Machine Learning

Supervised machine learning plays a pivotal role in healthcare by leveraging labeled data to develop predictive models for accurate diagnosis. Abdulhadi and A. Al-Mousa developed ML models for predicting the presence of diabetics using the Pima Indian dataset [9]. They were able to achieve an accuracy of 82% using a random forest classifier. Furthermore, they also developed a voting classifier using logistic regression, linear discriminant analysis, and random forest classifier which also achieved an accuracy of 80%. However, their research was performed on a specific ethnic group with a small sample size of 768 instances. J. Wu and C. Hicks implemented four different algorithms including SVM, KNN, NGB, and DT in classifying TNBC vs non-TNBC type breast cancer [10]. The SVM algorithm resulted in 90% accuracy, 87% recall, and 90% specificity. Nevertheless, they used a small and unbalanced sample size and the proposed machine learning model did not address the problem of classifying the multiple subtypes of TNBC and non-TNBC. D. A. Debal and T. M. Sitote developed ML for chronic kidney disease prediction with an accuracy of 99.8% for binary classification and 82.56% accuracy for multi-class classification [11]. Collecting patient data from St. Paulo's Hospital in Ethiopia, they used Univariate Feature Selection (UFS) and Recursive Feature Elimination with Cross-Validation (RFECV) to select the most relevant features. In their study, the SVM, RF with REFCV, and XGBoost machine learning models performed the best. They also identified features such as serum creatinine, blood urea nitrogen, and hemoglobin as the most important features. Nevertheless, their study was limited to particular area-based data and multi-class classification resulted much lower than binary-class classification. U. Nagavelli, D. Samanta, and P. Chakrabort were able to detect heart disease with an accuracy of 95.9% using the XGBoost algorithm [12]. They employed four different algorithms including weighted-based Naïve Bayes, SVM with Duality Optimization (DO), and XGBoost. However, their study faced some limitations including dataset size and diversity, limitations around data availability, and the need to expand the dataset with more attributes. H. M. Farghaly, M. Y. Shams, and T. Abd El-Hafeez used 859 patients data with 12 clinical and demographic features collected from Egypt to predict Hepatitis C Virus [13]. They implemented four different algorithms where the Random Forest model with hyperparameter tuning resulted in 94.04% accuracy without feature selection and 94.88% accuracy using only four features with the SFS feature selection technique. Nevertheless, their study used a relatively small dataset with only 11 features. More features may be needed to improve the prediction. M. N. Hossain et. al. developed Gradient Boosting and Random Forest-based ML models to predict lung cancer with an accuracy of 89% and 87% respectively [14]. They also implemented neural networks which reached 90% accuracy but presented interpretability limitations. They also identified features like smoking history, age, and family history as important predictors of lung cancer risk. However, their study lacks the advanced interpretability techniques for complex models like neural networks and the need for more diverse real-world clinical settings. S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian were able to predict the recurrence of thyroid cancer with an accuracy of 96.6% with neural network [15]. They implemented cross-validation with grid search to optimize the model performance. However, their model might potentially overfit due to the internal validation dataset, limitations of the ATA model, and fewer cases of specific pathological subtypes.

TABLE I
SUMMARY OF VARIOUS MACHINE LEARNING AND DEEP LEARNING APPLICATIONS IN HEALTHCARE

Author(s)	Application	Best Algorithm	Accuracy	Limitations
Abdulhadi et al. [9]	Diabetes Prediction	Random Forest	82%	Small sample size, specific ethnic group
J. Wu et al. [10]	Breast Cancer Classification (TNBC vs non-TNBC)	SVM	90%	Small, unbalanced sample, limited subtypes
D. A. Debal et al. [11]	Chronic Kidney Disease Prediction	SVM	99.8%	Limited to a specific area, lower accuracy for multi-class
U. Nagavelli et al. [12]	Heart Disease Detection	XGBoost	95.9%	Dataset size and diversity, limited attributes
H. M. Farghaly et al. [13]	Hepatitis C Virus Prediction	Random Forest	94.88%	Small dataset, limited features
M. N. Hossain et al. [14]	Lung Cancer Prediction	Neural Networks	90%	Lack of interpretability techniques, real-world diversity
S. Borzooei et al. [15]	Thyroid Cancer Recurrence Prediction	Neural Network	96.6%	Potential overfitting, limited pathological subtypes
T. D. Nguyen et al. [16]	Retinal Disease Classification	ResNet152	96.47%	Impact of preprocessing not mentioned, limited clinical outcome correlation
M. A. Talukder et al. [17]	COVID-19 Detection	EfficientNetB4	100%	Overfitting risk due to small dataset
A. U. Ibrahim et al. [18]	Pneumonia Detection	AlexNet	96%	Larger datasets and hybrid models needed
A. A. Shah et al. [19]	Lung Cancer Prediction	2D CNN	95%	Need for 3D-based CNN architecture and diverse data
S. K. Singh et al. [20]	Prostate Cancer Detection	3D CNN	87%	ROIAlign layer may degrade scale information
M. M. Rahman et al. [21]	Breast Cancer Detection	YOLO-based CNN	93.0%	Small dataset, limited generalization
S. R. Krishna et al. [22]	Skin Cancer Diagnosis	IMLT-DL	99.7%	Computational complexity, potential biases
M. E. Basiri et al. [23]	Sentiment Analysis of COVID-19 Tweets	NLP-based DL models	-	Did not include global COVID-19 news and statistics
P. Mukherjee et al. [24]	ASD Symptom Detection	BERT	83%	Need for quantum models and more data
Y. Yang et al. [25]	Cancer Trial Eligibility Classification	BERT	99%	Small sample size for exclusion criteria
H. Xu et al. [26]	Early Detection of Dementia	Pre-trained NLP Models	87.5%	Small sample size, potential sampling bias
K. Zandbiglari et al. [27]	ADRD and Suicide Ideation Detection	BERT	99%	No mention of data limitations or biases

B. Deep Learning and Medical Imaging

Beside Supervised Machine Learning, Deep Learning, or specifically CNN has been broadly used for image classification purposes in medical. T. D. Nguyen et al. proposed DL models for retinal disease classification including the ResNet152 model with an AUC score of 96.47% using ultra-wide-field fundus images [16]. However, the impact of the preprocessing steps on the model accuracy was not mentioned. The research focused on technical aspects of model development, and the direct correlation with clinical outcomes is limited. M. A. Talukder et. al. fined tuned the EfficientNetB4 model to achieve an impressive accuracy of 100% in detecting COVID-19 cases from X-ray images [17]. Their study highlighted the fine-tuning process was crucial in enhancing the performance of the deep learning models. A. U. Ibrahim et. al. proposed a pre-trained AlexNet-based model for pneumonia prediction with an accuracy of over 96% across various cases [18]. Their model performed efficient performance across four-way classification tasks for pneumonia. However, the use of larger datasets and hybrid CNN models can improve performance. A. A. Shah et. al. proposed 2D based CNN model with an accuracy of 95% to predict lung cancer [19]. The authors combined three different CNN models to improve the accuracy and reduce the false positive rate. The authors also suggested using 3D-based CNN architecture and including more diverse data to improve model accuracy. S. K. Singh et. al. proposed a 3D CNN-based model which achieved an accuracy of 87%, specificity of 85%, and sensitivity of 89% to successfully predict prostate cancer based on MRI images [20]. They combined different MRI modalities to improve the model performance. Despite that, the "ROIAlign" layer may result in a degradation of scale information that could be relevant for histopathology and is a potential area for improvement. M. M. Rahman et. al. proposed sixteen hidden layer-based custom CNN models based on the YOLO model which was able to predict breast cancer with an accuracy of 93.0% [21]. However, their model is trained on the MIAS dataset with only 330 images, which may limit the model's generalization ability. S. R. Krishna et. al. proposed a novel deep learning model named IMLT-DL which combines multilevel deep learning techniques to diagnose skin cancer with an accuracy of 99.7% [22]. The model outperformed other pre-trained models by 0.992%. However, the potential biases in the model and applications were not addressed and the computational complexity of the IMLT-DT model is very high as shown in the study.

C. Sentimental Analysis with NLP

M. E. Basiri et. al. used four Nmodelssed models for the sentimental analysis of COVID-19 tweets collected from different countries [23]. Their main findings were the rise of negative sentiment happened around the rise of new cases of COVID-19 with every country having unique sentiment patterns. Although, their study did not consider the global COVID-19 news and statistics. P. Mukherjee et. al. used BERT and ChatGPT models to detect ASD symptoms using dialogues from parents to children with ASD using social media [24]. The model BERT achieved an accuracy of 83%. The authors also suggested to use of quantum machine learning models and to collection of more accurate ASD-related data. Y. Yang et. al. implemented six BERT-based models on the PROTECTOR1 database to classify cancer trial eligibility [25]. Their study resulted in an accuracy of 99% in text classification. The key finding of the study is that many trials did not disclose certain key exclusion criteria on ClinicalTrials.gov, but only mentioned them in study protocols or publications. The authors mentioned the small sample size for some exclusion criteria, which may limit generalizability. H. Xu et. al. used statistical analysis and pre-trained NLP models to early detect dementia using data from audio and video [26]. The study achieved 87.5% accuracy in distinguishing individuals with probable Alzheimer's disease from a control group of the same age. The paper emphasizes the potential of using linguistic markers and the new "information unit" feature. Despite that, their study contains a small sample size and potential sampling bias. K. Zandbiglari et. al. identified ADRD patients and those with suicide ideation from the MIMIC-III and MIMIC-IV datasets using relevant ICD codes using BERT [27]. BERT achieved better than other models with precision of 99% and recall of 98%.

IV. Conclusion and Future Direction

This study highlights recent research on ML applications in disease diagnosis, prediction, and medical sentiment analysis, focusing on their findings, implications, and limitations. It emphasizes the potential of ensemble-based supervised learning models, which have often outperformed other algorithms in disease prediction, and the efficiency of pre-trained deep learning models like AlexNet and ResNet52 in image classification, particularly when proper image preprocessing is applied. Additionally, BERT-based models have shown significant promise in sentiment analysis, enhancing the ability to analyze med-ical texts and patient feedback. Despite these advancements, challenges such as small sample sizes, limited data diversity, and overfitting remain common, which hinder model gener-alization and interpretability. Future research should prioritize the use of diverse datasets, implement techniques to address class imbalance, and explore multi-modal learning approaches to enhance model robustness. Moreover, ensuring data privacy and security through methods like federated learning is crucial to maintaining patient confidentiality while utilizing decentral-ized data. Additionally, more emphasis is needed on improving model interpretability, particularly for complex DL models, to foster trust and acceptance in clinical settings.

References

[1] C. Zhang and Y. Lu, "Study on Artificial Intelligence: The State of the Art and Future Prospects," Journal of Industrial Information Integration, vol. 23, p. 100224, 2021.

[2] K. Sharifani and M. Amini, "Machine Learning and Deep Learning: A Review of Methods and Applications," World Information Technology and Engineering Journal, vol. 10, no. 7, pp. 3897–3904, 2023.

[3] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications, and Research Directions," SN Computer Science, vol. 2, no. 3, p. 160, 2021.

[4] A. Alanazi, "Using Machine Learning for Healthcare Challenges and Opportunities," Informatics in Medicine Unlocked, vol. 30, p. 100924, 2022.

[5] L. Hong, M. Luo, R. Wang, P. Lu, W. Lu, and L. Lu, "Big Data in Health Care: Applications and Challenges," Data and Information Management, vol. 2, no. 3, pp. 175–197, 2018.

[6] B. Mahesh and A. Batta, "Machine learning algorithms-a review," Int. J. Sci. Res. [Online], vol. 9, no. 1, pp. 381-386.

[7] S. Suganyadevi, V. Seethalakshmi, and K. Balasamy, "A review on deep learning in medical image analysis," Int. J. Multimed. Inf. Retrieval, vol. 11, no. 1, pp. 19-38, 2022.

[8] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," Multimedia Tools Appl., vol. 82, no. 3, pp. 3713-3744, 2023.

[9] N. Abdulhadi and A. Al-Mousa, "Diabetes detection using machine learning classification methods," in 2021 International Conference on Information Technology (ICIT), 2021, pp. 350–354.

[10] J. Wu and C. Hicks, "Breast cancer type classification using machine learning," Journal of Personalized Medicine, vol. 11, no. 2, p. 61, 2021.

[11] D. A. Debal and T. M. Sitote, "Chronic kidney disease prediction using machine learning techniques," Journal of Big Data, vol. 9, no. 1, p. 109, 2022.

[12] U. Nagavelli, D. Samanta, and P. Chakraborty, "Machine Learning Technology-Based Heart Disease Detection Models," Journal of Healthcare Engineering, vol. 2022, no. 1, p. 7351061, 2022.

[13] H. M. Farghaly, M. Y. Shams, and T. Abd El-Hafeez, "Hepatitis C Virus prediction based on machine learning framework: a real-world case study in Egypt," Knowledge and Information Systems, vol. 65, no. 6, pp. 2595–2617, 2023.

[14] M. N. Hossain, N. Anjum, M. Alam, M. H. Rahman, M. S. Taluckder, M. N. V. Al Bony, S. S. I. Rishad, and A. H. Jui, "Performance of machine learning algorithms for lung cancer prediction: A comparative study," International Journal of Medical Science and Public Health Research, vol. 5, no. 11, pp. 41–55, 2024.

[15] S. Borzooei, G. Briganti, M. Golparian, J. R. Lechien, and A. Tarokhian, "Machine learning for risk stratification of thyroid cancer patients: a 15-year cohort study," European Archives of Oto-Rhino-Laryngology, vol. 281, no. 4, pp. 2095–2104, 2024.

[16] T. D. Nguyen, D.-T. Le, J. Bum, S. Kim, S. J. Song, and H. Choo, "Retinal disease diagnosis using deep learning on ultra-wide-field fundus images," Diagnostics, vol. 14, no. 1, p. 105, 2024.

[17] M. A. Talukder, M. A. Layek, M. Kazi, M. A. Uddin, and S. Aryal, "Empowering COVID-19 detection: Optimizing performance through fine-tuned EfficientNet deep learning architecture," Computers in Biology and Medicine, vol. 168, p. 107789, 2024.

[18] A. U. Ibrahim, M. Ozsoz, S. Serte, F. Al-Turjman, and P. S. Yakoi, "Pneumonia classification using deep learning from chest X-ray images during COVID-19," Cognitive Computation, vol. 16, no. 4, pp. 1589–1601, 2024.

[19] A. A. Shah, H. A. M. Malik, A. M. Muhammad, A. Alourani, and Z. A. Butt, "Deep learning ensemble 2D CNN approach towards the detection of lung cancer," Scientific Reports, vol. 13, no. 1, p. 2987, 2023.

[20] S. K. Singh, A. Sinha, H. Singh, A. Mahanti, A. Patel, S. Mahajan, A. K. Pandit, and V. Varadarajan, "A novel deep learning-based technique for detecting prostate cancer in MRI images," Multimedia Tools and Applications, vol. 83, no. 5, pp. 14173–14187, 2024.

[21] M. M. Rahman, M. Z. B. Jahangir, A. Rahman, M. Akter, M. A. N. Al Nasim, K. D. Gupta, and R. George, "Breast cancer detection and localizing the mass area using deep learning," Big Data and Cognitive Computing, vol. 8, no. 7, p. 80, 2024.

[22] S. R. Krishna, A. Gudur, S. Jain, S. Deivasigamani, M. Tiwari, K. G. S. Venkatesan, and others, "Deep learning for automatic diagnosis of skin cancer using dermoscopic images," Journal of Artificial Intelligence and Technology, vol. 4, no. 2, pp. 114–123, 2024.

[23] M. E. Basiri, S. Nemati, M. Abdar, S. Asadi, and U. R. Acharrya, "A novel fusion-based deep learning model for sentiment analysis of COVID-19 tweets," Knowledge-Based Systems, vol. 228, p. 107242, 2021.

[24] P. Mukherjee, R. S. Gokul, S. Sadhukhan, M. Godse, and B. Chakraborty, "Detection of Autism Spectrum Disorder (ASD) from natural language text using BERT and ChatGPT models," Int. J. Adv. Comput. Sci. Appl., vol. 14, no. 10, 2023.

[25] Y. Yang, S. Jayaraj, E. Ludmir, and K. Roberts, "Text classification of cancer clinical trial eligibility criteria," in AMIA Annu. Symp. Proc., vol. 2023, pp. 1304, 2024.

[26] H. Xu, X. Wu, and X. Liu, "A measurement method for mental health based on dynamic multimodal feature recognition," Frontiers in Public Health, vol. 10, p. 990235, 2022.

[27] K. Zandbiglari, H. R. Hasanzadeh, P. Kotecha, R. Sajdeya, A. J. Goodin, T. Jiao, F. I. Adiba, M. T. Mardini, J. Bian, and M. Rouhizadeh, "A natural language processing algorithm for classifying suicidal behaviors in Alzheimer's disease and related dementia patients: development and validation using electronic health records data," medRxiv, pp. 2023–07, 2023.