Design And Development of Interpretable Machine Learning Models for Healthcare Prediction

Rajeev Ranjan
Research Scholar, Department of Computer Science
V. K. S. University, Ara, Bihar

Abstract*—The ever-increasing volume of healthcare data and the development of smart technology to aid in early assessment and good professional decision-making. Machine learning's (ML) capacity to sift through complicated medical information has made it an invaluable tool for healthcare forecasting. Results from many recent ML models are typically unclear, despite their excellence in prediction. This complicates their use, as trust and understanding are fundamental in practical therapeutic settings. The main objective of this project is to create an XML model for predictive healthcare with the goal of making medical decision support systems more accurate and easier to comprehend. The main goal of this effort is to construct a strong machine learning system that can accurately forecast serious health issues and provide thorough explanations for such predictions. Advanced ensemble learning methods are used in the study to enhance the prediction accuracy. The study also looks at common problems with healthcare datasets such uneven distribution of classes, data loss, and data unpredictability. The proposed framework makes use of Explainable Artificial Intelligence (XAI) techniques to ensure that it is easy to understand. Methods like feature value analysis and model explanation tools fall within this category. We evaluate the suggested model using conventional performance measures including accuracy, precision, recall, F1-score, and AUC-ROC to make sure it works and is reliable. The results show that the model's capacity for future prediction and patient comprehension is enhanced when explainability is included in ensemble learning. This makes it more applicable to real-life healthcare environments. By connecting theoretically solid prediction models with their practical use in healthcare, this study adds to the growing field of AI-driven healthcare. Improved decision-making speed, accuracy, and clarity might be achieved by healthcare practitioners using the explainable ML model that was built. More trust in AI-driven healthcare systems and better health outcomes for patients should be the final product.

I. INTRODUCTION

The artificial intelligence and machine learning have been incorporated into healthcare systems, which has resulted in a shift in the analysis, understanding, and use of medical data for the purpose of therapeutic decision-making. Predictive healthcare is an attractive new field of research that makes use of machine learning models to examine historical and current medical data in order to make predictions about the progression of illnesses, the frequency with which patients will need readmission, the accuracy of diagnosis, and the efficacy of treatments. In spite of the fact that these models have achieved significant advancements in terms of accuracy and speed, there is still a significant issue: the manner in which they arrive at choices is not transparent or simple to comprehend. The development of Explainable Machine Learning (XML) models is becoming more popular as a solution to this challenge. This is particularly true in the highly sensitive and high-stakes industry of healthcare, where it is just as vital to understand the "why" behind a prognosis as it is to understand the prediction itself. Explainable machine learning is a term that describes the approaches and tactics that assist anyone, particularly medical professionals, nurses, and patients, in comprehending the operation of complicated machine learning models and the results that they generate. When it comes to machine learning, classical models, particularly deep learning and ensemble approaches, are sometimes referred to as "black boxes" since they provide very accurate answers but do not provide an explanation as to how they arrived at those conclusions. Because of this lack of clarity, trust can be damaged, people may be dissuaded from adopting new technologies, and ethical and legal issues may arise in the healthcare industry.

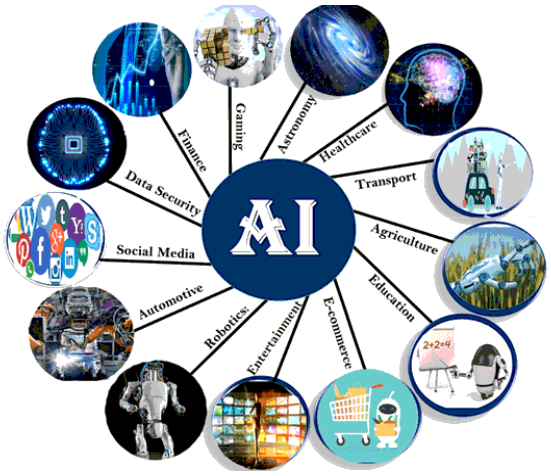

Figure.1. Use of AI in different sectors

Machine Learning Techniques

The term "machine learning algorithms" refers to a collection of mathematical and statistical techniques that computers may employ to "learn" from data, discover patterns within it, and make predictions or judgments without the need for human interaction. For the purpose of computing, the conventional technique entails the formulation of rules for each and every conceivable result. Artificial intelligence algorithms, on the other hand, are able to learn from their errors and make use of vast quantities of data in order to develop their own rules. The purpose of these algorithms is to discover links between the parameters that were utilized and the outcomes by learning from data that has been tagged or that has been obtained in the past. Immediately after the completion of the training process, the model may start the generation of predictions based on newly gathered data. One such example is a machine learning system that learns to recognize spam emails by analyzing patterns in spam emails that have been classified in the past.

There are three primary categories of algorithms that are used in machine learning:

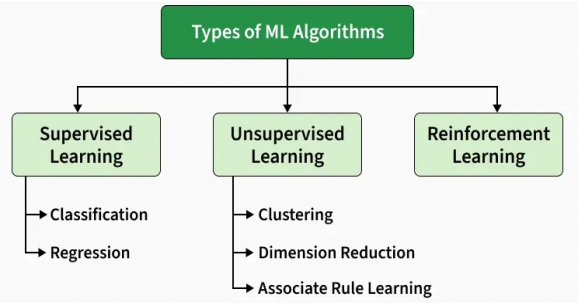

Figure.2. Machine Learning Algorithms

Supervised Learning Algorithms

The use of datasets is necessary in order to achieve the objective of training supervised learning algorithms. The label is a term that is used to refer to the target or response variable that is connected with each sample that is included within these datasets. In order for the model to be able to generate correct predictions on data that it has not yet seen, the goal is to achieve the learning of a mapping function from the input data to the matching output labels. This will allow the model to learn how to map the input data to the output labels. Because of this, the model will be able to provide accurate predictions about the future.

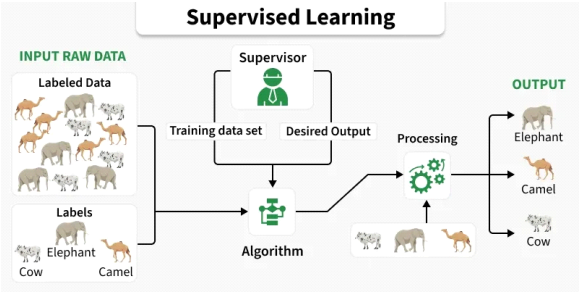

Figure.3. Supervised Learning Algorithms

Objectives of the study

Due to the foregoing obstacles, this study has the following objectives:

- To create an explainable predictive healthcare machine learning model that balances accuracy and transparency while addressing interpretability.
- To provide explanation methods that identify biases and connect model outputs with clinical reasoning to increase healthcare professionals' trust and usability.
- To assess the model's prediction performance, fairness, and integration capabilities to ensure its practicality in real-world healthcare systems and ethical compliance.

II. REVIEW OF LITERATURE

It is impossible to do research without first conducting a thorough literature review, which summarizes all prior work on the topic, the underlying ideas, and the research methods. Reviewing prior research is crucial since it helps to find research gaps, stops studies from being duplicated, and gives a good conceptual groundwork for the present investigation. Beyond this, it gives the researcher a chance to think about what

other people have found, assess other perspectives, and explain the study's significance within a broader academic framework. There have been major shifts in the analysis and understanding of medical data due to the advent of machine learning (ML). Machine learning methods have allowed computers to learn from complicated information, which has led to an increase in the precision of diagnoses, predictions, and personalized treatment regimens. Machine learning has come a long way from its humble beginnings as a collection of basic statistical models; now, it encompasses complex algorithms like deep learning and mixed learning, which can handle clinical data in real-time. Because of this change, medical personnel are now better able to make informed decisions, the frequency of diagnostic mistakes has decreased, and early disease detection has become much easier. Hence, it is crucial to research this topic historically to comprehend the development of machine learning applications and their potential for healthcare improvement, especially in healthcare systems that are anticipated.

Medical Machine Learning

In modern healthcare systems, machine learning (ML) has become an indispensable tool. As a consequence, medical experts are able to make better diagnoses and prognoses, and massive amounts of clinical data are easier to handle. Machine learning (ML) technologies have been more widely used in healthcare systems over the last decade to improve clinical decision support system performance, patient monitoring, and disease prediction accuracy. By using machine learning in healthcare, conventional doctors may move away from relying on anecdotes and instead make recommendations based on hard facts. Scientists from across the world have studied machine learning's various medical uses, with a focus on intensive care units (ICUs) and other critical care settings where quick and accurate decisions are life-or-death for patients.

Predictive Healthcare Models

In contemporary medical treatment, the emphasis is gradually shifting away from using a reactive, symptom-based strategy and toward using a proactive, data-driven approach. Prediction healthcare models that are based on machine learning (ML) are at the forefront of this trend. These models have the ability to forecast detrimental clinical outcomes and events before they take place. Through the compilation of complex, high-dimensional, and often continuous patient data, these models make an effort to provide physicians with information that is to their benefit. Early Warning Systems (EWS), Sepsis prediction models, and predicting mortality in the Intensive Care Unit (ICU) are three incredibly major uses of machine learning in the healthcare business that have shown to be very successful in the real world. These applications have been proved to contribute significantly to the field of medicine. In this part of the article, we will take a comprehensive look at the many research that have been carried out on these three distinct elements of prediction.

Class Imbalance Problem in Healthcare

Literature also makes a distinction between intrinsic imbalance and extrinsic imbalance. Intrinsic imbalance is caused by things like the low base rate of a rare cancer, while extrinsic imbalance is caused by problems with data collection, like lost lab samples or incomplete Electronic Health Record (EHR) documentation (He & Garcia, 2009). In predicted healthcare, experts are mostly looking at imbalances that happen naturally. This difference is important because innate imbalance can't be "fixed" by just gathering more data; the machine learning process needs to be completely changed to make sure that the minority class is properly reflected and learned from during model training.

Techniques for Handling Class Imbalance

He et al. (2008) offered an alternative to the traditional method of uniform synthetic generation called the Adaptive Synthetic Sampling Approach (ADASYN). Instead of concentrating on creating a uniform distribution of synthetic samples, ADASYN is devoted to creating an adjustable strategy based on density. For every minority occurrence, it determines a weight by comparing it to the ratio of nearby majority instances, and then it assigns that weight. More synthetic samples are synthesized in the region around minority samples when there is a high density of majority samples (sometimes called "hard-to-learn" circumstances). This is because minority samples in such areas are given greater weights. This is due to the fact that learning from majority samples is more challenging. When readily differentiated from

majority collections, minority samples have fewer synthetic neighbors than the majority samples.

Machine Learning Algorithms Used in Healthcare

In the field of medical predictive modeling, the Logistic Regression (LR) method continues to be the undisputed baseline approach. Through the process of fitting a logistic function to a linear combination of predictor variables, the logistic regression (LR) model, which functions as an extended linear model, calculates the likelihood of a binary clinical outcome (for example, illness vs no disease). Interpretability is the major benefit of LR in the healthcare industry since it is intrinsic and cannot be contested. Clinicians are able to readily compute Odds Ratios (OR) for certain biomarkers or demographic variables because to the fact that the model's coefficients directly correlate to the log-odds of the outcome (Kwon et al., 2021).

Feature Engineering in Healthcare Data

It is possible that the phase of feature selection is the most important one to consider when examining the engineering process through the lens of Explainable Machine Learning. It has been determined by Lipton (2018) that the most efficient method for making a model interpretable is to reduce the amount of characteristics that are input into the model. In comparison to a model that makes use of 500 points from electronic health records, a model that forecasts sepsis based on 15 essential clinical signs is intrinsically far simpler to explain to a physician.

Artificial Intelligence that Can Be Explained (XAI)

"Automation bias" is a phenomena in which physicians blindly accept automated advice, which leads to the atrophy of clinical reasoning abilities (Goddard et al., 2012). Without explainability, healthcare systems run the danger of falling prey to this phenomenon. In the other direction, a lack of transparency might result in algorithmic rejection; if physicians are unable to justify the output of an artificial intelligence, they will simply disregard it, which will make the technology worthless regardless of how accurate it is statistically.

III. RESEARCH METHODOLOGY

The development of computers themselves, the concept of creating tools that are intelligent and aware of themselves has gradually matured over the course of many years. Charles Babbage developed the Analytical Engine in the 1800s, and Ada Lovelace is credited with writing what is considered to be the first computer program. These two events are considered to be the origins of modern computers. These early breakthroughs served as a source of motivation for scientists and intellectuals, which significantly increased the likelihood that computers may one day be intelligent and capable of thinking for themselves. Approximately around the middle of the twentieth century, Alan Turing expanded upon this concept by establishing the foundation for theoretical computer science, algorithms, formal languages, machine learning, and artificial intelligence. Our current era is marked by an abundance of data, frequently called the "age of data" or the "information age," brought about by the proliferation of fast computers, enhanced processing capacity, and vast amounts of storage space. Massive volumes of data are generated and managed on a daily basis by both people and professional organizations. In order to make sense of this mountain of data, organizations are increasingly turning to approaches from the fields of data science, artificial intelligence, and machine learning. The use of these technologies allows for the identification of important patterns, insights, and forecasts.

Algorithms

An algorithm is a well-defined, step-by-step technique or collection of instructions that is intended to carry out a certain task or solve a specific problem. Its purpose is to accomplish either of these specific goals. By providing a basic method to the processing of data, the execution of calculations, and the formulation of choices, algorithms are fundamental components in the areas of computer science and programming.

Input

An input may be zero, one, or more, and a computer program has to be able to accept all of them. The term "inputs" refers to the information or values that are provided to the algorithm before the algorithm begins to carry out its operations. One or more of the following might be the source of the input: people, files, sensors, or even the programs themselves.

Output

It is essential for an algorithm to provide at least one output in order to function properly. The output is the

result that is obtained after all of the steps of the algorithm have been carried out. There are many phases in the method.

Finiteness

Before the execution of an algorithm may be regarded to be finished, it is always necessary to finish a certain number of steps. As a consequence of this, it is difficult to imagine that the algorithm might continue to function eternally because of it. It is necessary for the algorithm to arrive at a conclusion and terminate its execution for any input that is regarded to be legitimate after a certain length of time has elapsed after the completion of the algorithm's execution.

Definiteness

It is of the utmost importance that each and every level of the algorithm be entirely obvious, exact, and free of any ambiguity. This may be interpreted to suggest that every single order must be easily understood and should not allow any space for interpretation. This is the meaning that can be inferred from this. In the event that a significant number of individuals adhere to the identical process, it is fair to anticipate that they will all arrive at the same outcomes.

Effectiveness

It is necessary for an algorithm to be efficient, which means that each step must be straightforward, fundamental, and able to be carried out within a certain period of time. In order to be practicable and viable to carry out using the resources that are available, the instructions should be. It is also expected that the algorithm would give accurate findings in a timely manner. A good example of an algorithm that is not regarded effective is one that demands an endless level of accuracy or an unreasonable amount of computer power.

Additional Characteristics of a Good Algorithm

- Correctness – The algorithm should produce correct results for all valid inputs.
- Efficiency – It should use minimum time and memory resources.
- Generality – The algorithm should solve a broad range of problems rather than a single instance.
- Simplicity – The algorithm should be easy to understand and implement.

In its approach to the solving of problems, an algorithm is a procedure that is both ordered and systematic in its implementation. Algorithms are transformed into reliable tools for computing and decision-making in the area of computer science as well as in applications that are based on the real world. Through the ensuring of qualities such as input, output, finiteness, definiteness, and effectiveness, algorithms become trustworthy instruments.

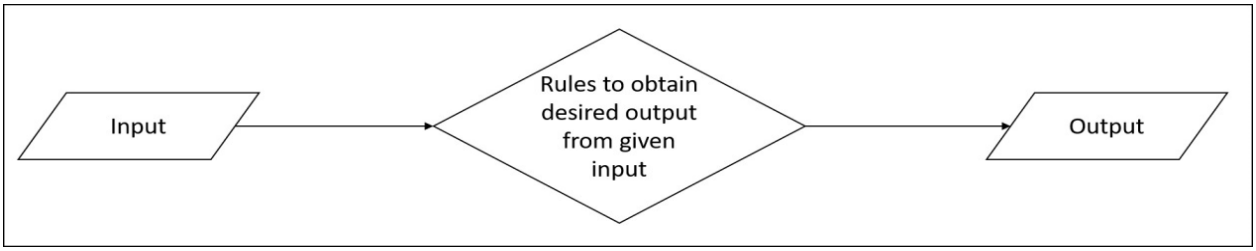

Figure.4. Data flow of an algorithm

Types of Machine learning algorithms

There are many different categories that machine learning algorithms fall into, and these categories are determined by the features of the learning process as well as the output that is meant to be produced. When all of these factors are taken into account, it becomes much simpler to understand how computers learn from data and how they arrive at predictions or decisions. For the purpose of machine learning, it is feasible to classify algorithms into a large range of different categories.

Supervised Learning:

The field of machine learning encompasses a wide variety of methodologies, but supervised learning is among the most prevalent of these techniques. A method of learning known as supervised learning includes feeding the algorithm with data that has been tagged and giving it both input and output values so that it may learn. The development of a function that properly correlates the variables that are input into the system with the outputs that are intended is one of the primary goals of supervised learning. In the course of

its training, the algorithm investigates input-output pairings and identifies patterns that have the potential to be used to the venture of forecasting fresh data.

Unsupervised Learning:
Unsupervised learning is used in situations when labeled data is not accessible. Using this approach, the algorithm searches for hidden correlations, patterns, or structures in the incoming data. When compared to supervised learning, unsupervised learning does not rely on labels that have already been defined for the output. Instead, the algorithm classifies data into groups based on commonalities. Few supervised learning methods are as prevalent as clustering, dimensionality reduction, and association rule learning.

Semi-Supervised Learning:
When it comes to semi-supervised learning, it is essential to make use of both supervised and unsupervised data types. In many situations that occur in the real world, there is a lack of tagged data despite the fact that there is a large amount of data that is not labeled. An improvement in learning accuracy may be achieved by the use of semi-supervised learning, which involves the use of both small amounts of labeled data and huge volumes of unlabeled data.

Reinforcement Learning:
Algorithms that benefit from reinforcement learning gain knowledge via seeing and responding to real-world events. In this approach, the algorithm learns a policy that controls how it acts in certain situations. The environment guides the learning process by rewarding or penalizing the algorithm as it completes its tasks. Through the gradual elimination of penalties and the gradual increase of incentives, algorithms that effectively use reinforcement learning gradually improve their performance.

Transduction Learning
Despite their similarities, supervised and transduction learning vary in that the former does not involve the explicit creation of a generic function. To make predictions, however, it relies on training data and outcomes to anticipate how certain inputs would behave. The goal of the method in transduction is to generate predictions from the provided information instead than constructing a generic model.

Learning to Learn (Meta-Learning)
Among the many forms of advanced machine learning, meta-learning stands out. The algorithm learns from its own performance history and adds to its existing body of information in this kind of learning. This approach improves the algorithm's ability to learn new tasks by combining its previous knowledge with its own inductive bias. Because of meta-learning, systems may learn to adapt to new environments and solve new problems with less training data. Applications like artificial intelligence, automated machine learning, and adaptive systems often use this approach.

Computational Learning Theory
Computational learning theory is a subfield of statistics concerned with the study of accurate and efficient computational analysis of machine learning methods. This field mainly focuses on studying learning algorithms and their complexity, accuracy, and efficiency. Research on learning data needs, algorithm convergence speed, and data generalizability is conducted. Computational learning theory may help researchers understand the limitations and theoretical foundations of machine learning algorithms, which can lead to the development of more effective algorithms.

Concept of Learning in Machine Learning
Building algorithms that can learn new things from data is at the core of machine learning. When it comes to machine learning, a human-level of consciousness or intelligence is not necessary for learning. The major areas of interest instead are on statistical regularities, trends, and correlations. Machine learning algorithms and human learners often use quite distinct approaches to knowledge acquisition. In contrast, these algorithms are masters at sifting through enormous information in search of patterns that humans might overlook. Learning algorithms also show how challenging learning tasks are in different contexts, which helps in building more robust systems.

IV. DATA ANALYSIS

The methods of exploratory data analysis are often used in the time leading up to the introduction of formal modeling practices. It is possible that the information required for the construction of very

complex statistical models might be improved by the use of these approaches. Additionally, exploratory data analysis techniques are necessary for removing or refining plausible assumptions about the world that may be addressed by the data. This may be accomplished by removing or refining the assumptions. The reason for this is because the data may be used to substantiate the hypothesis. In the following parts, we will discuss the particulars of the exploratory data analysis that was carried out for the purpose of this research.

Import Libraries

The data preparation, feature engineering, model creation, and performance assessment components of this investigation were carried out with the assistance of a number of different Python modules. It is much simpler to do activities involving data analysis and machine learning thanks to the wide ecosystem of modules that are available for the programming language Python. Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, and Imbalanced-learn (Imblearn) are some of the important libraries that were applied in this investigation.

Data Structure Operations: Pandas

The BSD license governs the use of the Pandas library, which is a free and open-source software package. Python programmers are provided with high-performance data structures and tools for data analysis by this framework. Additionally, making it simple to use was a primary goal throughout its development. Utilizing this technology for exploratory data analysis, data purification and transformation, and structured data management are just some of the many typical applications of this technology.

A fundamental level of support is offered by Pandas for two different data structures:

- Series – One-dimensional labeled array
- DataFrame – Two-dimensional labeled data structure

Through the use of these data structures, the process of preserving, filtering, and analyzing datasets is simplified significantly. When using Pandas, even the most fundamental statistical analysis, data merging, data aggregation, and management of missing values are as easy as a child's play. It is possible to locate these functions inside the library. Getting a dataset suitable for use in machine learning applications requires the use of Pandas (https://pandas.pydata.org), which is an important component. mainly due to the fact that Pandas is capable of operating independently of any other software.

Import the datasets

Several distinct techniques for importing datasets in a wide variety of formats have been made available by the Pandas library so far. For the purpose of this investigation, the read_csv method was used. read_csv requires the file name as an argument. As a separator, it makes use of commas. Additionally, the option sep has to be set to the proper character in any other separator. There is an expectation that the heading will be the first line in the dataset. If this is not the case, the header parameter must be set to None.

	HR	O2Sat	Temp	SBP	MAP	DBP	Resp	EtCO2	BaseExcess	HCO3	FiO2	pH	PaCO2	SaO2	AST	BUN	Alkalinephos	Calcium	Chloride	Creatinine	E
0	99.0	100.0	NaN	NaN	71.0	NaN	13.5	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
1	97.5	100.0	NaN	NaN	NaN	NaN	15.0	NaN	-16.0	16.0	NaN	7.19	25.0	NaN	72.5	13.5	58.0	8.1	112.5	1.7	
2	96.0	100.0	NaN	NaN	70.0	NaN	13.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	100.0	99.0	NaN	NaN	75.0	NaN	14.0	NaN	-15.0	11.0	NaN	7.24	24.0	96.0	NaN	12.0	NaN	7.7	113.0	1.6	
4	102.0	100.0	36.00	NaN	74.0	NaN	17.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
5	92.0	100.0	NaN	NaN	79.0	NaN	14.0	NaN	-13.0	10.0	NaN	7.29	22.0	NaN	68.0	11.0	49.0	7.3	115.0	1.5	
6	90.0	100.0	NaN	NaN	62.0	NaN	15.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
7	95.0	99.0	NaN	151.0	93.0	67.0	21.0	NaN	-11.0	NaN	NaN	7.36	20.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
8	92.0	100.0	36.89	94.0	64.0	49.0	16.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
9	91.0	100.0	36.72	106.0	68.0	51.0	15.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

Figure.5. Displaying first few observations of the Data Frame

V. CONCLUSION

In the context of the area of predictive healthcare, the primary objective of this research was to develop an Explainable Machine Learning Model with the intention of the early identification and prediction of

sepsis via the use of clinical data. The primary purpose of the research was to determine the effectiveness of a number of different machine learning algorithms by using strategies for class imbalance control and a structured four-phase predictive modeling framework. The findings of the research indicate that the improvement of predictive performance in healthcare applications is based on two factors: the algorithms that are used for machine learning and the approaches that are utilized to address class imbalances. When it comes to generating predictions regarding healthcare analytics, the inclusion of datasets that are not perfectly balanced is a big cause for worry. When the number of patients who do not have sepsis greatly exceeds the number of cases who have sepsis, the situation becomes very serious. A typical consequence of such an imbalance is the identification of events that involve minority classes that are not accurate. This imbalance often results in biased models that are advantageous to the dominant class. Getting rid of the imbalance between classes is very necessary in order to improve the accuracy and reliability of forecasts. This is because sepsis must be recognized as soon as possible in order to prevent further complications in healthcare institutions. SMOTE, which stands for the Synthetic Minority Oversampling Technique, was one of the three principal class imbalance treatment strategies that were used in this inquiry. The other two were random under-sampling and random over-sampling. A number of different machine learning techniques were used in order to analyze various tactics. various tools included Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), XG Boost, AdaBoost (ADA), and hybrid models such as SPMPH and 2-level SPMPH.

REFERENCES

[1] Alistarh, D., et al., "Temporal convolutional networks and deep learning for clinical time-series in critical care," *Journal of Biomedical Informatics*, vol. 118, p. 103803, 2021, doi: 10.1016/j.jbi.2021.103803.

[2] Amann, J., A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," BMC Medical Informatics and Decision Making, vol. 20, no. 1, p. 310, 2020, doi: 10.1186/s12911-020-01332-6.

[3] Batista, G. E., R. C. Prati, and M. C. Monard, "A study of the behavior of several methods for balancing machine learning training data," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 20–29, 2004, doi: 10.1145/1007730.1007735.

[4] Churpek, M. M., T. C. Yuen, C. Winslow, D. O. Meltzer, M. W. Kattan, and D. P. Edelson, "Multicenter comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards," Critical Care Medicine, vol. 44, no. 2, pp. 368–374, 2016, doi: 10.1097/CCM.0000000000001571.

[5] Fernández, A., S. García, F. Herrera, and N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary," Journal of Artificial Intelligence Research, vol. 61, pp. 863–905, 2018, doi: 10.1613/jair.1.11192.

[6] Goh, Y. S., Y. T. Ng, A. Mitani, C. M. Loo, and C. T. Lim, "Using natural language processing to identify early signs of sepsis from clinical notes in the electronic health record," Journal of the American Medical Informatics Association, vol. 28, no. 4, pp. 739–747, 2021, doi: 10.1093/jamia/ocaa270.

[7] Harutyunyan, H., J. Zech, M. Cornilescu, and A. Goldenberg, "Multitask learning and benchmarking with clinical time series data," Scientific Data, vol. 6, no. 1, p. 96, 2019, doi: 10.1038/s41597-019-0103-8.

[8] He, H., and E. A. Garcia, "Learning from imbalanced data," IEEE Transactions on Knowledge and Data Engineering, vol. 21, no. 9, pp. 1263–1284, 2009, doi: 10.1109/TKDE.2008.239.

[9] Hug, C. W., A. Sharma, and L. A. Celi, "Deep learning for ICU mortality prediction: A systematic review of model architectures and clinical utility," Critical Care, vol. 25, no. 1, p. 145, 2021, doi: 10.1186/s13054-021-03586-2.

[10] Japkowicz, N., and S. Stephen, "The class imbalance problem: A systematic study," Intelligent Data Analysis, vol. 6, no. 5, pp. 429–449, 2002, doi: 10.3233/IDA-2002-6504.

[11] Johnson, J. M., and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," Journal of Big Data, vol. 6, no. 1, p. 27, 2019, doi: 10.1186/s40537-019-0192-5.

[12] Kipnis, P., B. J. Turk, D. A. Wulf, K. D. Liu, V. Liu, and G. J. Escobar, "A pragmatic electronic health record-based early warning score for the general ward," Journal of Hospital Medicine, vol. 11, no. 12, pp. 847–853, 2016, doi: 10.1002/jhm.2634.

[13] Lundberg, S. M., G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S. I. Lee, "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, no. 1, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.

[14] Mandel, J. C., S. Khor, H. Dalianis, and S. Velupillai, "Generalizability and bias in machine learning algorithms for sepsis prediction: A systematic review," Journal of the American Medical Informatics Association, vol. 28, no. 8, pp. 1690–1700, 2021, doi: 10.1093/jamia/ocaa312.

[15] Nemati, S., A. Holder, F. Razmi, M. D. Stanley, G. D. Clifford, and T. G. Buchman, "An artificial intelligence system for predicting the onset of sepsis in the ICU," Nature Medicine, vol. 24, no. 11, pp. 1706–1711, 2018, doi: 10.1038/s41591-018-0213-y.

[16] Pirracchio, R., M. L. Petersen, G. Caruana, M. R. Rigon, S. Chevret, and M. J. van der Laan, "Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): A population-based study," The Lancet Respiratory Medicine, vol. 3, no. 1, pp. 42–52, 2015, doi: 10.1016/S2213-2600(14)70239-5.

[17] Rajkomar, A., E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Z. Zhang, M. Blau, D. Duggan, J. Guevara, M. Hardy, et al., "Scalable and accurate deep learning with electronic health records," NPJ Digital Medicine, vol. 1, no. 1, p. 18, 2018, doi: 10.1038/s41746-018-0029-1.

[18] Rajpurkar, P., J. Han, R. Ghosh, M. Mannion, and E. J. Topol, "Deep learning in medical diagnostics: Addressing the imbalance problem in rare diseases," NPJ Digital Medicine, vol. 1, no. 1, p. 45, 2018, doi: 10.1038/s41746-018-0052-2.

[19] Sendelbach, S., "Alarm fatigue: A threat to patient safety in the intensive care unit," AACN Advanced Critical Care, vol. 31, no. 1, pp. 81–86, 2020, doi: 10.4037/aacnacc2020824.

[20] Seymour, C. W., F. Gesten, H. C. Prescott, M. E. Friedrich, T. J. Iwashyna, G. S. Phillips, S. Lemeshow, T. Osborn, K. M. Terry, M. M. Levy, G. L. Allen, C. L. Hough, D. Clapp, K. Hargett, C. M. Lilly, A. Barche, M. D. Howell, D. P. Edelson, et al., "Time to treatment and mortality during mandated emergency care for sepsis," The New England Journal of Medicine, vol. 376, no. 23, pp. 2235–2244, 2017, doi: 10.1056/NEJMoa1703058.

[21] Smith, G. B., D. R. Prytherch, P. Meredith, P. E. Schmidt, and P. I. Featherstone, "The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death," Resuscitation, vol. 137, pp. 166–167, 2019, doi: 10.1016/j.resuscitation.2019.02.020.

[22] Weiss, G. M., "Mining with rarity: A unifying framework," ACM SIGKDD Explorations Newsletter, vol. 6, no. 1, pp. 7–19, 2004, doi: 10.1145/1007730.1007734.