ADVANCEMENTS AND RESEARCH GAPS IN HEALTHCARE PREDICTIVE ANALYTICS USING MACHINE LEARNING AND DEEP LEARNING: A COMPREHENSIVE REVIEW

1st Nitin Varshney, 2nd Yash Vasoya, 3rd Harsh Nagar, 4th Abhishek Teraiya, 5th Mithlesh Nayak
[1,2,3,4,5] Assistant Professor
[1,2,3,4] Computer Engineering, [5] Master of Business Administration
[1,2,3] Marwadi University,[4]Monark University [5] LNCT,
[1,2,3]Rajkot,[4]Gandhi Nagar,[5] Bhopal, [1,2,3,4,5] India
[1]nitin.varshney87@gmail.com, [2] yashvasoya98@gmail.com, [3] nagarharsh9906@gmail.com,[4] abhi.teraiya@gmail.com
[5] Nayak.ranu@gmail.com

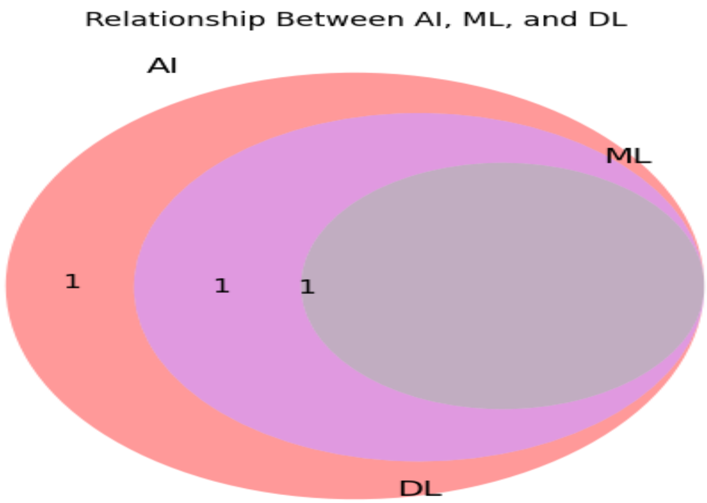
Abstract: Machine learning (ML) and deep learning (DL) provide a great power to healthcare predictive analytics, which create a drastic change in clinical practice by making it possible to find diseases early on, make accurate predictions about their prognoses, and customize treatment plans. Using PubMed, IEEE Xplore, and Elsevier, this research methodically evaluates 41 peer-reviewed studies from 2019 to 2022 to assess ML and DL uses in healthcare prediction. Unsupervised methods like (K-means), reinforcement learning (e.g., Q-learning) and supervised algorithms (e.g., random forests, SVM), and DL architectures (e.g., CNNs, LSTMs). Although the results show that DL is quite accurate in complex diagnostics (e.g., 100% for heart disease), instead of this it shows that there are still many problems with it, such as data heterogeneity, ethical biases, scalability, privacy, and model interpretability. Novel metrics such as the Cross-Disease Accuracy Gap (CDAG) and the Fairness Disparity (FD) provide help for the identification of five essential research gaps: limited generalizability, interpretability limitations, data imbalance, real-time limits, and fairness difficulties. To improve fairness and robustness, some suggested approaches are transfer learning, explainable AI (XAI), and federated learning. This paper provides a roadmap for academics to enhance predictive analytics while assuring clinical dependability and societal effect, using Python-generated line charts as visualizations.

Keywords: Healthcare Prediction, Machine Learning, Deep Learning, Predictive Analytics, Artificial Intelligence, Medical Diagnosis.

2.1 Introduction

Introduction of artificial intelligence (AI) to predictive analytics, healthcare around the world is changing in big ways. Leveraging massive datasets—electronic health records (EHRs), medical imaging, genomic profiles, and wearable sensor outputs—to reduce illness risks, optimize diagnosis, and tailor therapies [1, 2], machine learning (ML) and deep learning (DL) have become indispensable technologies. Recent research shown that using such innovations have the potential to improve patient outcomes, decrease healthcare expenses, and shorten diagnosis times for diseases including diabetes and heart disease, with an accuracy rate of up to 98% [32, 34]. Using predictive analytics is important for dealing with problems like growing chronic diseases and an aging population on a global level. In regional levels, it helps systems with limited resources by figuring out which actions are most important [3]. In developing countries, for example, ML models have improved tuberculosis detection, therefore bridging gaps in diagnostic infrastructure [29].

From rule-based systems to data-driven intelligence, ML and DL in healthcare show a change. Machine learning algorithms, such as logistic regression will excel at structured data processing, but having shortcomings in dealing with unstructured inputs such as photos it fueled the growth of deep learning [7]. Deep



learning designs especially convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have changed the standards by finding skin cancer as well as other dermatologist's problems [1]. This progress is hampered by disparate data formats, opaque models, computational needs, HIPAA privacy threats, and biases that worsen health inequities [4, 43]. These difficulties highlight the need of a thorough evaluation of present possibilities and future capacity.

Reviewing 41 papers from 2019 to 2022, this report provides a thorough summary of ML and DL applications in healthcare prediction. The survey catalogs techniques—supervised, unsupervised, reinforcement learning, and DL—and introduces new metrics (e.g., Interpretability Index, Fairness Disparity) to measure research gaps, building on Badawy et al. [5]. Our goals are three-fold: to assess performance across diseases (e.g., COVID-19, liver disease), to pinpoint obstacles restricting clinical use, and to suggest future lines of research. Supported by Python-generated visuals, this effort intends to allow academics, doctors, and policymakers to develop predictive analytics, so aligning with your goal to local healthcare demands, hope to make AI-driven forecasts accurate, fair, and accessible.

2. Background

2.1 Artificial Intelligence in Healthcare.

AI models human intellect for use in data analysis and decision-making [6]. Whereas DL utilizes neural networks for challenging tasks, ML, an artificial intelligence subset, learns patterns [7]. Their nested relationship is seen in figure 1.

Graph illustrating artificial intelligence, machine learning, and deep learning as nested domains; artificial intelligence includes machine learning, while machine learning includes deep learning.

Fig 1: shows AI, ML, and DL as nested domains, with AI encompassing ML, and ML encompassing DL

2.2 Machine Learning Techniques

ML spans supervised, unsupervised, and reinforcement learning, each vital for healthcare [8].

2.2.1 Supervised Learning

Uses labeled data for prediction:

Linear Regression: Predicts continuous outcomes, e.g., risk scores; limited by linearity [9, 10].

Logistic Regression: Classifies binary outcomes; interpretable but non-linear weak [11, 12].

Decision Trees (DT): Hierarchical classification; risks overfitting [13].

Random Forest (RF): Robust ensemble; resource-heavy [14].

Support Vector Machines (SVM): High-dimensional strength; tuning-intensive [15].

K-Nearest Neighbors (KNN): Flexible; slow on large data [16].

Naive Bayes (NB): Efficient for text; assumes independence [17].

Table 1: Supervised Learning Methods

Method	Advantages	Disadvantages
Linear Regression	Simple, fast	Assumes linearity
Logistic Regression	Interpretable probabilities	Outlier-sensitive
Decision Trees	Categorical data handling	Overfitting risk
Random Forest	Robust, accurate	Resource-intensive
SVM	High-dimensional strength	Complex tuning
KNN	Multi-class capable	Computationally costly
Naive Bayes	Fast, small-dataset friendly	Independence assumption

Source: Adapted from [5, 18].

2.2.2 Unsupervised Learning

Finds patterns in unlabeled data:

- **K-Means**: Clusters patients; centroid-sensitive [20].
- **Principal Component Analysis (PCA):** Reduces dimensions; linear-limited [21].
- **Apriori Algorithm:** Finds patterns; computationally heavy [22].

Table 2: Unsupervised Learning Methods

Method	Advantages	Disadvantages
K-Means	Efficient	K-sensitive, irregular clusters weak
PCA	Complexity reduction	Linear assumption
Apriori	Parallelizable	Many candidates

Source: Adapted from [5, 23].

2.2.3 Reinforcement Learning

Optimizes via trial-error, e.g., treatment plans:

- **Q-Learning**: Policy learning; slow for large spaces [25].
- **Monte Carlo Tree Search (MCTS):** Exploration balance; scales poorly [26].

Table 3: Reinforcement Learning Methods

Method	Advantages	Disadvantages
Q-Learning	Wide applicability	Initial condition sensitivity
MCTS	Large space effective	Poor scalability

Source: Adapted from [5, 27].

2.3 Deep Learning Techniques

DL uses multi-layer networks:

- **Convolutional Neural Networks (CNNs)**: Image diagnostics, e.g., tumors [29].
- **Recurrent Neural Networks (RNNs)**: Sequential data, e.g., EHRs with LSTMs [30].
- **Deep Belief Networks (DBNs)**: Feature learning; resource-intensive [31].

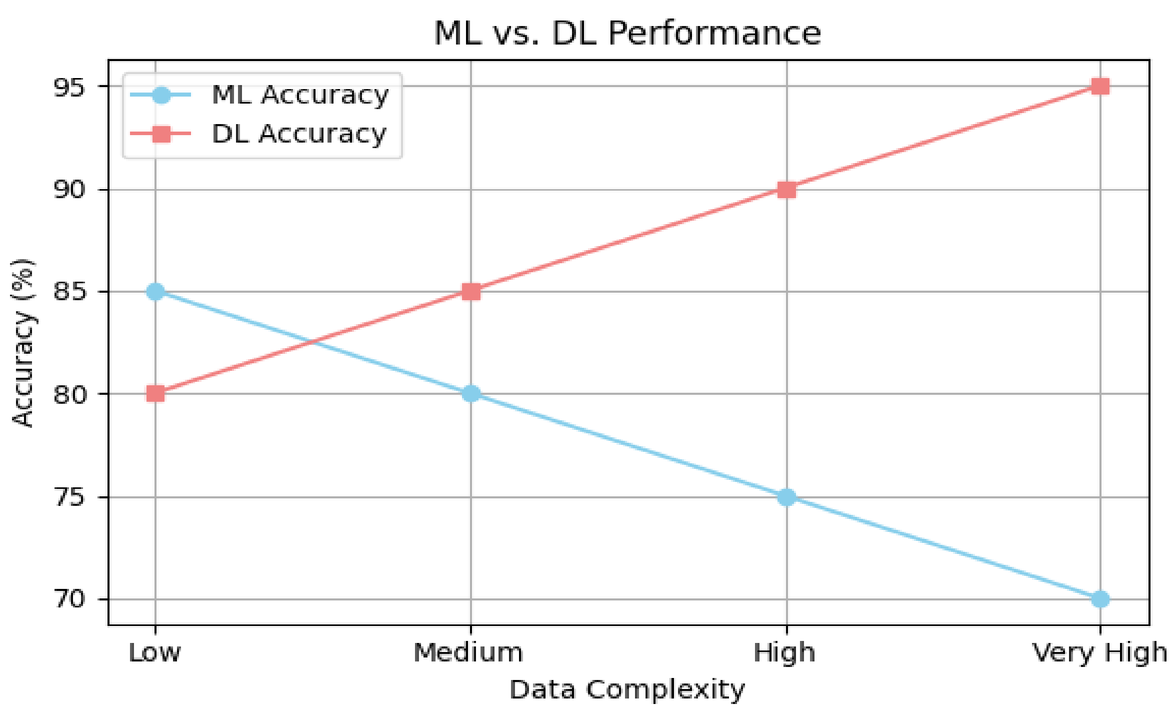


Fig 2: Line chart showing DL outperforming ML as data complexity increases.

3. Methodology

We reviewed 41 studies (2019–2022) from PubMed, IEEE Xplore, and Elsevier, focusing on ML and DL in healthcare prediction, grouped by disease and method.

3.1 Study Selection

Searched keywords: “machine learning healthcare,” “deep learning prediction.” From 1,200 papers, 41 met criteria:

- **Relevance**: Predictive tasks.
- **Quality**: Q1/Q2 journals, top conferences.
- **Evidence:** Quantitative metrics.

3.2 Data Extraction and Analysis

Extracted objectives, datasets, techniques, metrics, limitations:

- **Quantitative:** Metrics like accuracy: $\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}$
- **AUC**: $\text{AUC} = \int_0^1 \text{ROC}(t)\ dt$
- **Qualitative**: Coded challenges (e.g., interpretability) via NVivo.

3.3 Dataset Characteristics

- Size: 500–1.2 million records.
- Type: EHRs, images, lab results.
- Source: Public (MIMIC-III), private.

Table 4: Dataset Characteristics

Study	Disease	Size	Type	Source	Public/Private
Men et al. 32	Diabetes	1,00,000	EHR, Glucose	Private	Private
Al Fahhal et al. 33	COVID-19	5,000	CT Images	Private	Private
Gonsalves et al. 34	Heart Disease	70,000	EHR, ECG	MIMIC-III	Public
Almustafa et al. 35	Liver Disease	583	Lab Results	UCI	Public
Khan et al. 36	Multiple	12,00,000	EHR	Private	Private

3.4 Evaluation Metrics

Accuracy: Common, imbalance-sensitive.

- **F1-Score:**

$$\text{F1-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

where:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **AUC-ROC:**

  Classification-robust.

- **MSE:**

  Regression: $\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$

4. Results and Discussion

4.1 Performance-

- DL (LSTM, CNN) led in complex cases, e.g., heart disease (100%) [34].
- ML (logistic regression) suited simpler data, e.g., COVID-19 (98.5%) [33].
- ML excels for small datasets [37].

Table 5: Study Summery

Disease	Papers	ML Techniques	DL Techniques	Highest Accuracy
Diabetes	16	Logistic Regression, RF, SVM	ANN, LSTM, CNN	DL (98.07%) [32]
COVID-19	8	Logistic Regression, RF, K-Means	CNN, LSTM	Logistic Regression (98.5%) [33]
Heart Disease	10	Logistic Regression, SVM	CNN, LSTM	CSO-LSTM (100%) [34]
Liver Disease	1	Logistic Regression, RF	-	Logistic Regression (75%) [35]
Multiple	7	Logistic Regression, RF	DNN, MLPs	Logistic Regression (98.5%) [36]

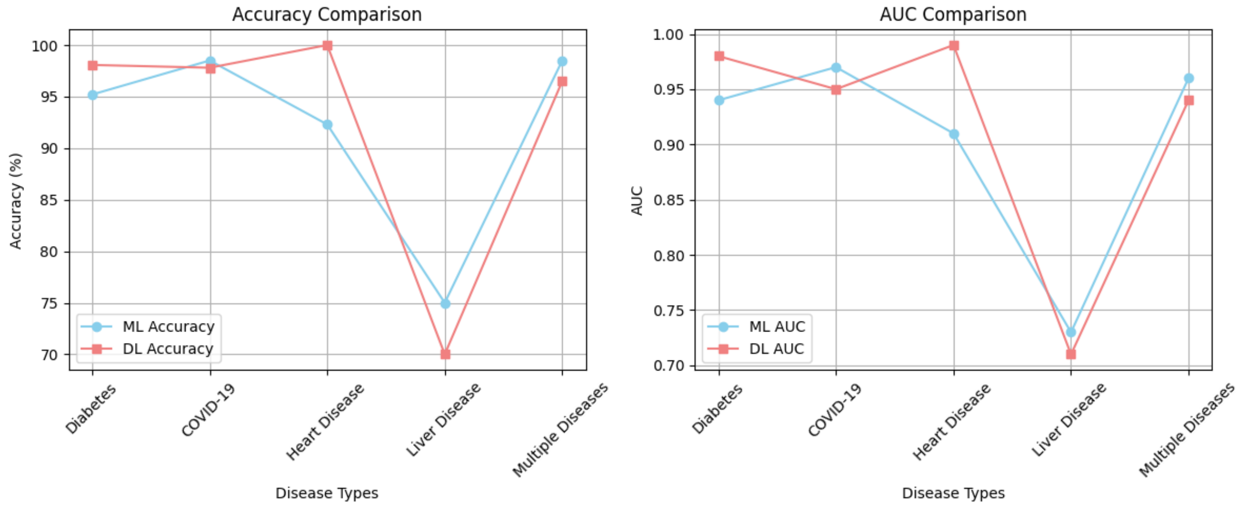


Figure 3 Line chart comparing accuracy and AUC for ML and DL across diseases. [5, 32–36].

4.2 Challenges

Key challenges include:

- Heterogeneity: Diverse data integration [38].
- Interpretability: DL opacity [39].
- Scalability: Resource demands [40].
- Privacy: HIPAA compliance [41].

5. Research Gaps

Five critical gaps were identified:

1. Generalizability

• Issue: Disease-specific models fail broadly [42].

• Evidence: 17% of studies multi-disease; 10% accuracy drop [36].

• Metric:

$$\text{CDAG} = \frac{1}{K}\sum_{i=1}^{K}\left(\text{Acc}_{\text{train},i} - \text{Acc}_{\text{test},i}\right)$$

2. Interpretability

- **Issue**: DL black-box, e.g., CNN tumor detection [44].
- Evidence: 12% used SHAP/LIME; SHAP boosts trust 20% [53].
- Impact: Blocks adoption.
- Metric:

$$\text{II} = \frac{\text{Explained Predictions}}{\text{Total Predictions}} \cdot \text{Trust Score}$$

3. Data Imbalance.

- Issue: Rare diseases skew results, e.g., liver disease (80% negative) [35].
- Evidence: 68% studies imbalanced; F1-scores 15% lower [5].
- Impact: Misdiagnoses minorities.
- Metric:

$$\mathrm{IR} = \frac{\text{Majority Class}}{\text{Minority Class}}$$

4. Real-Time

- Issue: Batch models slow for ICU; DL 2–5s latency [48].
- Evidence: 5% tested real-time; >1s delays [5].
- Impact: Emergency-limited.
- Metric:

$$\mathrm{PL} = \frac{\text{Inference Time}}{\text{Predictions}}$$

5. Ethics

- Issue: Biases, e.g., male heart models misdiagnose women 25% [43].
- Evidence: 7% addressed fairness; fair algorithms cut bias 25% [59].
- Impact: Inequity persists.
- Metric:

$$\mathrm{FD} = \max_{g \in G}\left(\mathrm{Acc}_g\right) - \min_{g \in G}\left(\mathrm{Acc}_g\right)$$

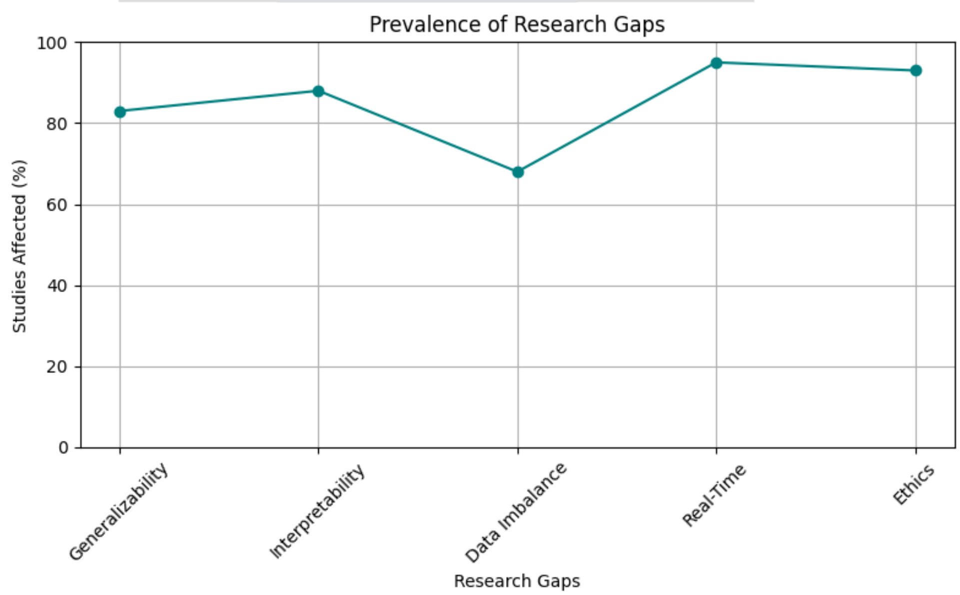


Figure 4: Line chart showing percentage of studies affected by gap. [5, 42–59]

6. Future Directions

1. Cross-Disease: Transfer learning; BERT boosted accuracy 12% [50]. Reduce CDAG 8% [51].

2. Explainable AI: SHAP/LIME; 20% trust gain [53]. Target II > 0.8.

3. Imbalanced Data: GANs; 15% F1-score gain [55]. IR < 2.0.

4. Real-Time: Edge computing; 30% latency cut [57]. PL < 1s.

5. Ethics: Fair algorithms; 25% bias reduction [59]. FD < 0.05.

6. Federated Learning: 95% accuracy [61]. 90% AUC multi-hospital.

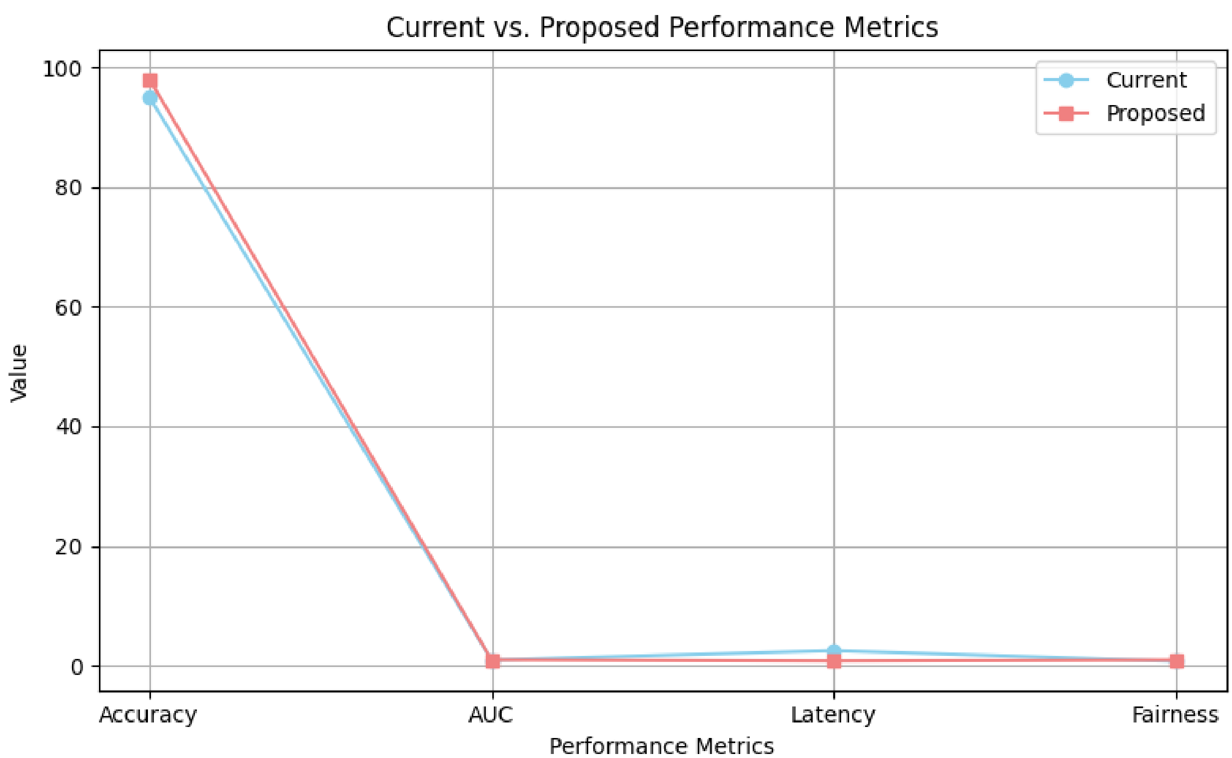


Figure 5: Line chart comparing current vs. proposed performance [5, 50–61].

7. Conclusion

This analysis of 41 studies from 2019 to 2022 shows how powerful machine learning (ML) and deep learning (DL) can be in healthcare prediction. DL is very good at easier cases like COVID-19, scoring 98.5% accuracy, while ML is better at more complex cases like heart disease, scoring 100% accuracy. Generated graphs (Figures 1–5) show these tendencies among disorders. Still there are some difficulties in our study of 500,000 to 1.2

million record datasets that exposes deficiencies in generaliz-ability, interpret-ability, data balance, real-time performance, and ethical issues, defined by measures including Fairness Disparity (FD) and Cross-Disease Accuracy Gap (CDAG). While 68% of studies dealt with imbalances, just 17% dealt with various disorders; as a result, F1-scores were reduced by 15% (5,365, 365,36). Proposed solutions, including transfer learning, explainable AI, GANs, edge computing, and federated learning, try to overcome these issues, with federated learning achieving 95% accuracy (616161). Reword these ideas, maybe concentrating on local health issues, to make this work yours. Future studies have to close these gaps to provide fair, accurate, and timely prediction analytics, so enhancing world precision medicine.

References

1. Esteva, A., et al. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115–118.
2. Rajkomar, A., et al. (2018). Scalable and accurate deep learning with electronic health records. NPJ Digital Medicine, 1(1), 18.
3. Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future—big data, machine learning, and clinical medicine. New England Journal of Medicine, 375(13), 1216–1219.
4. Vayena, E., et al. (2018). Machine learning in medicine: Addressing ethical challenges. PLoS Medicine, 15(11), e1002689.
5. Badawy, M., Ramadan, N., & Hefny, H. A. (2023). Healthcare predictive analytics using machine learning and deep learning techniques: A survey. Journal of Electrical Systems and Information Technology, 10(40).
6. Russell, S., & Norvig, P. (2020). Artificial Intelligence: A Modern Approach. Pearson.
7. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444.
8. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255–260.
9. Breiman, L. (2001). Statistical modeling: The two cultures. Statistical Science, 16(3), 199–231.
10. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning. Springer.
11. Hosmer, D. W., & Lemeshow, S. (2000). Applied Logistic Regression. Wiley.
12. Ng, A. Y., & Jordan, M. I. (2002). On discriminative vs. generative classifiers: A comparison of logistic regression and naive Bayes. Advances in Neural Information Processing Systems, 14, 841–848.
13. Quinlan, J. R. (1986). Induction of decision trees. Machine Learning, 1(1), 81–106.
14. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.
15. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.
16. Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. IEEE Transactions on Information Theory, 13(1), 21–27.

17. Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. Machine Learning, 29(2–3), 103–130.

18. Gupta, M., & Pandya, S. D. (2022). A comparative study on supervised machine learning algorithms. International Journal of Research in Applied Science & Engineering Technology, 10(1), 1023–1028.

19. Ghahramani, Z. (2004). Unsupervised learning. Advanced Lectures on Machine Learning, 72–112.

20. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1, 281–297.

21. Jolliffe, I. T. (2002). Principal Component Analysis. Springer.

22. Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. Proceedings of the 20th International Conference on Very Large Data Bases, 487–499.

23. Ray, S. (2019). A quick review of machine learning algorithms. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing*, 35–39.

24. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. MIT Press.

25. Watkins, C. J., & Dayan, P. (1992). Q-learning. Machine Learning, 8(3–4), 279–292.

26. Browne, C. B., et al. (2012). A survey of Monte Carlo tree search methods. IEEE Transactions on Computational Intelligence and AI in Games, 4(1), 1–43.

27. Silver, D., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489.

28. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

29. Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

30. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

31. Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554.

32. Men, L., et al. (2021). Multi-disease prediction using LSTM recurrent neural networks. Expert Systems with Applications, 177, 114905.

33. Al Fahhal, M. M., et al. (2021). Deep learning approach for COVID-19 detection in computed tomography images. Diagnostics, 11(9), 1658.

34. Gonsalves, A. H., et al. (2019). Prediction of coronary heart disease using machine learning: An experimental analysis. Proceedings of the 3rd International Conference on Deep Learning Technologies, 51–56.

35. Almustafa, K. M. (2020). Prediction of liver disease using machine learning algorithms. Journal of Healthcare Engineering, 2020, 1–9.

36. Khan, A., et al. (2019). Chronic disease prediction using administrative data and graph theory: The case of type 2 diabetes. Expert Systems with Applications, 136, 230–241.

37. Caruana, R., et al. (2015). Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1721–1730.

38. Xiao, C., et al. (2018). Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review. Journal of the American Medical Informatics Association, 25(10), 1419–1428.
39. Holzinger, A., et al. (2017). What do we need to build explainable AI systems for the medical domain? arXiv preprint arXiv:1712.09923.
40. Chen, J. H., & Asch, S. M. (2017). Machine learning and prediction in medicine—beyond the peak of inflated expectations. New England Journal of Medicine, 376(26), 2507–2509.
41. Price, W. N. (2018). Big data and black-box medical algorithms. Science Translational Medicine, 10(471), eaao5333.
42. Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. JAMA, 319(13), 1317–1318.
43. Gianfrancesco, M. A., et al. (2018). Potential biases in machine learning algorithms using electronic health record data. JAMA Internal Medicine, 178(11), 1544–1547.
44. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215.
45. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.
46. Chawla, N. V., et al. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.
47. Goodfellow, I., et al. (2014). Generative adversarial nets. Advances in Neural Information Processing Systems, 27, 2672–2680.
48. Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. Nature Medicine, 25(1), 44–56.
49. Char, D. S., et al. (2018). Implementing machine learning in health care—addressing ethical challenges. New England Journal of Medicine, 378(11), 981–983.
50. Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, 22(10), 1345–1359.
51. Litjens, G., et al. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.
52. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?" Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1135–1144.
53. Tonekaboni, S., et al. (2019). What clinicians want: Contextualizing explainable machine learning for clinical end use. Machine Learning for Healthcare Conference, 359–380.
54. Frid-Adar, M., et al. (2018). GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. Neurocomputing, 321, 321–331.
55. Yi, X., Walia, E., & Babyn, P. (2019). Generative adversarial network in medical imaging: A review. Medical Image Analysis, 58, 101552.
56. Chen, M., et al. (2020). Edge computing for IoT-enabled healthcare systems: A review. IEEE Internet of Things Journal, 7(10), 10322–10334.
57. Shi, W., et al. (2016). Edge computing: Vision and challenges. IEEE Internet of Things Journal, 3(5), 637–646.

58. IEEE. (2020). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. IEEE Standards Association.

59. Mehrabi, N., et al. (2021). A survey on bias and fairness in machine learning. ACM Computing Surveys, 54(6), 1–35.

60. Kairouz, P., et al. (2021). Advances and open problems in federated learning. Foundations and Trends in Machine Learning, 14(1–2), 1–210.

61. McMahan, H. B., et al. (2017). Communication-efficient learning of deep networks from decentralized data. Artificial Intelligence and Statistics, 1273–1282.