IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 14 December 2023, accepted 3 January 2024, date of publication 9 January 2024, date of current version 23 January 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3351688

RESEARCH ARTICLE

Optimizing Patient Recruitment for Clinical Trials: A Hybrid Classification Model and Game-Theoretic Approach for Strategic Interaction

C. U. OM KUMAR, ISSHAAN SINGH, AND M. SUGUNA
School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, Tamil Nadu 600127, India
Corresponding author: C. U. Om Kumar (omkumar.cu@vit.ac.in)

ABSTRACT This research is imperative due to the pressing need for improved patient recruitment in clinical trials, addressing challenges such as delays and high costs. By introducing a classification model and a game theoretic approach for clinical trial setting, we aim to boost trial efficiency, advance healthcare research, and enhance patient outcomes. This research is critical for revolutionizing recruitment strategies and accelerating medical progress. In this paper, we present a classification model that has been specifically designed to address this issue effectively. The proposed model employs an Autoencoder, augmented by a super classification model that merges Logistic Regression, Support Vector Machines, Random Forest Trees, and Decision Trees using a stacking classifier. The output of the super classifier is further processed by a meta classifier to obtain the final result. Notably, the model achieves a training accuracy of 99.576% and a validation accuracy of 83.45%, illustrating its robust classification performance and its potential to streamline patient recruitment, reducing delays and resource consumption. In addition to the classification model, this study formulates a three-layer game theoretic model involving Patients, Doctors or Clinical Investigators, and Research Firms. Within this static repeated game setting, players sequentially strategize to optimize their recruitment strategies, while research firms aim to optimize their overall interaction. The paper proposes a novel optimal solution that strikingly balances the payoffs of all three players. Moreover, the work presents a necessary condition and closed form for the existence of an equilibrium in the game, offering a strategic approach to recruitment optimization, and striking a balance between stakeholders. This equilibrium-seeking solution has the potential to revolutionize recruitment dynamics and foster collaboration. Additionally, the study's theoretical contributions lay the groundwork for future research in this critical healthcare domain.

INDEX TERMS Clinical trials, decision-making, ensemble learning, feature extraction, game theory, healthcare analytics, stacked ensemble, soft computing.

I. INTRODUCTION

Clinical trials are essential for the development of new treatments and cures for diseases. However, patient recruitment is often a major challenge for clinical trials, especially for rare diseases or those with complex eligibility criteria. This can lead to delays in trial initiation and completion, and can ultimately reduce the chances of success. Several studies have highlighted the challenges in patient accrual for clinical trials, particularly in rare diseases [1]. Additionally, the use of convenience-enhancing solutions, such as new technologies, can help improve patient engagement and retention in clinical trials [2]. Furthermore, the classification of clinical trials as 'negative' based solely on statistical significance does not accurately reflect the complexities and nuances of trial outcomes. Reclassifying trials into categories that consider factors such as power, termination, and initiation can provide a more accurate assessment of trial characteristics and

The associate editor coordinating the review of this manuscript and approving it for publication was Cesar Vargas-Rosales.

be more useful for patients and healthcare providers [3]. Recruitment for clinical trials continues to be a challenge, with increasing costs and complexity. Industry-sponsored phase III clinical trials have seen a significant increase in recruitment duration over the past 12 years, indicating less effectiveness compared to before [4], [5]. Game theory and ensemble learning have the potential to improve patient recruitment in clinical trials. Game theory can model interactions between patients, investigators, and sponsors, identifying strategies for more efficient recruitment [6]. Ensemble learning can combine predictions from multiple machine-learning models, improving the accuracy of patient recruitment predictions [7].

A. MOTIVATION

Recruiting patients for clinical trials in the context of rare diseases is challenging due to the limited number of eligible participants, geographical dispersion, and lack of awareness. These factors often lead to delays and lower success rates, as smaller patient pools make it harder to gather statistically significant data. Therefore, innovative approaches and collaborative efforts are crucial to address these obstacles and improve the efficiency of clinical trial recruitment in rare disease research. Recent technological advancements and the recognition of the need for a more nuanced approach to trial classification have created a ripe environment for improvement. Two promising avenues are game theory and ensemble learning. Game theory offers a strategic framework to optimize healthcare decisions and it could extend to patient-centric considerations, such as healthcare facility preferences, fostering trust and efficiency in healthcare. Meanwhile, ensemble learning demonstrates its potential in healthcare by amalgamating models to enhance predictive accuracy, consistently outperforming traditional methods in domains like disease classification with models like LSTM and random forests. Game theory models can be complex and difficult to develop, requiring careful consideration of parameters to ensure accurate predictions. Ensemble learning models, on the other hand, require large amounts of data to train effectively. Additionally, both game theory and ensemble learning models can be sensitive to the choice of parameters, which can further complicate the process of producing accurate predictions [8]. Nevertheless, leveraging these techniques effectively for clinical trial recruitment remains a challenge. These challenges highlight the need for further research and development to overcome these limitations and fully harness the potential of game theory and ensemble learning in clinical trial patient recruitment. In response, this paper introduces a hybrid model and game-theoretic approach to comprehensively address these challenges, offering a transformative solution to the intricacies of patient recruitment in clinical trials.

Major Contributions:

1) Collaborative Approach for Patient Recruitment in Clinical Trials: The paper introduces a novel two-step approach to optimize the patient recruitment process in clinical trials. In the first phase, a machine learning (ML) methodology is employed to assess and predict the likelihood of an individual being recruited for a clinical trial. This ML-driven process leverages key medical factors, such as blood parameters, to effectively identify suitable candidates. Following the ML-based patient selection, the second phase of the framework involves the application of game theory principles. Game theory is utilized to construct an optimization model that considers patients' beliefs, incentives, and overall strategic interactions. This step aims to enhance the efficiency of the patient recruitment process by strategically aligning incentives and fostering a cooperative environment. By distinctly implementing ML and game theory in separate phases, our approach maximizes the strengths of each methodology. The ML phase ensures the identification of eligible candidates based on medical criteria, while the subsequent game theory phase refines the patient recruitment strategy by addressing individual beliefs and optimizing the overall game dynamics. This dual-phase framework places patients at the forefront, prioritizing both medical suitability and strategic interactions to enhance the overall effectiveness of clinical trial recruitment.

2) Machine Learning Integration and Feature Extraction: The proposed approach utilizes machine learning techniques, specifically an autoencoder model, for feature extraction from a preferred dataset. This step aims to capture the relevant patterns and information from the data, which is essential for subsequent decision-making processes.

3) Stacked Ensemble Classifier for Enhanced Decision-Making: The stacked ensemble classifier harnesses the individual strengths of Logistic Regression's simplicity and interpretability, Support Vector Machines' capacity for handling complex data, Random Forest Trees' resilience to noisy data, and Decision Trees' intuitive decision rules. This collaborative approach creates a powerful decision-making tool for patient recruitment, enhancing accuracy by capturing both linear and non-linear relationships, managing complex and noisy data, and providing interpretable insights, ultimately improving the efficacy of clinical trial predictions.

4) Dynamic Interplay of Stakeholders in Game Theoretic Mode and Optimized Decision-Making Strategies for Stakeholders: The proposed 3-layer game theoretic model provides a comprehensive representation of the interactions between Patients, Doctors/Clinical Investigators, and Research Firms. By considering the dynamic interplay among these stakeholders, the model aims to optimize decision-making processes related to patient recruitment in clinical trials. This includes devising strategies that enhance patient recruitment efficiency while increasing the payoff for Doctors/Clinical Investigators.

B. PAPER ORGANIZATION

The paper's structure is organized into distinct sections to comprehensively address the challenges and solutions in patient recruitment for clinical trials within the healthcare

analytics domain. In Section II, we present a review of related works, focusing on prior research where machine learning and game theory have been applied, either individually or collaboratively, to tackle healthcare analytics problems or any other domain where they have combined to solve a problem. Section III elucidates the technical aspects of our methodology, elucidating the techniques employed to optimize patient recruitment in a generalized healthcare setting. This encompasses a detailed exposition of our hybrid classification model, game-theoretic framework, and their integration. Moving to Section IV, we delve into the evaluation metrics, conducting a rigorous comparative analysis to assess the performance of our proposed methods. We engage in an in-depth discussion and analysis of the results, providing insights into the strengths and limitations of our approaches, from an observational point of view. Section V serves as the conclusion, summarizing our research's key findings and their implications in the context of patient recruitment for clinical trials. Additionally, it highlights the novel contributions of our work. Section VI further delineates future research directions and acknowledges the limitations, charting a path for further advancements in patient recruitment optimization and healthcare analytics, while considering the challenges we encountered in this study.

II. RELATED WORKS

Patient recruitment for clinical trials continues to be a major challenge. Virtual or decentralized methods have shown promise in improving recruitment, retention, and diversity in clinical trials, with some studies reporting better results compared to traditional methods [9]. Additionally, decentralized methods have also been associated with improved participant retention [10]. Informatics interventions, such as natural language processing, have also been used to improve the efficiency and accuracy of eligibility determination and trial recruitment [11]. These interventions have shown potential in streamlining the process and reducing the labor-intensive nature of identifying eligible patients. Overall, the use of decentralized methods and informatics interventions can help address the barriers to patient recruitment in clinical trials, such being the methods of game theory and machine learning techniques.

Game theory is being increasingly recognized as a valuable tool in healthcare work [12]. It allows for the optimization of multiple objectives simultaneously, even when there are conflicting interests. Game theory also offers a framework for decision-making in clinical settings, such as clinical decision support systems, by considering the interactions between patients and healthcare providers [13]. Bilal et al. [14] proposed one such framework, which uses a Bayesian game-theoretic approach to optimize disease classification in healthcare systems. Game theoretic frameworks have worked around the ideal care of patients as well. Udok et al. [15] signified the patient-centric influence. The patients' preferences for healthcare facilities were assessed based on factors such as the costs of services and the attitude of healthcare providers, which were used to determine the optimal strategy and the value of the game. Game theory can be applied in healthcare to promote fair competition, trust, flexibility, and transformational leadership. The five principles necessary to lead an infinite game in healthcare are a just cause, a trusting blame-free team culture, acknowledging adversaries, existential flexibility, and courageous leadership [16]. Incorporating game theory into healthcare can enhance resource utilization, accuracy in disease detection, and overall efficiency in healthcare delivery [17].

Significant advancements have been made in recent years to optimize patient recruitment for clinical trials using machine learning methods [18]. Machine learning methods have also been applied to conduct response-adaptive randomization in clinical trials, resulting in more personalized optimal treatment assignments and higher overall response rates among trial participants [19]. Clinical trial recruitment is a continuing challenge, but efforts to improve study recruitment have been informed by theories of human decision-making and behavior change, such as shared decision-making (SDM) and the Theoretical Domains Framework (TDF) [20]. Automating the patient recruitment task using natural language processing and machine learning techniques has been explored, achieving high accuracy in determining patient eligibility for clinical trials [21]. A general framework for the practical application of XAI in medical research that can inform clinicians and validate and explain cancer biomarkers has been introduced as well [22].

Ensemble learning is a promising approach in healthcare that combines multiple models to improve prediction and diagnostic performance. It has been applied in various healthcare domains such as medication adherence prediction [23], classification of diabetic disease [24], prediction of post-partum hemorrhage [25], and diagnosis of hepatitis C [26]. These studies used ensemble learning techniques such as deep ensemble learning (DEL), gradient deep learning boosting, and memory-based and gradient boosting-based methods. The results show that ensemble models outperform traditional machine learning and deep learning techniques, achieving higher accuracy, recall, F1-score, and area under the curve values. The Shapley value, a concept from game theory, has been used to quantify the importance of models in ensemble games. Additionally, learning automata, a reinforcement learning technique, has been used to dynamically assign coefficients of influence to base learners in an ensemble. These approaches have shown promising results in improving the performance of ensemble learning models [27], [28]. While game theory and ensemble learning show promise for patient recruitment in clinical trials, there are still challenges that need to be addressed.

The introduction of feature extraction methods from machine learning models, particularly those based on data augmentation, has significantly transformed the landscape of data analysis and predictive modeling. In recent research, diverse methods have been explored to address challenges

in natural language processing (NLP) and text analysis. One approach involves combining graph-based neural networks and genetic algorithms, creating a method (GTR-GA) that generates high-quality augmented text data, overcoming issues related to data scarcity in NLP [29]. Additionally, hierarchical graph-based text classification frameworks, incorporating contextual node embedding and BERT-based dynamic fusion, have been proposed to enhance text classification accuracy [30]. For sentiment analysis, a novel bidirectional convolutional recurrent neural network architecture with a group-wise enhancement mechanism has shown superior performance, outperforming existing methods and proving beneficial for tasks requiring context understanding [31]. To tackle the labor-intensive process of creating labeled data, researchers have developed automated methods like SRL-ACO, leveraging Semantic Role Labeling and Ant Colony Optimization. These techniques generate additional training data for various NLP models, enhancing accuracy without the need for manual annotation [32]. Furthermore, in bibliometric data analysis and topic extraction from scientific literature, an innovative two-stage framework combining word embedding schemes and ensemble clustering methods has exhibited improved performance, surpassing baseline methods in predictive measures and accuracy [33]. These advancements underscore the diverse and impactful strategies emerging in the field of NLP and text analysis.

Combining game theory and machine learning for decision-making frameworks has been explored in several papers. One approach is to integrate learning with computational game theory to address societal challenges such as security and sustainability [34]. Combining game theory and machine learning in healthcare can enable personalized, adaptive treatment strategies, ensuring tailored and effective patient care. By leveraging the strengths of both game theory and machine learning, these frameworks offer promising solutions for decision-making problems.

The literature review of previous methods is summarized in Table 1, providing concise descriptions of the employed techniques, research advantages, and identified limitations.

A. PROBLEM STATEMENT

Patient recruitment for clinical trials represents a critical bottleneck in healthcare research, characterized by inefficiencies that can lead to substantial delays, escalated costs, and compromised trial outcomes. The need for an effective and efficient patient recruitment strategy is underscored by several key factors. Firstly, the inherent complexity of modern clinical trials, particularly those focused on rare diseases or therapies with stringent eligibility criteria, poses a formidable challenge to identifying and enrolling suitable participants. This complexity often results in prolonged recruitment timelines and can jeopardize the statistical power and integrity of the trials. Secondly, the existing paradigms for classifying clinical trials, primarily based on statistical significance, are overly simplistic and fail to capture the multifaceted nature of trial outcomes. Relying solely on binary outcomes, such as 'successful' or 'failed,' hinders a nuanced understanding of the trial's dynamics, including power, termination rates, and initiation criteria. Furthermore, the intersection of machine learning and game theory in the context of healthcare analytics presents a promising avenue to address these recruitment challenges. However, there remain limitations and challenges to be overcome. Existing studies often lack a comprehensive hybrid model that effectively integrates machine learning techniques and game-theoretic approaches, creating an opportunity for more advanced and synergistic solutions. Moreover, the practical implementation of these sophisticated methodologies requires careful consideration of real-world constraints, including data availability, computational resources, and ethical considerations, which necessitates a tailored and pragmatic approach. This study endeavors to bridge these gaps by introducing a hybrid classification model and a game-theoretic framework. This unified approach seeks to optimize patient recruitment for clinical trials while accounting for the intricacies of real-world healthcare settings and the complexities inherent in-patient selection. By seamlessly merging the realms of machine learning and game theory, this study endeavors to offer a comprehensive and pragmatic answer to elevate the efficiency and efficacy of patient recruitment within the intricate landscape of clinical trials. This innovative approach leverages generic parameters, offering adaptability for diverse organizations to tailor the methodology to their specific needs seamlessly.

III. METHODOLOGY

A synergistic approach has been developed to optimize the generic patient recruitment process for advanced clinical trials. This multifaceted framework uses machine learning techniques and basic game theory principles and applications. We have selected a preferred dataset, having blood parameters for patient eligibility in clinical trials, which is akin to employing a comprehensive set of features or variables for a machine learning classification task. Each blood parameter is analogous to a feature, much like pixel values in an image or data points in a dataset. These parameters provide a multidimensional representation of a patient's health status, including various biomarkers, concentrations of substances, and physiological indicators. Machine learning models can then process and analyze these parameters collectively to make accurate predictions regarding a patient's suitability for a clinical trial. The models use the patterns and relationships between these parameters, essentially learning to differentiate between eligible and ineligible patients. This approach allows researchers to leverage the extensive information contained in blood parameters to enhance the precision and objectivity of patient classification, similar to how machine learning models use multiple features to classify data points in various other domains. The extracted features are then input into a super-classification model, which includes Logistic Regression, Support Vector Machines, Random Forest Trees, and Decision Trees. This orchestrates a stacked ensemble

TABLE 1. Literature survey.

Ref. No	Techniques Used	Benefits	Challenges
[11]	NLP driven Cognitive walkthrough with a think-aloud protocol and a Post-Study System Usability Questionnaire	The NLP-driven tool has gained acceptance among clinical research staff, demonstrating successful usability. It autonomously generated queries to identify eligible participants from clinical databases, showcasing semi-autonomous functionality. Notably, the tool showed high usability, especially in the final iteration cycles. These results highlight the importance of employing an iterative usability evaluation approach, emphasizing its role in refining NLP systems. This iterative process contributes to improving the overall usability and acceptance of such systems among end-users.	The findings suggest the tool is user-friendly for clinical research staff, prompting a call for input from a wider audience. This input is vital for refining the Natural Language Processing (NLP) tool for eligibility prescreening. Given the complexities of eligibility criteria and the risk of imprecise cohort definitions with NLP alone, the authors endorse an approach that combines human and machine intelligence, forming a dynamic decision-making framework.
[12]	A Game Theory Based Clustering Scheme	It suggests a clustering scheme grounded in game theory coupled with the TDMA method, enhancing network longevity, data transmission, and reducing overall energy consumption. This paves the way for integrating game theory into scenarios requiring optimization.	A notable limitation lies in assuming rational behavior among nodes, as game theory relies heavily on participants making rational decisions. In real-world scenarios, nodes may not always act rationally due to factors like resource constraints, environmental conditions, or faulty hardware. Moreover, the effectiveness of the GCS heavily relies on the accuracy of information shared among nodes during clustering, and inaccurate or incomplete data may lead to suboptimal clustering decisions, impacting overall network performance. Additionally, the proposed GCS may encounter challenges in dynamic and unpredictable environments, as clustering decisions made during setup phases may not adapt well to rapidly changing network conditions. This underscores the need for a dynamic, mathematically based reasoning model to address uncertainties and adapt to varying conditions.
[13]	Decentralized Game Theoretic Mobile Edge Computing Enabled 5G Health Monitoring	The research explores the utilization of MEC-enabled 5G for in-home health monitoring in IoMT, addressing both intra-WBANs and beyond-WBANs. In intra-WBANs, the study formulates a bandwidth scheduling problem as a bargaining game, employing the Nash bargaining solution for optimal allocation. For beyond-WBANs, a decentralized approach based on weighted potential games resolves the non-cooperative game, achieving a Nash Equilibrium. The study provides theoretical insights into the upper bounds of time complexity and the number of patients benefiting from MEC. Through performance evaluations, the research underscores the solution's efficacy in minimizing system-wide costs and maximizing patient benefits from MEC, contributing significantly to healthcare advancements.	The algorithm claims Nash equilibrium, but practical challenges in real-world scenarios, such as diverse patient behaviors, dynamic network conditions, or unforeseen technological constraints, may disrupt its attainment. This highlights a limitation in adaptability to health monitoring systems' complexities, compromising its effectiveness in unpredictable environments. To overcome this, essential integrations include machine learning (ML) and dynamic learning mechanisms. ML analyzes historical data, enhancing decision-making in novel scenarios, while dynamic learning ensures continuous adaptation to changing patient behaviors and network conditions, making the algorithm more robust and versatile.
[14]	Co-operative Bayesian game-theoretic model	The model's use of Bayesian Nash Equilibrium offers an advantage in handling incomplete information between the involved parties, enhancing its accuracy in healthcare decision-making. The model's enhanced accuracy reduces the likelihood of forgery. Additionally, thorough statistical analysis ensures results are unbiased, emphasizing Bayesian models without prior knowledge.	The current method faces limitations in accuracy and interpretability. Future improvements suggest incorporating diverse methods, including polynomial mutation, simulated binary crossover, and game theory. The objective is to investigate their effects, anticipating multiple reference points for enhanced decision-making.
[15]	Cross-sectional descriptive study and purposive sampling technique	It allows researchers to obtain information on respondents' demographic data and the questions about the study. The findings indicated that public hospitals drew more patients based on service costs, while private hospitals attracted patients due to healthcare providers' attitudes. Patients favored public hospitals over private ones.	The study lacks information on the generalizability of findings to other healthcare facilities in Akwa Ibom State or Nigeria. Using a descriptive cross-sectional design with questionnaires for grouping respondents into public and private hospital preferences, the dataset includes 9976 responses out of 10,000 distributed questionnaires, indicating a 0.24% non-response rate. This limited approach may not fully represent the population, impacting the findings' generalizability. To enhance the study's robustness, there is a need for a more comprehensive and diverse dataset, employing randomized sampling methods to draw more conclusive conclusions about patients' hospital preferences.
[17]	Internet of Things (IoT) technology to acquire real-time ambient data inside	Utilizing embedded IoT sensors, it enables real-time acquisition and assessment of healthcare resources, offering a valuable solution for dynamic healthcare environments. The proposed two-player game model	Relying on simulated validation raises concerns about the model's real-world adaptability, warranting scrutiny for practical implementation. The effectiveness claim, supported by comparative analysis, demands a deeper

TABLE 1. *(Continued.)* **Literature survey.**

	smart hospitals	introduces an effective decision-modeling technique, providing structure and insight into healthcare resource management. Rigorous validation in a simulated environment showcases high classification efficacy (93.74%) and efficiency metrics, establishing its potential for accurate and efficient decision-making.	exploration of healthcare complexities for robustness. This necessitates refined approaches with machine learning (ML) and other techniques to optimize parameters and enhance the game theory model's efficacy. Furthermore, future research recommendations on data security and network bandwidth optimization lack in-depth insights into challenges or methodologies, presenting opportunities for further exploration and development.
[18]	Two baseline models for constant study-site level enrollment rate prediction, with one based on historical enrollment rates and the other an XGBoost model, and three monthly enrollment prediction models employing LightGBM, Zero-Inflated Poisson (ZIP) regression, and a family of hurdle models with specific count distributions,	Improved prediction accuracy over traditional baselines, especially in terms of study-level MAE and MSE, which are crucial for study enrollment planning. Additionally, the model selection is based on both predictive performance and efficient training times, resulting in a practical and effective solution for clinical trial enrollment prediction.	Certain models, such as the ZIP model, require time-intensive training, potentially affecting real-time decision-making with durations exceeding a day. The heavy reliance on simulated environments for validation raises concerns about the model's adaptability to real-world scenarios and potential variations. The research lacks a comprehensive exploration of intricate complexities in diverse healthcare settings, limiting its applicability. Focusing on specific large pharmaceutical companies may constrain the generalizability of findings to a broader healthcare context. Assumptions about constant enrollment rates may not align with the dynamic nature of patient recruitment in practical clinical scenarios. Finally, data limitations pose challenges in generating meaningful enrollment rate estimates, impacting the overall robustness of the model.
[20]	Shared Decision-Making (SDM) and Theoretical Domains Framework (TDF), to guide the investigation of recruitment strategies	Applying the Theoretical Domains Framework (TDF), the research employs a theory-informed approach to categorize factors, revealing the complexity of recruitment issues. Findings demonstrate the relevance of nearly all TDF domains to trial recruitment behaviors, uncovering nuances overlooked in existing survey instruments. The TDF approach suggests specific interventions to overcome identified barriers by linking domains to behavior change techniques. The study proposes targeted strategies like problem-solving, social support, instruction, goal setting, and environmental restructuring for improving recruitment in the SAVER trial. This method could serve as a foundation for developing formal training for recruiters in future trial phases or other clinical trials, with recommendations tailored to specific domains.	The study's limitations include a small sampling frame consisting only of non-physician recruiters from a single multisite study, limiting generalizability. Potential issues related to data saturation, sampling, and participant understanding are acknowledged, urging the need for broader research with diverse samples and physician recruiters. While following the 10 plus 3 rule for data saturation, the study recognizes the possibility of new themes emerging with additional interviews. Moreover, challenges associated with the language and concepts of Shared Decision Making (SDM) literature were identified, indicating potential ambiguity and unfamiliarity among recruiters, suggesting the need for refining language in future research.
[21]	Optimal patient enrollment in clinical trials with interim analyses	This study tackles optimal patient enrollment in clinical trials using dynamic programming with Bayesian updates. It introduces backward approximate dynamic programming, demonstrating monotonicity in the value function and establishing an optimal enrollment policy based on average treatment responses. Illustrated with a clinical trial program, the study conducts sensitivity analyses and reduces computation time through innovative algorithms. The findings provide valuable insights for trial management, potentially boosting drug profit for pharmaceutical firms implementing the optimal policy.	The model's reliability in practical applications depends on high-quality data and sensitivity to input parameters, requiring continuous adjustments to changing trial conditions. Insights from the optimal policy and sensitivity analyses rely on assumed values for the dropout rate and prior probability distribution, impacting the model's robustness. Illustrations with a specific firm's trial program raise questions about the generalizability of optimal policies to diverse trial contexts. Users should exercise caution in directly applying ADP algorithms to patient enrollment strategies, validating efficacy in diverse settings due to potential variations in trial characteristics.
[22]	Explainable AI	Demonstrate ML pipeline efficacy, consistency, and advantages over traditional statistics in biomarker discovery. The robust ML pipeline with Shapley analysis aligns with conventional analyses, enhancing interpretability and transparency. Shapley values contribute to outcome transparency, fostering trust and offering valuable insights. The study acknowledges ongoing research efforts to improve Shapley values, ensuring methodological advancement. The combined use of Shapley values and other aspects provides a cohesive approach for medical research, emphasizing explainability to address accountability concerns and promote understanding and trust, particularly among domain experts.	Despite the pipeline's demonstrated effectiveness, challenges include higher computational costs, especially in extensive cohorts. Shapley values enhance result explainability but require domain expertise for insightful interpretation. The lack of a universal definition for explainability in medical applications poses challenges, relying on subjective evaluations. Shapley analysis offers correlation insights, emphasizing caution in inferring causal relationships or 'true intentions' of ML models. Its post-hoc nature focuses on explaining diagnoses rather than improving model classification. The study's simplified ovarian cancer model raises concerns about applicability to diverse disease stages, urging awareness of potential clinical limitations. The absence of a direct comparison with existing literature and variation in AUC-ROC

TABLE 1. *(Continued.)* **Literature survey.**

			scores emphasizes the need for cautious interpretation and further validation in diverse datasets and methodologies.
[24]	ML ensemble-based pipelining	The study introduces a focused predictive analytics methodology for identifying future non-adherence in patients. Utilizing IoT-enabled sharps bins ensures data consistency and reduces human errors, offering generalizability. The absence of ground truth labels minimizes label noise, enhancing model reliability. The research highlights potential targeted patient interventions, avoiding unnecessary interventions for those managing their treatment independently. Predictive analytics offer valuable insights for personalized healthcare strategies from a patient-centric perspective.	The reliance on historic injection disposal records from sharps bins as the primary data source may omit patient-related features such as gender, age, and race, which have been considered in previous studies. The study recognizes the need for further exploration of explainability approaches, emphasizing the importance of building trust in the context of AI applications in healthcare.
[25]	Machine learning based ensemble approach consisting of combinations of memory based and gradient boosting-based methods	The study introduces an ensemble classifier with promising results, achieving 81.8% classification efficiency. Evaluation metrics include accuracy, precision, recall, and ROC AUC. Incorporating diverse learning algorithms enhances model performance. 5-fold cross-validation ensures system reliability, robustness, and generalizability. The ROC curve and AUC analysis visually represent the classifier's discriminatory power, providing a comprehensive evaluation.	The study overlooks extrinsic factors influencing diabetes, potentially limiting system value. Mentioning future incorporation of microRNA (miRNA) levels introduces a promising avenue, but specific profiles and validation need exploration. Lack of details on the ensemble method's computational aspects hinders assessing reproducibility. The proposed collaboration with LifeGene Bioscience lab for miRNA validation warrants careful ethical and procedural consideration to ensure findings' robustness and reliability.
[26]	Ensemble Learning	The study introduces an ensemble learning (EL) method for predictive diagnosis in postpartum hemorrhage (PPH) and disseminated intravascular coagulation (DIC), comparing its performance with logistical regression (LR) and artificial neural networks (ANN). The large clinical dataset (3842 records for PPH and 212 for DIC) contributes to study robustness. The ensemble method integrates four basis methods, achieving high accuracy (96.7% for PPH and over 90% for DIC). The model showcases efficacy in predicting PPH risk and assessing critical levels in PPH patients, highlighting the capabilities of ensemble methods.	The study lacks detailed insights into the interpretability and explainability of the ensemble learning model, hindering a comprehensive understanding of its decision-making process. The absence of specific information on hyperparameter tuning for individual basis methods raises questions about reproducibility and optimization. While the comparative analysis with LR and ANN highlights the superior performance of the EL method, it lacks a deeper exploration of the reasons behind these differences. A decision-making framework is needed to validate the scalability of the system.
[27]	A fuzzy ensemble of deep learning models	The study contributes to breast cancer histology image classification by integrating deep convolutional neural networks (DCNNs) and a fuzzy ensemble approach. Introducing Choquet Integral for fusion innovates over traditional aggregation methods. The novel method for calculating fuzzy measures enhances ensemble methods' robustness and efficiency. Transparent use of the publicly available BACH dataset ensures experiment reproducibility and validation. The study sets a benchmark in breast cancer histology image classification, providing valuable insights for future research.	The proposed method, involving fusion of multiple deep learning models, may demand significant computational resources and time. The study could offer more in-depth insights into the interpretability of Choquet Integral and the ensemble method's decision-making process for practical understanding. Additionally, exploring future directions and challenges in implementing the method at a larger scale or in clinical settings could enhance its practical applicability.
[28]	Reinforcement Learning for Ensemble based ML models	The article introduces an innovative model selection method for electricity peak load forecasting using the Double DQN algorithm. Incorporating individual machine learning models like ANN, SVR, and DBN enhances forecast accuracy comprehensively. The efficient hyperparameter tuning method based on fractional factorial design reduces computational efforts compared to traditional methods. Empirical testing on real data from Thailand's electricity peak load demonstrates the superior performance of the proposed Double DQN over individual models and other ensemble models. Focusing on Mean Squared Error (MSE) affirms the practical significance of the approach in enhancing forecasting precision.	The proposed future direction of prioritizing forecast values higher than the actual peak load for sufficient power supply suggests potential limitations in the current strategy for specific forecasting scenarios. The call to expand the pool of forecasting models to include other time series and deep learning models highlights a potential limitation in the model's applicability to diverse datasets. Applying the proposed method to forecast total electricity consumption instead of peak load adds complexity and potential challenges, necessitating cautious generalization. Lastly, testing the model on data from Thailand prompts further exploration of its performance across datasets from other countries or those with different characteristics to establish broader applicability.
[29]	Graph neural networks and Genetic Algorithm	Capable of generating high-quality augmented text data, improving the performance of various downstream NLP tasks such as sentiment analysis and text classification	Cannot address the issue of missing classes in small datasets. For instance, if a dataset lacks certain classes, such as golden retrievers in a dog breed classification task, Data Augmentation cannot generate those missing classes, highlighting its limitation in such cases.
[30]	Hierarchical graph-based text classification	Framework excels in achieving high accuracy, leveraging the strengths of combining diverse approaches, and effectively handling complex datasets, offering valuable	Models utilizing graph-based approaches performed less effectively than those without, indicating potential limitations in graph-based text classification models.

TABLE 1. *(Continued.)* **Literature survey.**

	framework with contextual node embedding and BERT-based dynamic fusion	insights for the development of accurate text classification models in natural language processing applications.	Additionally, the study highlighted that the performance of text classification models can vary significantly depending on dataset characteristics, implying the need for tailored approaches for specific datasets.
[31]	Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism	The article empirically evaluates fourteen deep learning architectures and six variants for sentiment analysis across eleven datasets. Word2vec outperforms fastText and GloVe in the embedding layer. Introducing a group-wise enhancement mechanism with spatial attention demonstrates a novel approach, outperforming traditional attention mechanisms. The RCNNGWE architecture stands out, achieving the highest predictive accuracies across multiple datasets in sentiment analysis. Insights into bidirectional GRUs and LSTMs emphasize their efficacy in capturing global semantic features and modeling long-term dependencies. The RCNNGWE architecture delivers high predictive performance for both short and long review documents, improving with larger training sets.	While the focus on classification metrics is crucial, discussing potential trade-offs or challenging scenarios could enhance the article's completeness. The claim of RCNNGWE outperforming state-of-the-art results needs further scrutiny considering dataset characteristics. The assertion that higher predictive performances with larger training sets implies resource dependency that may not be universally applicable. The empirical results mention percentage increase rates for the MR dataset, but a more comprehensive discussion on their significance would strengthen the article. The conclusion lacks explicit mention of potential future directions or areas of improvement for RCNNGWE, leaving the reader curious about enhancements or refinements.
[32]	A text augmentation framework based on semantic role labeling and ant colony optimization	The empirical results highlight the crucial role of augmentation methods in significantly improving the predictive performance of deep learning models in text classification tasks. SRL-ACO, AMR-ACO, and SRL-PSO consistently demonstrate superior performance. The article underscores the importance of selecting the right dataset and architecture, offering nuanced insights into LSTM and GRU performance. The impact of training size on performance varies across datasets and architectures. SRL-ACO is presented as a valuable contribution, effectively addressing the challenge of limited training data in NLP tasks. Its effectiveness in generating high-quality synthetic data, stability across architectures, and potential for interpretability are significant advantages. The conclusion aptly summarizes the importance of text augmentation, introduces SRL-ACO, and provides compelling evidence of its effectiveness in improving NLP model performance.	The claim that SRL-ACO outperforms other augmentation techniques warrants further scrutiny, considering potential nuances in task-specific scenarios or dataset characteristics. Mentioning the impact of SRL-ACO on other NLP tasks as a future direction, the article could explore potential challenges or considerations when extending the method. While stability across architectures is noted as a strength, a nuanced discussion about architecture-specific nuances or scenarios where performance gains may vary would enhance completeness. Exploring the computational complexity or resource requirements of the SRL-ACO framework could provide a more comprehensive understanding for practitioners implementing the proposed method.
[33]	Two-Stage Topic Extraction Model for Bibliometric Data Analysis Based on Word Embeddings and Clustering	The claim that word embedding schemes combined with clustering approaches outperform baseline results warrants attention, considering potential nuances in task-specific scenarios or dataset characteristics. While highlighting the effectiveness of the latent Dirichlet allocation (LDA) method in topic extraction, the study could discuss potential challenges or considerations in using LDA compared to conventional clustering algorithms. The empirical comparison of word embedding schemes reveals the excellence of the LDA2vec scheme, but potential limitations or scenarios where it may not be the best choice could enhance the article's completeness. The analysis of vector size, projection layer dimension, and average words per sentence provides valuable insights, but a discussion on the practical implications and potential trade-offs of these findings would strengthen the article.	The research overlooks computational complexity and scalability concerns, potentially limiting its utility for larger datasets. The study lacks a thorough discussion of potential limitations or challenges associated with the proposed two-stage topic extraction model. The claim that the clustering ensemble framework outperforms conventional clustering baselines needs a more nuanced exploration, considering the diversity of clustering tasks and datasets. While the study suggests the effectiveness of the ensemble word embedding scheme, further investigation into its robustness across different domains or task-specific scenarios would strengthen the argument.
[34]	Game Theoretic Model for Explainability of ML models	The article introduces an explanation game contributing to normative and descriptive dimensions for interpretable machine learning. The formal introduction of the "explanation game" as a collaborative framework reflects a proactive stance, fostering dialogue in algorithmic development. The optimistic outlook on its contributions to ethical, epistemological, and scientific discussions indicates a forward-looking approach to challenges posed by black box algorithms.	The study hints at challenges in relaxing foundational assumptions, requiring further research. While emphasizing interpretability, it acknowledges not providing a comprehensive solution for practical machine learning complexities. Normative challenges like fairness and transparency are touched upon, but explicit strategies or solutions are lacking. Detailed insights into how the explanation game addresses these challenges would strengthen practical implications. The future work lacks a clear timeline and specific milestones, making proposed advancements in the explanation game somewhat ambiguous. Brief mention of prior work on algorithmic fairness optimization could be expanded for a comprehensive understanding of the existing research landscape.

mechanism that harnesses the strengths of each classifier. In simpler terms, we combine the best parameters or capabilities of each classifier to produce a stacking classifier that imputes the final result in the form of numerical odds of the binary. Furthermore, after the implementation of the machine learning phase, our interdisciplinary study delves into the domain of game theory. We propose a 3-layer game approach motivated by the framework proposed by Rawat et al [35]. The 3-layer game theoretic model weaves together the roles and duties of Patients, Doctors/Clinical Investigators, and Research Firms. In this dynamic interplay, strategies and formulas are devised to optimize the decision-making process for effectively recruiting patients. The ML Model proposed will increase the payoff from the Doctors' side by helping doctors effectively recruit patients based on medical parameters first, enhancing the recruitment process and identifying those who are most likely interested in trials too. The architecture diagram for the proposed framework is given in Figure 1.

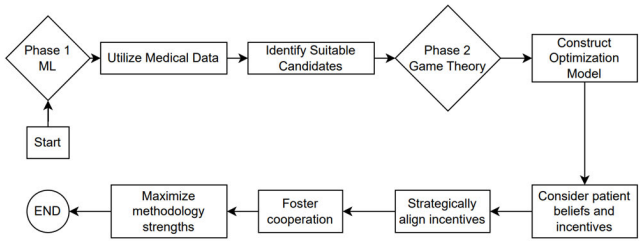


FIGURE 1. The proposed framework architecture.

Note: *is important to note that within each discipline—machine learning methodology and game theory methodology—there may be instances where certain variables or concepts are used. These variables are specific to the content and scope of their designated discipline and should not be confused with variables from different sections of the methodology. Each discipline employs its own set of variables and methodologies tailored to its objectives and analysis, ensuring a clear and focused approach within its respective context. This separation of variables maintains clarity and prevents any cross-disciplinary confusion.*

A. PRELIMINARIES

Before delving into the details of our novel approach to optimizing patient recruitment for clinical trials using a hybrid classification model and game-theoretic framework, it is essential to establish the foundational concepts and components that underpin this research. These preliminaries provide the necessary background knowledge and context for a comprehensive understanding of our methodology.

1) CLINICAL TRIALS

Clinical trials are fundamental to the development of new medical treatments and therapies. These controlled studies involve human participants and are designed to evaluate the safety and efficacy of new drugs, therapies, or medical interventions. Understanding the various phases of clinical trials, ethical considerations, and the significance of patient recruitment is vital for comprehending the challenges addressed in this research.

2) PATIENT RECRUITMENT CHALLENGES

Efficient patient recruitment is a critical factor in the success of clinical trials. Challenges in identifying and enrolling eligible participants can lead to delays, increased costs, and compromised trial outcomes. These challenges are exacerbated in the context of rare diseases and trials with stringent eligibility criteria.

3) MACHINE LEARNING

Machine learning is a subset of artificial intelligence (AI) that focuses on developing algorithms and models capable of learning from data to make predictions or decisions without explicit programming. Familiarity with machine learning concepts, such as supervised learning, ensemble methods, and deep learning, is essential to grasp the intricacies of our hybrid classification model. Classification models are a category of machine learning algorithms used to categorize data into predefined classes or labels. Understanding various classification techniques, including logistic regression, support vector machines, random forests, and decision trees, is crucial for comprehending the components of our hybrid classification model. Ensemble learning involves combining multiple machine learning models to improve predictive performance. Concepts such as model aggregation and stacking play a pivotal role in our proposed work.

4) GAME THEORY

Game theory is a mathematical framework used to analyze and model strategic interactions among rational agents. In our research, game theory is employed to model the interactions between patients, investigators, and research firms in the context of patient recruitment for clinical trials.

B. MACHINE LEARNING IMPLEMENTATION

In this machine learning implementation, we employ the Hybrid Classification Model Training algorithm (Algorithm 1) to tackle the critical challenge of optimizing patient recruitment for clinical trials. The algorithm utilizes an ensemble of base classification models, including Decision Tree, Logistic Regression, Support Vector Machine, and Random Forest, in conjunction with feature encoding via Autoencoder (AE). We initialize essential components such as encoded dimensions, stacking classifiers, and meta-classifiers. The training process involves iterative hyperparameter tuning for each base model, with a focus on finding the optimal configurations. During this process, we train each base model, store them in the stacking classifier, collect their predictions on the input data, and compute losses and gradients for fine-tuning. Finally, we train the meta-classifier to effectively combine the base models' predictions and make the binary patient recruitment decision. This comprehensive machine learning implementation is designed to enhance

patient recruitment efficiency for clinical trials through the fusion of feature engineering and ensemble learning techniques. Subsequent sections will evaluate and discuss the effectiveness of this approach.

Our Hybrid Classification Model Training approach stands out for its precision in feature encoding, leveraging Autoencoders for accurate dimensionality reduction. The integration of ensemble learning enhances prediction accuracy and reduces model bias, bolstered by meticulous hyperparameter optimization. An adaptive meta-classifier dynamically combines base model predictions, expediting patient recruitment and minimizing trial delays. This data-driven approach empowers informed decision-making, leading to more successful clinical trials and improved healthcare outcomes. Figure 2 displays an illustration representing the suggested machine learning framework.

Algorithm 1 Hybrid Classification Model Training

Input:
- Input Data: input_data
- Data encoded using Autoencoder (AE)
- Classification Models: Decision Tree (DT), Logistic Regression (LR), Support Vector Machine (SVM), Random Forest (RF)
- Hyperparameter Settings for each model

Output:
- Stacking Classifier: stacking_classifier, predicting binary labels

Initialization:
- encoded_dimesnions of AE
- Stacking Classifier: stacking_classifier
- Meta Classifier: meta_classifier

Preprocess input_data
For each model in [DT, LR, SVM, RF]:
  While hyperparameter settings are not exhausted:
    Train Classifier:
      - **Train** model with hyperparameters
      - **Store** model in stacking_classifier
      - **Collect** predictions for input_data
      - **Compute** loss
      - **Compute** gradient
  End while
End for

Train Meta Classifier:
- **Train** meta_classifier with predictions from stacking_classifier

End Algorithm

1) AUTOENCODERS FOR FEATURE EXTRACTION

Autoencoders play a pivotal role in the realm of feature extraction, especially when it comes to complex datasets, such as those entailing intricate blood parameters for clinical trial selection. They are indispensable due to their unique capability to discern and highlight relationships within high-dimensional data. By converting the convoluted web of input features into a concise feature vector, autoencoders pave the way for efficient and accurate processing. These reduced-dimension representations serve as a crucial bridge, facilitating the seamless integration of machine learning algorithms and classification systems.

The significance of autoencoders becomes even more apparent when considering their capacity to outperform raw datasets in predictive tasks. These neural networks are not just about dimensionality reduction; they excel at capturing intricate patterns, subtle correlations, and latent relationships within the data. This proficiency is invaluable in the context of clinical trials, where precise patient selection based on blood parameters is paramount. The methodology behind autoencoders is grounded in their ability to learn from the input data itself. Through a training process, they encode the data and subsequently decode it to minimize reconstruction errors. The learned weights of the encoder layers, which effectively distill the most relevant information from the input, are then harnessed as the extracted features. In the context of clinical trial patient selection, this means that autoencoders can identify and encapsulate the crucial blood parameter interactions and attributes essential for making informed decisions. Their role in feature extraction is indispensable, as it empowers healthcare researchers to navigate complex datasets and unveil hidden insights that might otherwise remain concealed.

Autoencoders are neural networks that may be applied to a variety of tasks, such as text data processing, to extract features. They can convert high-dimensional input data into a feature vector with fewer dimensions so that it may be fed into classification or machine learning algorithms. Numerous research has demonstrated the efficacy of using autoencoders for feature extraction [36] Even when the extracted features have dimension sizes that are only a small portion of the total number of features, autoencoders can nevertheless perform predictions well. Autoencoders have even occasionally outperformed using the source datasets in some circumstances [37]. The methodology for using autoencoders for feature extraction involves training the neural network to reconstruct the input data and then using the learned weights of the encoder layers as the extracted features [38].

Encoder and Decoder: Autoencoders are made up of two independent components: the encoder, which assigns encodings to each input, and the decoder, which reconstructs the original input using the encoded input data. Pre-processed data is inputted into the input layer, and the output layer, also known as the decoded layer, should closely match the original input data with a minimum amount of error. The encoder condenses each blood parameter into a compact encoding, while the decoder reconstructs the original data, aiming to minimize errors. This encoding captures vital relationships within the data. The decoder, using these encoded representations, recreates the original data, ensuring it closely matches the input. By minimizing the loss function, the encoder-decoder system

fine-tunes parameters for accurate data representation, facilitating precise clinical trial patient selection. By minimizing the supplied loss function in generalized form, parameters are permitted to fit into the data in

$$J(i) = \Sigma x L(x, f_i(x)) \quad (1)$$

- J(i) is the overall loss or cost function that you want to minimize.
- Σx denotes the summation over all instances in your dataset.
- $L(x, f_i(x))$ represents the individual loss for a particular instance x_i, where $f_i(x_i)$ is the predicted output generated by a model with parameters i.

Encoded Dimensional Layout: Judging from the data, we can keep our favored size of the encoded variable dimensions, along with some more viable layouts [39].

2) STACKING CLASSIFIER-ENSEMBLE METHOD

To validate the encoded data produced by the autoencoder model, Decision Trees, Random Forests, Logistic Regression and Support Vector Machines were used too. The purpose of autoencoders is to abstract and meaningfully represent the encoded dimensional layout by storing the most complex and important aspects. Encoded data is not only simpler to comprehend than the original data but also smaller or has fewer dimensions. The classifier generalizes the representation to test the brand-new, untested data. To fine-tune the hyperparameters and perform k-fold Random Cross Validation for each model, RandomizedSearchCV [40] was required to produce the best classification model results. We create a hyperparameter search space for each classification model using a dictionary, where we store a range of potential values for various hyperparameters that we want to modify. For future calculations, the optimal hyperparameter configuration is saved. Table 2 illustrates the hyperparameter configurations for each model, showcasing the specific parameter values that led to optimal individual model performance.

We employ a heuristic technique to combine these models, placing them in a stack-type structure, to increase the model's performance because each model has a unique manner of computing the result and we can't truly determine what goes on inside the processing and model training process. It consists of four base classifiers—sometimes referred to as level-0 classifiers—Decision Tree Model, Logistic Regression Model, Support Vector Machines Model, and Random Forest Model, as well as one meta classifier that combines all of the level-0 classifiers' predictions with their optimal hyperparameter settings. The structural diagram of the suggested model is shown in Figure 2.

C. GAME FORMULATION

Following the successful implementation of machine learning methods to classify patient eligibility based on blood parameters, we now shift our focus to the realm of game theory. This independent phase, occurring immediately after the machine learning process, revolves around the dynamic interactions and decision-making of key stakeholders, including doctors, patients, and research firms. It is here that we construct a robust framework to strategically plan and manage patient participation in clinical trials, transcending the confines of data-driven eligibility assessments.

This paper presents a 3-layer game approach incorporating a static-repetitive manner to effectively study the interactions between the players, as well as the immediate gains and long-term incentives. The suggested framework would work on a "Patient-Centric Stratified Influence", to prioritize

TABLE 2. Hyperparameter configuration for each model.

Decision tree classifier	min_samples	20
	max_leaf_nodes	256
	max_features	0.2
	max_depth	12
	criterion	gini
	class_weight	{0:1, 1:3}
logistic regression	penality	12
	max_iteration	100
	fit_intercept	TRUE
	class_weight	{0:1, 1:1}
	C	1
Support vector classifier	Kernal	rbf
	gamma	1
	degree	4
	C	100
Random forest classifier	n-estimators	150
	min_samples_split	2
	min_samples_leaf	2
	max_features	sqrt
	max_depth	20
	bootstrap	FALSE

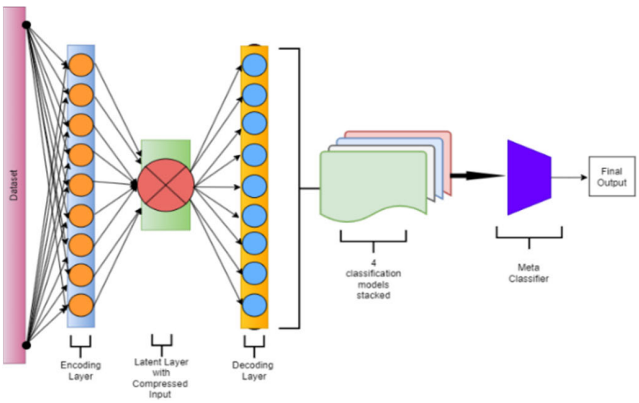


FIGURE 2. Structural diagram of the ML model implemented.

patients, then the doctors, and finally, the research firms. Patients, Doctors, and Research firms, in our setting, would work on strategic profiles of Producers, Intermediaries, and Consumers respectively.

Patients would have the most direct influence based on personal health, risks, benefits, and convenience. Research firms would be the ones managing and designing trials, setting certain criteria for eligibility, and providing incentives for participation if needed. Doctors or Clinical Investigators would help identify potential candidates adhering to the trial practice protocols. They would be obliged to communicate with patients regarding all the trial practices, guiding and supporting them along the way.

1) NORMAL GAME FORMULATION

Judging from the influence, patients are the primary stakeholders. They have the liberty to decide if they want to continue in the trial or not, any time in the trial process. Research firms are the organizations which would fund and conduct the trials. These organizations would ensure ethics and would determine steps to keep safety and efficiency of treatments. Doctors or Clinical Investigators will be responsible for providing medical care and collect data on trials, for research purposes.

TABLE 3. Normal form game matrix for the clinical investigators or doctors vs patients.

∨Patients|Doctors>	P1	P2
P1	++	+-
P2	-+	--

From the Table 3, P1 would denote the action of engaging in a trial whereas P2 will denote the action of not engaging in a trial. Normal game form between Patients, Doctors and Patients, Research Firms as their decision structure aligns, only if the game is in cooperative manner, for the Doctors and the Research Firms. Each of the payoff representation in the matrix would be interpreted as:

1) ++: Both Doctors and Patients are engaging in recruitment, resulting in great recruitment process, which would make the outcome desirable for both.
2) -+: Doctor engages but Patient doesn't. Recruitment rate would be low, affecting the Doctor's work and Patient will have net zero effect on its payoff.
3) +-: Doctors won't engage in the trial while Patient chooses to participate or engage in the trial, resulting in not selecting the Patient for a certain trial. Although, if Patient wants, they can participate in some other experiment trial, which may have high risk but high reward.
4) --: Both outcomes are unfavourable, which would hamper the clinical trial process.

To increase the payoff for doctors where patients are not engaging, one can use the naïve approach as well, which would include giving incentives or maybe some information for assurance of the trials, adhering to the risks associated with their health or their overall well-being. In a more intricate style, doctors can use methods like patient matching, personalized treatment recommendations, risk assessments, patient safety, informed consent, feedback analysis and patient monitoring to mitigate or level up the payoff, if patients are not engaging.

For the low payoff for patients, where doctors are not cooperative, one can use the targeted patient identification, predictive modelling, personalized patient outreach and real time monitoring. One such proposed model is mentioned in our paper as well, incorporating autoencoders and ensemble methods.

TABLE 4. Normal form game matrix for the doctors vs research firm.

∨Doctors|Research Firm>	Recruit	Fails to Recruit
Recommend	++	+-
Withold	-+	--

From the Table 4, 4 actions can be taken. Doctor can either recommend or withhold the recruitment. Whereas the research firm can either recruit or fail to recruit the apt patients. Each of the payoff representation in the matrix would be interpreted as:

1) *Doctor Recommends, Research Firm Recruits (Doctor: +, RF: +):* Doctor recommends the clinical trial to the patient and the Research Firm successfully recruits the patient.
2) *Doctor Recommends, Research Firm Fails to Recruit (Doctor: +, RF: -): Doctor* recommends the clinical trial to the patient and Research Firm fails to recruit the patient.
3) *Doctor Withholds, Research Firm Recruits (Doctor: -, RF: +):* Doctor withholds the recommendation for the clinical trial and Research Firm successfully recruits the patient.
4) *Doctor Withholds, Research Firm Fails to Recruit (Doctor: -, RF: -):* Doctor withholds the recommendation for the clinical trial and Research Firm fails to recruit the patient.

Clinical investigators or the doctors can maximize their payoff in a naïve approach, which may include financial compensation, career advancement, validation of trial outcomes and the quality of collaboration between the players. Whereas. Research firms can increase their payoff by effectively carrying out decisions related to trials. Research firm's main objective is to conduct a successful trial, which can be done by validating the trial outcomes, choosing quality data, checking cost-effective solutions and methods and also by maintaining a good relationship with the doctors.

By using machine learning methods, doctors can increase their payoff due to intervention of tools. Doctors can use these ML tools for identifying apt patients, engage in collaborative decision making, clinical support, remote collaboration and research facilitation. For Research firms, they can use ML tools to maneuver into drug discovery and development,

to optimize drug composition and application possibilities, and safety profiles too. Such tools can also optimize supply chain [41].

2) BACKWARD INDUCTIVE APPROACH

It is advantageous to know how many times a repeating game technique will be used until it achieves a Nash equilibrium, or to put it another way, an ideal solution or strategy. Backward propagation will only slow down the process if it is uncertain. However, even if there may be some uncertainty in the case of clinical trials, we may still estimate the number of occurrences. In order to employ the backward inductive approach, we would therefore assume that the organization is aware of the criteria so that they can successfully conduct the experiments. By taking into account research firm efforts, doctors can boost their compensation. Research firms boost their profits by conducting better studies, for which they must find the best patients. Figure 3 shows the influential model of the game proposed. Patients who choose to participate in trials will present conditions (or weights) to the study company. Doctors receive involvement through a working channel between clients and businesses in order to maximize their own financial gain through superior partnership. The firm's objective is to maximize profits with the least amount of work while rigorously following to medical regulations.

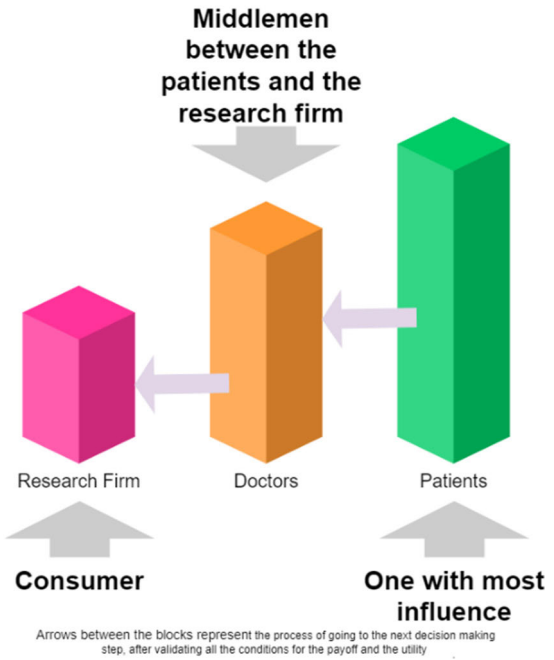


FIGURE 3. Interaction flow in the game between patients, doctors and research firms.

3) SAMPLE GAME INTERACTIONS

a: PATIENT-FIRM

- *Patient Awareness:* Firms aim to inform potential patients about clinical trials, benefits, risks and the eligibility criteria for the same [42].
- *Patient Screening:* Firms evaluate interested patients based on their medical history, eligibility criteria and suitability for the trial [43].
- *Informed Consent:* If the patient meets up with the eligibility criteria, firms present a consent form explaining everything. Patient can ask queries and decide whether to participate [44].
- *Patient Enrolment:* After consent, patients get enrolled. Firm then may collect information about the patients, perform baseline assessments and allocate patients to different treatment groups [45].

b: FIRM-DOCTOR

- *Study Design:* Firms and doctors collaborate to design trials. They select the effective endpoints, treatment protocols and recruitment strategies [46]. They also validate safety procedures and feasibility of the trials [47].
- *Recruitment Planning:* Firms and doctors discuss and plan recruitment strategies, considering factors such as patient demographics, geographic distribution, referral networks and competing trials. Firms provide doctors with recruitment materials, updates on enrolment status and guidance on engaging with potential participants [48].
- *Patient Referral:* Doctors refer eligible patients from their clinical practice as well [49].
- *Study Implementation:* Doctors collaborate with firms to ensure proper execution of study protocols, monitor patient progress, address events and collect data as per trial's requirements [50].

c: PATIENT-DOCTOR

- *Doctor's Recommendation:* Doctors interact with patients, providing insights into trial's response and purpose, potential benefits, risks and alternative treatments. Patients ask questions too.
- *Patient Decision Making:* Shared decision making is implemented to decide whether patient has to participate in the clinical trial or pursue any other options, acting upon doctor's recommendation by personal factors like health condition, preferences and logistical preferences.
- *Trial Participation:* If patients get enrolled, they collaborate with the doctors by attending regular check-ups, undergoing treatments, providing feedback etc.
- *Follow-up care:* Doctors provide post-trial care and monitor patient's health, ensuring smooth transitions back to regular clinical management [51], [52].

4) UTILITY FUNCTION AND PAYOFF MAXIMIZATION SUB-GAMES

a: UTILITY FUNCTIONS

Here, we build the groundwork by explaining the utilitarian roles of important participants, including patients, physicians, and research organizations. These functions are essential for measuring individual preferences, assessing decision-making

procedures, and eventually providing a thorough framework for examining interactions in the context of our study.

Patient Utility: The utility function of the patients would be defined as,

$$U_p = P_{gain} - P_{cost} = \sum_{d=1}^{D} C_d B_d - P_d \tag{2}$$

Here, P_{gain} represents the perceived benefits gained from participating in clinical trials by the patients, P_{cost} stands for the potential personal costs involved for the patients, C_d represents the specific benefits associated with trial aspect, from each doctor d, B_d signifies the intrinsic appeal of each benefit given to the patient, from the good degree of care by the doctors, and P_d represents the individual's own personal costs, raised due to hesitations or concerns, which were not addressed properly by the doctors, equivalent to P_{cost} and $\sum_{d=1}^{d} C_d B_d$ equivalent to P_{gain}.

This utility function encapsulates the patients' decision-making process by quantifying the balance between perceived benefits and concerns, allowing us to analyze their recruitment strategies in the context of doctors and research firms.

Doctor Utility: The utility function of the doctors would be defined as,

$$U_d = \sum_{a=1}^{Ad} K_{d,a} G_{d,a} - \sum_{p=1}^{P} C_p B_p \tag{3}$$

where U_d represents the doctor's utility, $K_{d,a}$ denotes the significance of engaging patients through various strategies by each doctor d and resources used by the doctor, with each usage of resource a, $G_{d,a}$ signifies the success achieved with each engagement strategy in a trial by each doctor d, using resource a, and C_p $^*B_p$ captures benefits and gains achieved by the doctors due to effective addressing of concerns or reservations, of each patient p.

Doctor's utility evaluates a doctor's utility by considering the success of patient engagement strategies, the resources used, and the effectiveness in addressing patient concerns. It provides a measure of how valuable a doctor's efforts are in clinical trial recruitment.

Research Firm Utility: The utility function of the research firms would be defined as,

$$U_r = G_{d,a} \quad s.t.\ \delta_a \geq \sum_{d=1}^{D} K_{d,a} G_{d,a} \tag{4}$$

where U_r represents the research firm's utility, $G_{d,a}$ signifies the benefits derived from successful patient recruitment using some resource a by some doctor d, δ_a reflects the research firm's efficacy in achieving recruitment goals using some resource a, $K_{d,a}$ $^*G_{d,a}$ represents the costs associated with patient engagement strategies, and $\sum_{d=1}^{D}$ encompasses all potential engagement strategies.

Research firm's computes the utility of a research firm by considering the benefits of successful patient recruitment, the costs of engagement strategies, and ensuring that the benefits surpass the total costs. It offers an assessment of the firm's effectiveness in patient recruitment.

b: PAYOFF MAXIMIZATION SUB-GAMES

Research Firm:

The clinical trial recruitment enhancement subgame for research firms, as outlined in the paper, aims to discover the optimal approach for research firms to maximize their utility function while collaborating with patients, doctors, and other research firms. The optimization challenge for augmenting the recruitment process is represented by the following

$$maximize_{\zeta_{d,a}} G_{d,a} = \alpha_{d,a} \beta_{d,a} \ln(1 + \zeta_{d,a}) \tag{5}$$

subjected to these constraints,

- $\sum_{d=1}^{D} K_{d,a} G_{d,a} \leq \delta_a$
- $\sum_{i=1}^{Ad} \alpha_{d,a} \beta_{d,a} \leq A$
- $\zeta_{d,a} \geq \zeta_{d,a}; \forall a$

where,

- $\zeta_{d,a}$ denotes the efficiency coefficient of the recruitment method employed by research firm with available resources a and doctor d.
- $\alpha_{d,a}$ represents the research firm's overall strategic effectiveness or capability in devising and implementing patient recruitment strategies. It encapsulates the firm's expertise, resources, and approach towards patient engagement and recruitment. Higher the value of $\alpha_{d,a}$ could indicate a more successful and resourceful research firm in terms of recruitment strategies.
- $\beta_{d,a}$ reflects the specific impact of the chosen recruitment strategy on each individual patient's willingness and likelihood to participate in the clinical trial. It takes into account factors such as the patient's preferences, concerns, medical history, and the alignment of the recruitment approach with their needs. A higher $\beta_{d,a}$ value suggests that the chosen strategy is more compelling and relevant to that particular patient.
- $\ln(1 + \zeta_{d,a})$ signifies that $G_{d,a}$ from the 5^{th} equation, increases more slowly when $\zeta_{d,a}$ increases.

The objective of the optimization problem is to enhance the recruitment process and maximize the utility. Constraints encompass factors such as the total resources allocated to patients, the overall capacity of research firms, and the prescribed limits for the recruitment approach. These constraints guarantee a feasible optimization and ensure that resources are not overspent during the recruitment process.

1) *First constraint,*

The constraint essentially states that the total quantity of components for payoff ($K_{d,a}$, $G_{d,a}$) needed to create all the patient recruitment efforts should not exceed the allocated efforts of the research firm(δ_a) for the trial. The research firm's commitment to engage with patients ($K_{d,a}$) for crafting patient recruitment process, should contribute to the overall benefits, while staying within the constraints of the available time or efforts (δ_a) for the research firm's operations.

2) *Second constraint,*

The constraint essentially states that the sum of the effectiveness ($\alpha_{d,a}$) of designing individual recruitment strategies for patients, while considering their preferences ($\beta_{d,a}$),

should not exceed the available resources (A) of the research firm. The constraint helps illustrate how the constraint aligns with the concepts of architectural design and customization, as well as the research firm's strategic effectiveness, all while staying within the boundaries of their available resources.

3) *Third constraint,*

$\bar{\zeta}_{d,a}$ (Minimum Efficiency) represents the minimum level of efficiency that the research firm expects for each participation planned for each doctor, allotted with resources a. It's like setting a baseline standard for how well each patients' recruitment should be organized. The constraint essentially states that the efficiency of the $\zeta_{d,a}$ for each patient should not fall below the established minimum efficiency level ($\bar{\zeta}_{d,a}$). The constraint aligns with the idea of setting a certain benchmark for the efficiency of each personalized patient recruitment strategy. The research firm aims to maintain a consistent level of effectiveness in their recruitment methods for each patient.

Furthermore, $K_{d,a}$ can be represented as:

$$\frac{\hat{K}_{d,a}(\zeta_{d,a})}{\alpha_{d,a}\beta_{d,a}\ln\left(1+\zeta_{d,a}\right)} \tag{6}$$

Equation (5) can be rewritten as,

$$maximize_{\zeta_{d,a}} G_{d,a} = \alpha_{d,a}\beta_{d,a}\ln\left(1+\zeta_{d,a}\right) \tag{7}$$

subjected to constraint,

- $\zeta_{d,a} \geq \bar{\zeta}_{d,a}$; $\forall p, \forall a$ *where a denotes resources utilized*
- $\sum_{d=1}^{D} \hat{K}_{d,a}\zeta_{d,a} \leq \delta_a$

Equation (7) is convex in $\zeta_{d,a}$ [53] and it has a single optimal solution. Therefore, optimal $\zeta_{d,a}$'s payoff would be maximized and would satisfy the third constraint of equation (5).

The optimal $\zeta_{d,a}$ which will satisfy $\zeta_{d,a} > \bar{\zeta}_{d,a}$ for any resource a, for each doctor d will be given as,

$$\zeta_{d,a} = \frac{\sum_{d=1}^{D}\hat{K}_{d,a}+\delta_a}{D\hat{K}_{d,a}} - 1 \tag{8}$$

where D can be some maximum quantity of doctors participating, and when equality is satisfied of the third constraint of equation (5), we can rewrite equation (8) as,

$$\zeta_{d,a} = \max(\bar{\zeta}_{d,a}, (\frac{\sum_{d=1}^{D}\hat{K}_{d,a}+\delta_a}{D\hat{K}_{d,a}} - 1) \tag{9}$$

Upon seeing the equations, we can also propose that to satisfy the third constraint of equation (5), as

$$\delta_a + \sum_{i=1,i\neq d}^{D} \bar{K}_{i,a} \geq ((\bar{\zeta}_{d,a}+1)D-1)\hat{K}_{d,a} \quad \forall a, \forall p \tag{10}$$

Based on research firm's optimal strategy, we can move backward to Doctors' subgame, which would maximize its own payoff under the influence of various actions.

Doctors:

Payoff Maximization Sub-game can be defined as (D, $\{S_D\}$,$U_d(.)$) where,

1) D are the set of active doctors or Clinical Investigators
2) S_D are the set of strategies of each doctor
3) U_d:$\{S_1$ X $S_2 \ldots$ X $S_C\}$ is the payoff

Payoff/utility maximization sub-game for the doctor can be written as,

$$maximize_{K_{d,a}C_d} U_d = \sum_{a=1}^{Ad} K_{d,a}G_{d,a} - \sum_{p=1}^{P} C_pB_p \tag{11}$$

Subjected to:

- $\alpha_{d,a}\beta_{d,a} \leq \varepsilon_d A$
- $\sum_{a=1}^{Ad} G_{d,a} \leq \hat{G}_{d,a}$
- $K_{d,a} > 0$

Each constraint tells a different analogy.

1) *In the first constraint,* ε_d represents a factor that signifies the research firm's capacity to design and implement various patient recruitment strategies within the limitations of their allocated resources, by each doctor. A higher value of ε_d indicates a larger capacity to allocate resources for designing and implementing effective patient recruitment strategies, while a lower value implies a more constrained resource allocation. The first constraint of equation (11), reflects that the product of the research firm's strategic effectiveness ($\alpha_{d,a}$) and the specific impact of the recruitment strategy ($\beta_{d,a}$) for each patient should not exceed the firm's available resources (A), considering the capacity indicated by the coefficient ε_d. ε_d signifies the research firm's resource allocation capacity for designing tailored patient recruitment strategies.
2) *In the second constraint*, the constraint essentially states that the total success achieved across all engagement strategies should not exceed the predefined maximum success limit (unit $G_{d,a}$). The constraint aligns with the idea of maintaining a balanced and impactful collection of engagement strategies. The research firm aims to ensure that the overall success achieved through various engagement strategies stays below a predefined limit while effectively engaging patients in the context of clinical trial recruitment.
3) *In the third constraint,* the constraint essentially states that the significance assigned to each engagement strategy should be greater than zero, meaning that every strategy holds importance and contributes positively to the overall patient recruitment effort. The research firm assigns significance to each strategy with the expectation that they all contribute to the recruitment process.

Inspiring from the first two constraints of equation (11), we can say that

$$\sum_{\forall a} \zeta_{d,a} \leq S_D \tag{12}$$

where S_D is estimated to be $S_D = \frac{e^{G_d/(\alpha_{d,a}\beta_{d,a})}}{A_d}$ and optimization problem will be given as,

$$maximize_{K_{d,a}C_d} U_d = \sum_{a=1}^{Ad} \hat{K}_{d,a}\zeta_{d,a} - \sum_{p=1}^{P} C_pB_p \tag{13}$$

Subjected to:

- $\sum_{a=1}^{Ad} \zeta_{d,a} \leq S_D$
- $\hat{K}_{d,a} > 0$

Here, this maximization sub-game of doctors is a non-cooperative game where doctors would be interested only in maximizing its utility. Each doctor will start with a unit significant effort to each engagement strategy, and announces it to the research firm. Each doctor has some maximum numbers of hours allotted to them. Every doctor updates their unit effort based on their time and demand from the research firm as well, in the form of $\zeta_{d,a}$. Its given as,

$$\hat{K}_{d,a}(t+1) = \hat{K}_{d,a}(t) + \sigma_D\left(\sum_{\forall d}\zeta_{d,a} - S_D\right) \quad (14)$$

where $\sigma_D \ll 1$, where σ_D is the adjustment parameter introduced to fine-tune the changes in effort based on the difference between the research firm's perceived significance ($\zeta_{d,a}$) and the doctor's perception (S_D) for a specific engagement strategy. It is intentionally set to a small value ($\ll 1$) to ensure that the adjustments remain subtle and incremental, thus avoiding abrupt changes in effort allocation. σ_D is estimated as $\sigma_D = \frac{\sum_{\forall p} C_d}{2\sum_{\forall d,a} K_{d,a}}$. This process is repeated until it converses. In order to explore the presence of an equilibrium in the game and to discern the optimal response strategies employed by the players, we outline the subsequent precise definitions derived from game theory within the framework of our specific problem.

The set of all significance noted $\{\hat{K}_{1,a}, \hat{K}_{2,a}, \hat{K}_{3,a} \ldots \hat{K}_{D,a}\}$ is going to be equilibrium for the Doctor's Payoff Maximization Sub-game for each doctor, if we have,

$$U'_d\left(\hat{K}'_{d,a}, \hat{K}_{-d,a}\right) \geq U_d\left(\hat{K}_{d,a}, \hat{K}_{-d,a}\right), \forall \hat{K}'_{d,a} \epsilon S_d \quad (15)$$

where $\hat{K}_{-d,a} = \{\hat{K}_{1,a}, \ldots K_{d-1,a}, K_{d+1,a} \ldots K_{D,a}\}$ is the significance to be adhered by doctors. At the point of equilibrium, none of the parties – doctors, research firms, or patients – have any motivation to alter their strategies independently as it would not result in any advantageous outcome for them. At the equilibrium point, neither doctors, research firms nor the patients receive positive incentives by changing their strategies unilaterally. The doctor's payoff, as described in equation (13), is an increasing function with respect to $\hat{K}_{d,a}$, ensuring the presence of a single optimal solution.

Optimized unit $K_{d,a}$ which will maximize the payoff of each doctor will be updated as,

$$\hat{K}_{d,a} = \frac{\sum_{i=1,i!=d}^{D}\hat{K}_{I,a} + \sum_{a=1}^{A,d}\delta_a}{(A(D-1) + DS_D} \quad (16)$$

where A are the total allocated resources. The optimized $K_{d,a,}$ will be then given as,

$$K_{d,a} = \hat{K}_{d,a}\frac{\zeta_{d,a}}{\alpha_{d,a}\beta_{d,a}\ln(1+\zeta_{d,a})} \quad (17)$$

Next, based on Doctor's utility/payoff subgame, we can determine how we can determine the optimal parameters for patients to maximize their utilities/payoff.

Patients:

Main objective of the Patient Maximization Sub-game is to maximize the benefits or chance of getting recruited into the clinical trials and setting an optimal contract with the doctors and the research firm. Payoff optimization problem can be expressed as,

$$maximize_{C_d B_d} U_p = \sum_{d=1}^{D} C_d B_d - P_d \quad (18)$$

Subjecting to:

- $\alpha_{d,a}\beta_{d,a} \leq \varepsilon_d A$
- $\sum_{\forall D}\alpha_{d,a}\beta_{d,a} \leq A$
- $\sum_{a=1}^{A,d} G_{d,a} \leq \hat{G_{d,a}}$
- $C_d > 0$

For the above constraints, first, second and the third are already mentioned from the constraints of equation (11), (5). The fourth constraint, reflects the idea that patients are motivated by positive benefits, and trial aspects must provide meaningful, positive advantages to be considered in their decision to participate in clinical trials.

For patients, taking benefits for a certain trial, from d doctors, would be,

$$C_d(t+1) = C_d(t) + \rho(\sum_{\forall d} C_d - P_d) \quad (19)$$

where "ρ" in this equation represents a parameter or a constant that determines the rate at which the patients' benefits (C_d) for a certain trial increase based on their interactions with doctors (P_d). C_d which can maximize the payoff for any patient, can be

$$C_d = \frac{\sum_{\forall d}\sum_{\forall a} K_{d,a} + \sum_{j=1,j\neq p}^{P} C_p}{P\hat{G}_{d,a} + D(P-1)} \quad (20)$$

c: BEST RESPONSE FOR THE THREE STAGE GAME AND UNIQUE EQUILIBRIUM

From the equation (20) and (16), we can say that it maximizes the payoffs of each player in their respective subgames, and then it is also the best responses of the game.

Proof: Let's imagine that there is an upward adjustment in the value of $\hat{K}_{d,a}$ by $\varepsilon_d > 0$, resulting in a revised substantial unit,

$$\hat{K}'_{d,a} = \hat{K}_{d,a} + \varepsilon_d \quad (21)$$

For research firm's subgame for maximizing the payoff, we can determine the new $\zeta'_{d,a}$ from equation (8) by putting $\hat{K}_{d,a}$with $\hat{K}'_{d,a}$,

$$\zeta'_{d,a} = \frac{\sum_{j=1,j\neq d}^{D} K_{j,a} + \hat{K}'_{d,a} + \delta_a}{D\hat{K}'_{d,a}} - 1 \quad (22)$$

Finding the difference between $\zeta_{d,a}$ and $\zeta'_{d,a}$ will be formulated as,

$$\zeta_{d,a} - \zeta'_{d,a} = \frac{\hat{K}'_{d,a} - \hat{K}_{d,a}}{\hat{K}_{d,a}\hat{K}'_{d,a}}\left(\frac{\sum_{j=1,j\neq d}^{D} K_{j,a} + \delta_a}{D}\right) \quad (23)$$

Equation (23) says that $\zeta_{d,a} - \zeta'_{d,a} > 0$ since $\hat{K}'_{d,a} > \hat{K}_{d,a}$ for $\varepsilon_d > 0$. This indicates that raising the significant unit does not lead to an increase in the players' payoff in the research firm's subgame.

In the next step, when focusing on the Doctor's subgame for maximizing their payoff, we'll examine the transition from the initial significance level, denoted as $\hat{K}_{d,a}$ to the updated significance level, denoted as $\hat{K}'_{d,a}$ for doctors. Due to increase in significance for given C_d and B_d, the payoff difference is,

$$U'_d\left(\hat{K}'_{d,a}, \hat{K}_{-d,a}\right) - U_d\left(\hat{K}_{d,a}, \hat{K}_{-d,a}\right) = \hat{K}'_{d,a}\sum_{\forall a}\zeta'_{d,a} - \hat{K}_{d,a}\sum_{\forall a}\zeta_{d,a} \quad (24)$$

Substituting from equations from (8), (22) and (16) into (24), we can formulate,

$$U'_d\left(\hat{K}'_{d,a}, \hat{K}_{-d,a}\right) - U_d\left(\hat{K}_{d,a}, \hat{K}_{-d,a}\right) = -\frac{D-1}{D}\varepsilon_d \quad (25)$$

From equation (25), it can be inferred that $U'_d\left(\hat{K}'_{d,a}, \hat{K}_{-d,a}\right) < U_d\left(\hat{K}_{d,a}, \hat{K}_{-d,a}\right)$ for $\varepsilon_d > 0$, which is contradicting equation (15), which is setting up the equilibrium point for the Doctor's subgame. Also, we can observe a decline in significance value, which will decrease the payoff due to the effect of keeping $\varepsilon_d > 0$.

In the same way, we can see the trend by keeping $K'_{d,a} = K_{d,a} + \varepsilon_d$ for patient's payoff, we can formulate the difference between U'_p for and U_p for, using equation (20). After some algebraic calculations, difference computed is,

$$U'_p - U_p \propto \varepsilon_d \zeta'_{d,a} c_d (\zeta'_{d,a} - \zeta_{d,a}) \quad (26)$$

where we can say that $U'_p - U_p < 0$ since $\varepsilon_d > 0$, $\zeta'_{d,a} > 0$, $C_d > 0$ and $\zeta'_{d,a} - \zeta_{d,a}$ is negative, from seeing equation (23). Similarly, we can deduce a comparable adjustment in C_d without resulting in an increased payoff for patients, doctors, and research firms. Due to space constraints, we haven't included the analysis for the change in C_d. Consequently, we can affirm that the three-way influence-centric game has an equilibrium point. Therefore, equation (16) represents the optimal response within the game. Furthermore, an equilibrium exists for patients and doctors in the significance selection game under the following conditions: a) The negotiation sets are non-empty, convex, and bounded subsets of Euclidean space, i.e., they should have finite space and b) Utilities are continuous and exhibit concavity (for both U_p and U_d).

Also, $\frac{\partial^2 U_p}{\partial (C_d)2} = 0, \forall p, and\ \frac{\partial^2 U_d}{\partial \hat{K}^2_{d,a}} = 0, \forall d$. Hence, payoff or the utility functions for Doctors and Patients are concave in significance, hereby stating that a unique equilibrium does exist.

Table 5 will furnish a concise exposition elucidating the purpose and role of each variable employed in formulating the game theoretic model for reference.

TABLE 5. Variables mentioned in the game theory methodology elucidated.

Symbol/ Character Assigned	Description
$U_{(any\ class\ of\ player)}$	Utility Function of a particular class of player, distinguished by a subscript, assigned for each class of players
P_{gain}	The perceived benefits gained from participating in clinical trials by the patients
P_{cost}	The potential personal costs involved for the patients
$C_{any\ class\ of\ player}$	The specific benefits associated with trial aspect
$B_{any\ class\ of\ player}$	The intrinsic appeal of each benefit given to the patient, from the good degree of care by the doctors
$P_{any\ class\ of\ players}$	The individual's own personal costs, raised due to hesitations or concerns, which were not addressed properly by the doctors
D, P, A, R	Each representing total count of all individuals in player class- doctors, patients, resources and research firms respectively
d,p,a,r	Each represented as a particular unit variable for classes of doctors, patients, resources and research firms respectively
K	The significance of engaging patients through various strategies
G	The success achieved with each engagement strategy in a trial
δ	The research firm's efficacy in achieving recruitment goals
ζ	The efficiency coefficient of the recruitment method employed by research firm
α	The research firm's overall strategic effectiveness or capability in devising and implementing patient recruitment strategies. It encapsulates the firm's expertise, resources, and approach towards patient engagement and recruitment
β	The specific impact of the chosen recruitment strategy on each individual patient's willingness and likelihood to participate in the clinical trial. It takes into account factors such as the patient's preferences, concerns, medical history, and the alignment of the recruitment approach with their needs
^ symbol over a variable	Defines a unit variable
¯Symbol over any variable	Defining some threshold for that variable
$S_{any\ class\ of\ players}$	Set of strategies set up for some particular class of player
ε	a factor that signifies the research firm's capacity to design and implement various patient recruitment strategies within the limitations of their allocated resources
σ	The adjustment parameter introduced to fine-tune the changes in effort based on the difference between the research firm's perceived significance ($\zeta_{d,a}$) and the doctor's perception (S_D) for a specific engagement strategy
'Symbol over any variable	Some updated variable of the same type, signifying a change
ρ	A parameter or a constant that determines the rate at which the patients' benefits for a certain trial increase based on their interactions with doctors

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL SETUP

Our empirical analysis of the proposed model, incorporating ensemble stacking methods and autoencoders, was conducted using PYTHON in a Windows 10 environment, supported by an 8GB RAM system with a Ryzen 5 5000 series CPU. The experiments were executed with Python version 3.10.

B. DATASET DESCRIPTION

The dataset utilized in this study originates from Electronic Health Records Prediction, obtained from a private hospital located in Indonesia [54]. The dataset comprises laboratory test results of patients, which play a pivotal role in guiding subsequent treatment decisions, whether involving patients receiving care within the hospital premises or those under outpatient care. The primary objective associated with the dataset involves classification prediction tasks. This real-world dataset has been out-sourced from Kaggle, a well-established platform for data science and analysis. Discrete variables are there as well, listed as follows-

- AGE denotes the patient's age.
- The SEX attribute is a binary nominal variable indicating the patient's gender.
- The SOURCE attribute is a nominal feature with two categories: ''in' represents patients receiving in-care, while 'out' indicates patients receiving out-care.

The 'SOURCE' attribute serves as the target column (i.e., 'in' or ''out') for classification tasks, making it a vital component for predictive modeling and healthcare analysis. The continuous variables in this dataset, namely HAEMATOCRIT, HAEMOGLOBINS, ERYTHROCYTE, LEUCOCYTE, THROMBOCYTE, MCH, MCHC, and MCV, represent essential blood parameters that play a crucial role in assessing a patient's overall health and suitability for participation in clinical trials. Here's a brief description for all the blood parameters considered for our experiment-

- *HAEMATOCRIT:* Haematocrit measures the volume of red blood cells in the blood. It helps evaluate a patient's blood composition and oxygen-carrying capacity.

  Example- In a clinical trial for patients with anemia, a minimum HAEMATOCRIT level may be required to ensure that participants have a sufficient number of red blood cells to assess the effectiveness of an anaemia treatment.
- *HAEMOGLOBINS:* Haemoglobins are responsible for transporting oxygen in the blood. The levels of haemoglobins can provide insights into a patient's oxygen-carrying capacity and overall health.

  Example- For a clinical trial studying the effects of a new blood transfusion therapy, patients with low HAEMOGLOBINS may be considered as they could benefit the most from the treatment.
- *ERYTHROCYTE:* Erythrocyte count refers to the number of red blood cells. It is vital for assessing a patient's ability to transport oxygen efficiently. In clinical trials, ERYTHROCYTE levels are important for determining a patient's suitability for studies focused on oxygen transport and blood health.

  Example- In a clinical trial investigating the effects of a drug on oxygen transport, participants with low ERYTHROCYTE counts may be chosen to assess the drug's potential to improve oxygen-carrying capacity.
- *LEUCOCYTE:* Leucocytes, or white blood cells, are crucial for the immune system's function. Their count can indicate a patient's immune response and overall health. In clinical trials, LEUCOCYTE levels play a role in assessing a patient's immune system health and response to treatments.

  Example- In an immunotherapy trial for cancer treatment, patients with normal or high LEUCOCYTE counts may be preferred to ensure that their immune systems can effectively respond to the treatment.
- *THROMBOCYTE:* Thrombocytes, or platelets, are essential for blood clotting. Their count is significant in evaluating a patient's ability to form blood clots. In clinical trials, THROMBOCYTE levels are important for assessing the suitability of patients for studies involving clotting disorders or treatments.

  Example - In a study on patients with bleeding disorders, individuals with low THROMBOCYTE counts may be selected to evaluate the efficacy of a new drug in improving blood clotting.
- *MCH, MCHC, and MCV:* These parameters are related to the size and hemoglobin content of red blood cells. They provide information about the quality and characteristics of red blood cells.

  Example- In clinical trials, MCH, MCHC, and MCV levels can be critical for assessing a patient's blood health and their eligibility for studies involving red blood cell disorders.

In the dataset, patient_id serves as a unique identifier for each of the 4412 patients. Notably, HAEMATOCRIT, representing haematocrit levels, ranges from 13.7 to 69.0, showing substantial variation. HAEMOGLOBINS, which indicate haemoglobin concentration, span from 3.8 to 18.9, reflecting diversity among patients. ERYTHROCYTE, measuring red blood cell count, varies between 1.48 and 7.86. LEUCOCYTE, denoting white blood cell count, ranges from 1.1 to 76.6. THROMBOCYTE, representing platelet count, extends from 8.0 to 1183.0, showing significant differences. MCH (Mean Corpuscular Haemoglobin) varies between 14.9 and 40.8, reflecting diversity in haemoglobin content within red blood cells. MCHC (Mean Corpuscular Haemoglobin Concentration) ranges from 26.0 to 39.0, showcasing variations in haemoglobin concentration. MCV (Mean Corpuscular Volume) varies from 54.0 to 115.6, indicating differences in red blood cell volume. Finally, AGE spans from 1.0 to 99.0, highlighting the age diversity within the patient population. These variations provide valuable insights for clinical trial patient recruitment and further analysis. Within the context of optimizing patient recruitment for advanced clinical trials, the dataset's parameters align with the eligibility criteria for participation. This includes discerning the eligibility of patients, both in terms of in-patient and out-patient care, for enrolment in advanced clinical trials.

Resampling methods, feature engineering and feature scaling were done as the part of the preprocessing. Dataset was

divided into 2 parts: 0.8 part of the dataset for the training data and the rest for the testing data.

C. AUTOENCODERS PERFORMANCE

Upon analyzing the encoded data, our autoencoder-based approach has impressively captured essential features and patterns within the patient dataset. The encoded data brings to light several crucial insights for optimizing patient selection. The shape of the encoded data, which is (4204, 700), indicates that while there wasn't an extensive reduction in dimensionality, it's essential to note that this subtle dimensionality reduction has uncovered intricate details and patterns among the blood parameters and other discrete variables within the dataset. Despite retaining a considerable number of features, the encoded data provides a clearer view of underlying relationships and interactions, contributing to a more refined understanding of patient characteristics and eligibility for clinical trials.

To quantify the accuracy of the autoencoder's data reconstruction, we employed the Mean Squared Error (MSE) metric. For each data point in our designated dataset, we obtained the reconstructed counterpart using the autoencoder model. The MSE was then computed by averaging the squared differences between the original data and its reconstructed form, revealing an average MSE value of 0.00022996273820050333 for the autoencoder's reconstructions on the test dataset. This MSE value provides a quantitative measure of the reconstruction accuracy achieved by the model. Low Mean Squared Error (MSE) during data reconstruction underscores the autoencoder's ability to maintain data integrity, ensuring the accurate identification of eligible patients without compromising clinical data quality. Figure 4 shows a graph of the restructured data points plotted against the range of MSE values.

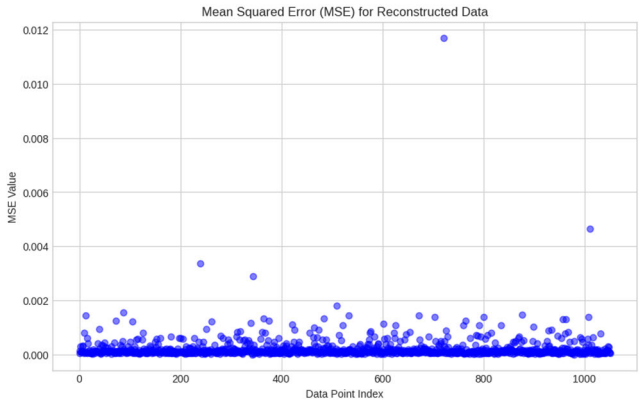


FIGURE 4. MSE for reconstructed data, made by the autoencoder.

D. CLASSIFICATION MODELS METRIC EVALUATION

In the proposed super-classification model, in the form of stacking classifier, 4 different classification models were trained with best hyperparameters, searched using RandomizedCVSearch. There are various performance metric tasks used in the domain of machine learning. For the classification, we have utilized accuracy, recall, precision, f1-score, support, macro average and weighted average.

TABLE 6. Comparison of ML classification on the basis of testing/validation accuracy and training accuracy.

	Testing/Validation Accuracy	Training Accuracy
DecisionTrees	70.81%	82.65%
Logistic Regression	71.29%	70.29%
SVM	73.85%	74.35%
Random Forest	82.88%	100%
Stacking Classifier	83.45%	99.95%

Table 7 shows the classification report of each model used, after the hyperparameter tuning. Table 6 shows training accuracy and validation accuracy of each classification used.

From Table 2, we can see the selection of hyperparameter values for the decision tree model used in the analysis of the encoded clinical trial dataset reflects a careful balance of various considerations. A minimum of 20 samples required for a split ensures that nodes have enough data for meaningful branching, preventing overfitting. A maximum of 256 leaf nodes controls the tree's complexity, keeping it manageable and interpretable. Using only 20% of features at each split introduces diversity, mitigating overfitting risks. The limitation of the tree depth to 12 strikes a balance between capturing patterns and model simplicity. The choice of the 'gini' criterion aligns with classification tasks, optimizing classification accuracy. The class weighting, with class 1 (eligible patients) assigned higher weight, emphasizes the importance of correctly identifying eligible patients. These settings are tailored to the clinical trial recruitment context, emphasizing the significance of accuracy, diversity, and model simplicity in identifying eligible patients while maintaining interpretability. The Decision Tree model, with its hierarchical and tree-like structure, excels at dividing the dataset into subsets based on the most informative features, providing an intuitive representation of decision-making. It identifies eligibility for clinical trials by considering key features, achieving a good balance between correctly identifying suitable patients and avoiding incorrect identifications.

The hyperparameter settings for our logistic regression model reflect a balanced approach, considering key factors for effective clinical trial patient recruitment. We employ an 'l2' penalty for regularization, striking a balance between model accuracy and simplicity. With 'max_iter' set to 100, we ensure efficient model training without excessive computational overhead. Enabling ''fit_intercept' aligns our model

TABLE 7. Classification report of each model, stating each classification model's accuracy, macro average, weighted average, precision, recall, support and F1-score.

		Precision	Recall	F1-Score	Support
Decision Tree	0	0.79	0.56	0.66	526
	1	0.66	0.85	0.75	526
	Accuracy			0.71	1052
	Macro Avg	0.73	0.71	0.7	1052
	Weighted Avg	0.73	0.71	0.7	1052
Logistic Regression	0	0.71	0.72	0.72	526
	1	0.72	0.7	0.71	526
	Accuracy			0.71	1052
	Macro Avg	0.71	0.71	0.71	1052
	Weighted Avg	0.71	0.71	0.71	1052
SVM	0	0.72	0.77	0.75	526
	1	0.75	0.71	0.73	526
	Accuracy			0.74	1052
	Macro Avg	0.74	0.74	0.74	1052
	Weighted Avg	0.74	0.74	0.74	1052
Random Forests	0	0.82	0.84	0.83	526
	1	0.83	0.82	0.83	526
	Accuracy			0.83	1052
	Macro Avg	0.83	0.83	0.83	1052
	Weighted Avg	0.83	0.83	0.83	1052
Stacking Classifier	0	0.82	0.85	0.83	526
	1	0.84	0.81	0.83	526
	Accuracy			0.83	1052
	Macro Avg	0.83	0.83	0.83	1052
	Weighted Avg	0.83	0.83	0.83	1052

with traditional logistic regression. Equal class weights (1:1) signify a fair treatment of both class 0 and class 1. The regularization strength 'C' is set to 1, emphasizing a balanced approach to prevent overfitting and maintain model flexibility. These hyperparameters collectively aim to achieve accuracy while keeping the model straightforward, crucial for the accurate identification of eligible patients in clinical trials. Logistic Regression, a linear classification method, demonstrates robust performance in identifying clinical trial eligibility. It leverages the linear combination of features to make predictions, resulting in balanced Precision, Recall, and F1-scores. While achieving commendable accuracy, it provides consistent results, making it a reliable choice for patient recruitment. However, it's noteworthy that Decision Trees performed slightly better, possibly due to their non-linear decision boundaries, which were better suited for capturing complex relationships in the data.

In our SVM (Support Vector Machine) model, the choice of hyperparameters is designed to balance complexity and accuracy for optimal clinical trial patient recruitment. We opt for the 'rbf' (Radial Basis Function) kernel, a versatile choice that can capture complex patterns in the data efficiently. A 'gamma' value of 1 emphasizes the significance of each data point, while a 'degree' of 4 indicates higher-order polynomial kernel usage. With a regularization parameter 'C' set to 100, we strike a balance between model complexity and accuracy, aiming to prevent overfitting and maintain the model's generalization capabilities. These hyperparameters collectively contribute to accurate patient selection for clinical trials while ensuring the model's robustness. SVM

meticulously identifies a boundary that maximizes the margin between eligible and ineligible patients. This margin represents the space that is least ambiguous for patient classification. SVM operates in a higher-dimensional space, where each feature of the patient dataset corresponds to a different dimension. SVM excels in finding non-linear decision boundaries, which is essential for capturing complex relationships between various blood parameters and patient attributes. In our analysis, SVM achieved an impressive test accuracy of 73.85%, indicating its proficiency in correctly categorizing patients. Its balanced Precision and Recall values signify that it effectively identifies both eligible and ineligible patients without compromising on quality. However, Decision Trees, with their adaptability to complex data patterns, slightly outperformed SVM in our specific clinical trial recruitment setting.

For our Random Forest model, we have thoughtfully selected hyperparameters to optimize patient recruitment for clinical trials. With 'n_estimators' set to 150, we consider a substantial number of decision trees to achieve robust predictions. 'min_samples_split' and 'min_samples_leaf' are both set to 2, promoting fine-grained data partitioning while preventing overfitting. 'max_features' is chosen as 'sqrt', which means it considers a square root of the total features at each split, offering a balance between diversity and correlation among decision trees. A 'max_depth' of 20 controls tree depth, enhancing model generalization. Additionally, we opt for 'bootstrap' as 'FALSE,' which allows us to build the forest using the entire dataset. These hyperparameters collectively enable accurate patient recruitment for clinical trials and a resilient model with minimized overfitting. Random Forest, a versatile ensemble learning technique, combines the power of multiple decision trees to make robust and accurate predictions. It's particularly useful in complex classification tasks, like predicting clinical trial eligibility. Random Forest excels at capturing intricate patterns in the data by aggregating multiple decision trees' results, thereby reducing the risk of overfitting. In our analysis, Random Forest achieved remarkable test accuracy of 82.88%. It maintains a fine balance between precision and recall, ensuring that both eligible and ineligible patients are correctly identified. This demonstrates its efficacy in our clinical trial recruitment context, where precision and recall are both critical for accurate patient selection. The model's excellent performance can be attributed to its ability to adapt to the intricacies of the patient dataset, showcasing its suitability for such applications.

As for the Stacking classifier, the Stacking Classifier strategically incorporates the strengths of each base model to enhance patient selection for clinical trials. Decision Trees help identify non-linear patterns in the dataset, making them useful for capturing complex relationships among blood parameters and patient attributes. Logistic Regression's efficiency is leveraged for straightforward classifications, ensuring swift decision-making. SVM excels in managing high-dimensional data and finding optimal decision boundaries, thus aiding in precise patient selection. The Random Forest's ability to handle noisy data and extract intricate patterns is crucial in maintaining data quality. By combining these models, the Stacking Classifier optimizes patient recruitment by leveraging their respective capabilities, resulting in a robust and accurate classification system tailored for clinical trial settings. The similarity between the Stacking Classifier and the Random Forest's results can be attributed to the significant influence of the Random Forest's strengths in the stacking ensemble. Random Forest excels in handling noisy data and capturing complex patterns in the clinical trial recruitment dataset, where intricate relationships among blood parameters and patient attributes are prevalent. The Stacking Classifier leverages this strength by assigning considerable importance to the Random Forest model within the ensemble. Consequently, it harnesses the Random Forest's ability to identify subtle patterns and maintain data quality, ensuring the precise selection of eligible patients for clinical trials.

Remarkably, the Stacking Classifier stands out as the top performer, achieving an impressive testing accuracy of 83.45%. Yet, the exceptional training accuracy of 99.95% raises considerations about potential overfitting, encouraging a thoughtful interpretation of its generalization capabilities. This observed difference between training and testing accuracies highlights the importance of fine-tuning and employing regularization strategies to delicately balance model intricacy for effective generalization.

In the landscape of clinical trials, envision a scenario where the Stacking Classifier boasts an impressive 83.45% testing accuracy. Picture a trial focused on identifying patients with rare genetic markers. The model's precision in recognizing these markers streamlines participant recruitment. However, the training accuracy at an exceptional 99.95% signals potential overfitting, demanding careful calibration. This conundrum highlights the critical need for fine-tuning and regularization techniques to ensure the model's reliability in identifying eligible participants with rare genetic markers, offering a nuanced perspective for further consideration.

Figure 5 shows ROC curves, which were used to assess their performance. The area under the curve (AUC) values demonstrated their discriminative power. Based on the provided ROC curve, it's evident that the Random Forest and Stacking models exhibit superior performance compared to the Tree, Logistic Regression, and SVM models. The Random Forest and Stacking models achieve an AUC of 0.89, indicating a high ability to distinguish between positive and negative cases. In contrast, the Tree model has an AUC of 0.78, followed by Logistic Regression (AUC = 0.76) and SVM (AUC = 0.69). The ROC curve demonstrates a trade-off between the True Positive Rate (TPR) and the False Positive Rate (FPR). A higher TPR indicates a better ability to correctly identify positive cases, while a lower FPR indicates a lower rate of mistakenly identifying negative cases as positive. The ideal model would have an AUC of 1.0, representing perfect classification accuracy. The Random Forest and Stacking models consistently achieve higher

TPR values across various FPR thresholds, suggesting their ability to accurately identify positive cases while minimizing false positives. In contrast, the Tree, Logistic Regression, and SVM models exhibit lower TPR values, indicating a higher likelihood of misclassifying positive cases. The ROC curve analysis highlights the superior performance of the Random Forest and Stacking models compared to the other models. These models demonstrate a strong ability to distinguish between positive and negative cases, making them suitable choices for classification tasks.

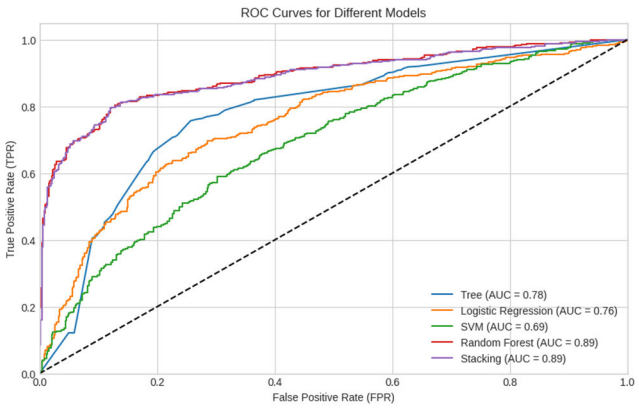


FIGURE 5. ROC curves of each model tested.

E. GAME THEORY MODEL PERFORMANCE EVALUATION

We conducted a thorough simulation employing randomization and value variation to assess the effectiveness of the suggested theoretic model in order to demonstrate and support our theoretical analysis about the equilibrium point mentioned in the above sections. Three patients, four doctors, six research firm officials (or any other authoritative body that could serve as a general decision-maker) have all been taken into consideration.

Initial unit variables associated with patients and doctors (unit significance of each strategy) for any number of doctors, patients and the allocated resources were chosen randomly from 0 to 1 and their initial values at given d were {0.61, 0.63, 0.70, 0.75, 0.80} with $K_{d,a}$ being 12% larger than C_d values, and the value of δ_a were selected from [1] and [6].

We plotted the fluctuation of many measures, including the value of research firms, doctors and patients vs. occurrences. First, we drew a plot of the research firm's variation in predicted utility, as seen in figure 6, where we could observe that each user's utility is rising until it converges. We next plotted, as seen in figure 7, the change in expected values of Doctors' unit importance towards each method vs. the occurrences. In order to offer competitive methods to influence additional research firm officials or more incentives or support from the research businesses, doctors are attempting to reduce their individual unit significance for the offered scenario.

Higher predicted payoffs for doctors are correlated with more research firms or more incentives from research firms, as demonstrated in figure 8. Finally, we examined the plot

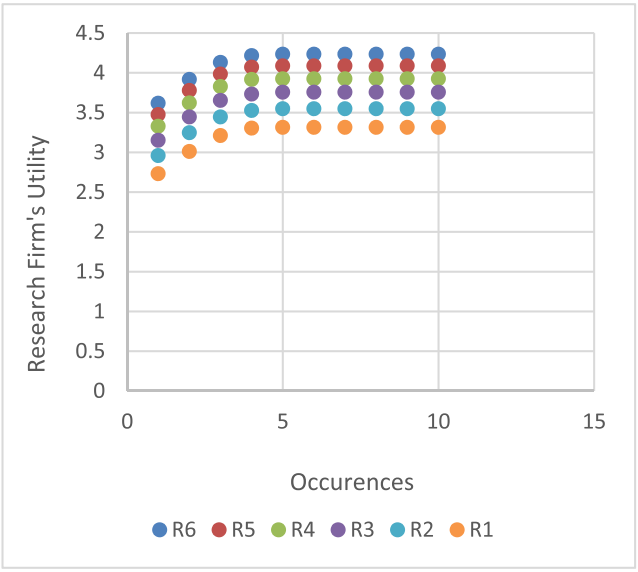


FIGURE 6. Variation of research firm's utility v/s occurrences.

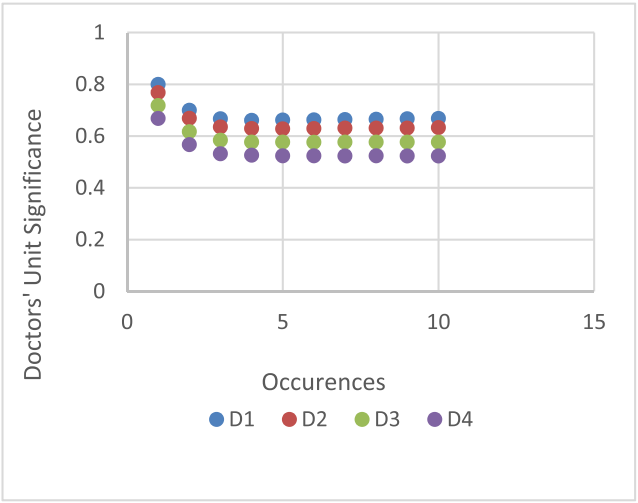


FIGURE 7. Doctor's unit significance v/s occurrences.

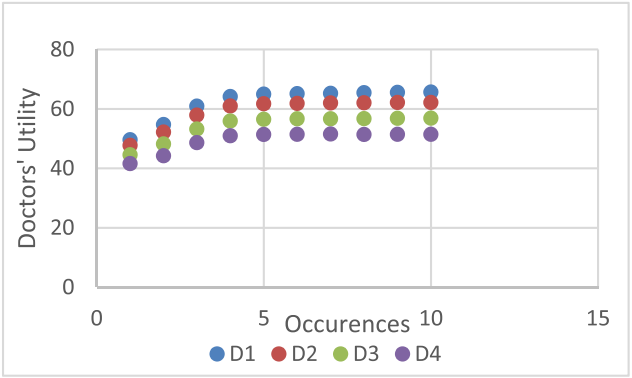


FIGURE 8. Doctor's utility v/s occurrences.

to determine the variation in the benefits for patients linked with each experiment compared to the occurrences, which is shown in figure 9.

In figure 9, we noticed that each patient was reducing their benefit in order to provide doctors more room to ask for benefits or clarification on trials. As a result, as seen in figure 10, Patients are receiving higher payment. These plots support our analysis from the previous section and demonstrate that the three-layer game reaches an ideal, singular equilibrium point where patients compete with doctors for their attention and doctors compete with research firms for their incentives. This results in high workability for the research firms and high payoffs for both doctors and patients.

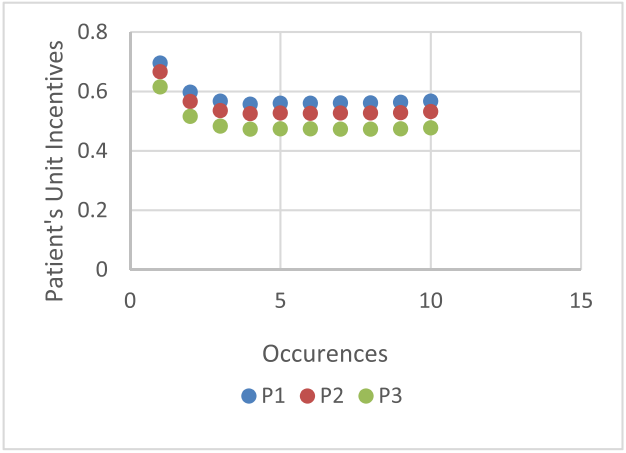


FIGURE 9. Patient's unit incentive v/s occurrences.

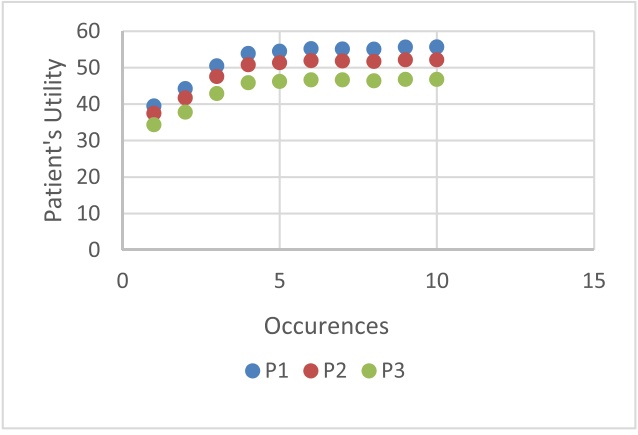


FIGURE 10. Patient's utility v/s occurrences.

V. CONCLUSION

This study has explored a wide range of techniques and interdisciplinary approaches with the goal of improving patient recruitment, diagnosis, and prediction performance in healthcare applications. Several interesting discoveries have been made as a result of an extensive literature review. The approaches from the literature review that were looked at have shown promise in improving the accuracy of diagnosis and prediction across a variety of data types.

The implementation of autoencoders in our machine learning approach has yielded impressive results, with an average Mean Squared Error (MSE) of 0.00022996273820050333. This signifies the model's capability to efficiently capture and represent complex relationships within the dataset. Notably, the shape of the encoded data, (4204, 700) as compared to dataset's actual size (4412,11), highlights the richness of insights extracted from the blood parameters. The choice to incorporate blood parameters into our analysis has proven to be judicious, as it serves as a valuable resource for uncovering hidden correlations and patterns. While it can be a daunting task for doctors to manually discern these intricate relationships, autoencoders step in to streamline the process. This emphasizes the synergistic potential of combining medical expertise with advanced machine learning techniques to empower clinical decision-making and facilitate medical research.

Our utilization of the stacking classifier underscores the power of combining different decision-making approaches to enhance the overall performance of our model. The very concept of ensemble learning draws inspiration from the idea that aggregating diverse viewpoints and strategies can lead to more robust and accurate decisions. In our specific case, we've harnessed the strengths of multiple machine learning models to study intricate relationships within our dataset, culminating in a comprehensive classification process for patient eligibility. This approach is akin to how a doctor leverages years of medical knowledge and diverse thought processes to make informed decisions for their patients. While we're limited to amalgamating the strengths of different models rather than the cognitive prowess of a human doctor, our remarkable testing accuracy of 83.07% and training accuracy of 99.95% underscores the effectiveness of this approach in the realm of machine learning and computational decision-making. Just as doctors employ a variety of techniques and insights to make critical decisions, our model showcases that similar principles can be applied in the domain of machine learning to deliver robust and reliable results.

This paper introduces a comprehensive game theoretic framework designed to facilitate more effective collaboration between doctors, research firms, and patients. The framework delves into the intricate interactions that take place among these three distinct groups of stakeholders. Through a meticulously designed three-layer patient-centric and influence-based game, this study explores a set of subgames that cater to the unique objectives of each player category. In these subgames, research firms take center stage as they endeavor to maximize their payoffs while adhering to stringent budget constraints. This entails making strategic decisions on how to allocate resources efficiently and engage with doctors and patients. On the other hand, doctors are in pursuit of optimizing their payoff, a goal achieved by not only attracting a higher number of patients but also by thoroughly understanding and addressing patients' individual needs and concerns. In essence, doctors must tailor their approaches to different patients to enhance their chances of recruitment. Patients, the third essential player group, strive to maximize their own payoff. Their success in this endeavour hinges on attracting

more requests from doctors, a goal that can be achieved by effectively conveying their needs and preferences.

Let's illustrate this with a simple example: Imagine a clinical trial scenario in which multiple doctors are competing to recruit patients. Research firms are responsible for overseeing the recruitment process efficiently, ensuring that resources are allocated judiciously. Doctors, on the other hand, are keen on recruiting as many patients as possible. Patients are looking to participate in trials that align with their preferences and concerns. Within this dynamic framework, the game theory model provides valuable insights into how these stakeholders can make strategic decisions, cooperate, and compete to achieve their respective objectives. As the subgames unfold, each class of player strives to maximize their utility, keeping a keen eye on their unique set of constraints and goals.

The best response of our game theoretic model signifies the optimal adjustment that each doctor should make to their recruitment strategy, given the strategies of other doctors and the available resources. It's a response that maximizes the doctor's utility, ensuring that their actions align with the current state of the game and the actions of other players. Equation (16) essentially takes into account the strategies employed by other doctors and the available resources allocated for that particular strategy. By adjusting their recruitment strategy based on these considerations, the doctor aims to optimize their payoff. An inference we can draw from this equation is that the best response represents a dynamic and adaptive strategy. Doctors continuously adapt their approach, considering the strategies employed by other doctors and the resources available. This adjustment ensures that the doctor is maximizing their effectiveness in attracting patients while staying within the constraints of available resources. For example, consider a scenario where one doctor observes that their recruitment strategy is performing better in comparison to other doctors, possibly due to its alignment with patient preferences. In such a case, the best response would involve allocating more resources to that strategy, increasing its significance. Conversely, if a strategy is less effective, the doctor might redistribute resources to more successful strategies. The equation provides a formalized way for doctors to adapt and optimize their recruitment strategies in response to changing dynamics within the game, ultimately contributing to a more efficient and effective recruitment process. This adaptability and optimization are key findings from our game theory model, demonstrating the benefits of dynamic decision-making strategies in complex interactions.

The proven benefits of this hybrid method highlight its potential to close the gap between real-world decision-making and predictive modelling, leading to better clinical trial patient recruitment outcomes. However, restrictions on dataset size and the requirement for thorough implementation guidance have been noted as potential roadblocks to their efficient application. Additionally, the computational requirements of the suggested approaches could be problematic, particularly in healthcare facilities with limited resources. This paper's unique strength is its multidisciplinary synergy, which combines machine learning methods with basic game theory ideas. Utilizing both sectors' best practices, this holistic approach makes it possible to improve patient recruiting procedures.

In diverse industries, a collaborative framework could be devised, similar to the one crafted for optimizing patient recruitment in clinical trials. This intricate methodology of ours, first utilizes machine learning techniques, merging them with fundamental game theory principles, extending its potential application beyond clinical trials. For instance, it could prove effective in detecting fraudulent financial transactions, ensuring precise defect identification in manufacturing processes, predicting customer churn in telecommunications, and facilitating personalized marketing campaigns in e-commerce based on user behavior. Following the machine learning phase, the framework transitions into the realm of game theory. For financial fraud detection, players could include Fraudulent Entities, Financial Institutions (Investigators), and Regulatory Bodies. In manufacturing quality control, the players could consist of Quality Control Systems, Production Line Managers (Inspectors), and Manufacturing Regulatory Bodies. Similarly, in customer retention for telecommunications, players may involve Telecommunication Service Providers, Subscribers (Customers), and Competing Telecommunication Companies. E-commerce personalized marketing engages E-commerce Platforms, Individual Customers, and Competing E-commerce Platforms. Lastly, in energy consumption optimization, players may include Energy Equipment and Systems, Maintenance Personnel, and Regulatory Bodies in the Energy Sector. In each industry, the two-step approach involves leveraging machine learning for effective classification based on relevant parameters and subsequently formulating a game theory model to optimize decision-making processes among the identified players. The framework's versatility extends beyond clinical trials, indicating its potential utility in addressing various challenges across different industries.

VI. FUTURE DIRECTION AND LIMITATION

Building upon this synergic and novel approach to optimize patient recruitment for clinical trials, there are several promising future directions to consider:

A. REFINEMENT OF HYBRID MODEL

Our focus centres on refining the hybrid model by incorporating cutting-edge machine learning techniques and classifiers to enhance the stacked ensemble's accuracy and robustness. An enticing avenue involves exploring graph neural networks, recognizing their potential to excel in decision-making scenarios, as many real-world choices can be aptly modelled within graph-like structures [55], [56]. This approach involves converting traditional tabular datasets into graph representations, offering a more comprehensive understanding of intricate relationships and dependencies within the data. Furthermore, the utilization of graph autoencoders enhances the depth and quality of information encoded

within complex graph structures. This augmented graph representation can be seamlessly integrated into our existing classification framework, bolstering our decision-making and eligibility predictions. In essence, this future work promises to empower our model with enhanced decision support capabilities through the adoption of advanced graph-based techniques.

B. DYNAMIC FEATURE SELECTION

We can explore dynamic feature selection techniques that adapt to the evolving landscape of the patient recruitment process [57]. By integrating methods like feature importance scores and recursive feature elimination, we can ensure that our model leverages the most pertinent features at each stage of recruitment, thereby enhancing both prediction accuracy and interpretability [58]. Additionally, the incorporation of deep reinforcement learning tactics could be a promising avenue [59]. For instance, an intelligent agent, equipped with a subset of blood parameters and a sample space of diseases, could dynamically determine the most relevant parameters for a given context. This agent might aim to maximize the number of correctly matched diseases with minimal parameters. These innovations, coupled with the existing methodology, have the potential to further elevate the efficiency and effectiveness of the patient recruitment process, facilitating better healthcare decisions. Exploring policy gradient methods, such as Proximal Policy Optimization (PPO) [60] and Trust Region Policy Optimization (TRPO) [61], can be instrumental in the task of dynamically selecting relevant blood parameters for specific diseases. These methods enable the development of intelligent agents that can adapt and learn optimal feature policies by interacting with the data, making them well-suited for the complex and evolving nature of healthcare data analysis. PPO and TRPO are known for their stability and sample efficiency, which makes them promising candidates for developing algorithms that can effectively choose the most appropriate blood parameters for accurate disease classification and healthcare decision-making.

C. REAL-TIME PATIENT RECRUITMENT OPTIMIZATION

Extending the 3-layer game theoretic model to incorporate real-time data and interactions opens up exciting possibilities for improving patient recruitment strategies in clinical trials. By embracing an unsupervised approach like association rule mining, we can extract valuable relations from evolving patient characteristics [62], clinical trial progress, and various external factors. This enables us to gain a deeper understanding of the ever-changing context and refine our recruitment strategies accordingly. In addition to emotional and incentive-based factors, exploring other relevant variables can provide a holistic view of patient behavior and preferences, facilitating more personalized and effective recruitment tactics. Investigating the feasibility of tailoring patient recruitment strategies at an individual level by considering patient demographics, medical history, and preferences can significantly enhance patient engagement and participation. This data-driven approach ensures that recruitment strategies remain adaptable and responsive, ultimately contributing to the success of clinical trials.

D. ETHICAL AND LEGAL CONSIDERATIONS

Incorporating the ethical and legal dimensions of this approach is paramount to ensure the responsible and secure deployment of our methodology. Collaboration with experts in healthcare ethics and data privacy is essential for addressing issues concerning patient consent, data protection, and compliance with healthcare regulations. Given the sensitive nature of data involved, particularly blood samples and clinical trial information, robust safeguards must be in place to prevent any malicious activities. To address these concerns, a federated learning approach can be adopted [63]. This decentralized process ensures that the real-time data processing, whether from blood parameters or individual patient profiles, occurs within distinct node structures. These nodes process data locally and share only valuable insights and results with a central hub. This way, the raw data remains secure and protected, while still allowing for advanced analyses and valuable inferences to be drawn, upholding both the ethical and legal aspects of our approach.

By exploring these future directions, our interdisciplinary approach that combines machine learning and game theory principles could pave the way for more efficient, effective, and patient-centric patient recruitment strategies in clinical trials, ultimately contributing to advancements in healthcare research and patient outcomes.

Regarding the limitations, it should be noted that the dataset, while valuable, it mostly focuses on laboratory test results and does not accurately reflect the variety of patient characteristics impacting trial eligibility. Despite being fundamental, the study's focus on classification prediction tasks could ignore the many patient features that affect recruiting choices. Even while it is approachable, the simplified methodology might not adequately capture the intricacies of the patient recruiting process, necessitating consideration of more sophisticated models. The model's applicability may also be impacted by the lack of real-world validation and any unexplained variables. Realizing these limitations highlights the need for future research to enhance the method. To ensure effective patient recruiting strategies for clinical trials, this additional study may involve larger datasets, more complex modelling techniques, and real-world testing. This work paves the way for improved patient recruitment tactics in healthcare research by addressing the issues raised in the literature review and providing a fresh interdisciplinary strategy. Further research built on this basis might help create a healthcare system that is more productive and effective.

ACKNOWLEDGMENT

The authors would like to take this opportunity to thank the management of Vellore Institute of Technology, for providing the APC and encouragement to carry out this work.

REFERENCES

[1] M. Sequeira, J. R. Almeida, and J. L. Oliveira, "A comparative analysis of data platforms for rare diseases," in *Proc. IEEE 34th Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2021, pp. 366–371.

[2] M. Amiridi, C. Qian, N. D. Sidiropoulos, and L. M. Glass, "Enrollment rate prediction in clinical trials based on CDF sketching and tensor factorization tools," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Jun. 2023, pp. 1–5.

[3] P. A. Torres-Saavedra and K. A. Winter, "An overview of phase 2 clinical trial designs," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 112, no. 1, pp. 22–29, Jan. 2022, doi: 10.1016/J.IJROBP.2021.07.1700.

[4] M. Paul, Y. Dishon-Benattar, Y. Dickstein, and D. Yahav, "Optimizing patient recruitment into clinical trials of antimicrobial-resistant pathogens," *JAC-Antimicrobial Resistance*, vol. 5, no. 1, pp. 1–6, Dec. 2022, doi: 10.1093/jacamr/dlad005.

[5] M. Brøgger-Mikkelsen, J. R. Zibert, A. D. Andersen, U. Lassen, Z. Ali, and S. F. Thomsen, "Changes in key recruitment performance metrics from 2008–2019 in industry-sponsored phase III clinical trials registered at ClinicalTrials.Gov," *PLoS ONE*, vol. 17, no. 7, Jul. 2022, Art. no. e0271819, doi: 10.1371/journal.pone.0271819.

[6] J. Liu, C. Yu, C. Li, and J. Han, "Cooperation or conflict in doctor-patient relationship? An analysis from the perspective of evolutionary game," *IEEE Access*, vol. 8, pp. 42898–42908, 2020.

[7] N. Liu, Z. X. Koh, E. C. Chua, L. M. Tan, Z. Lin, B. Mirza, and M. E. H. Ong, "Risk scoring for prediction of acute cardiac complications from imbalanced clinical data," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 6, pp. 1894–1902, Nov. 2014.

[8] R. Benedek and R. Sarkar, "The Shapley value of classifiers in ensemble games," in *Proc. 30th ACM Int. Conf. Inf. Knowl. Manag.*, 2021, pp. 1558–1567, doi: 10.1145/3459637.3482302.

[9] B. Miyata, B. Tafuto, and N. Jose, "76 methods and perceptions of success for patient recruitment in decentralized clinical trials," *J. Clin. Transl. Sci.*, vol. 7, no. s1, p. 21, Apr. 2023, doi: 10.1017/cts.2023.160.

[10] C. Weng and P. J. Embi, "Informatics approaches to participant recruitment," in *Clinical Research Informatics*. Cham, Switzerland: Springer, 2023, pp. 219–229.

[11] B. Idnay, Y. Fang, C. Dreisbach, K. Marder, C. Weng, and R. Schnall, "Clinical research staff perceptions on a natural language processing-driven tool for eligibility prescreening: An iterative usability assessment," *Int. J. Med. Informat.*, vol. 171, Mar. 2023, Art. no. 104985, doi: 10.1016/j.ijmedinf.2023.104985.

[12] A. Ahad, M. Tahir, M. A. S. Sheikh, N. Hassan, K. I. Ahmed, and A. Mughees, "A game theory based clustering scheme (GCS) for 5G-based smart healthcare," in *Proc. IEEE 5th Int. Symp. Telecommun. Technol. (ISTT)*, Nov. 2020, pp. 157–161.

[13] Z. Ning, P. Dong, X. Wang, X. Hu, L. Guo, B. Hu, Y. Guo, T. Qiu, and R. Y. K. Kwok, "Mobile edge computing enabled 5G health monitoring for Internet of Medical Things: A decentralized game theoretic approach," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 2, pp. 463–478, Feb. 2021.

[14] B. S. Raja and S. Asghar, "Disease classification in health care systems with game theory approach," *IEEE Access*, vol. 8, pp. 83298–83311, 2020, doi: 10.1109/ACCESS.2020.2991016.

[15] V. U. Udeme and U. C. Orumie, "Patients' preferences of healthcare facilities for quality healthcare services in Akwa Ibom state: A game theory approach," *Amer. J. Oper. Res.*, vol. 11, no. 3, pp. 181–198, 2021, doi: 10.4236/AJOR.2021.113011.

[16] A. G. Powell, C. Bowman, C. Brown, R. J. Egan, and W. G. Lewis, "Team strategic philosophy: Requiem for the infinite game," *Postgraduate Med. J.*, vol. 96, no. 1136, pp. 310–312, Jun. 2020, doi: 10.1136/POSTGRADMEDJ-2019-137441.

[17] T. A. Ahanger, M. Bhatia, and A. Aljumah, "Game-theory-inspired novel mechanism for assessing healthcare quality," *IEEE Internet Things J.*, vol. 9, no. 24, pp. 24856–24865, Dec. 2022, doi: 10.1109/JIOT.2022.3194854.

[18] J. Liu, P. J. Allen, L. Benz, D. Blickstein, E. Okidi, and X. Shi, "A machine learning approach for recruitment prediction in clinical trial design," 2021, *arXiv:2111.07407*.

[19] D. S. Robertson, K. M. Lee, B. C. López-Kolkovska, and S. S. Villar, "Response-adaptive randomization in clinical trials: From myths to practical considerations," *Stat. Sci.*, vol. 38, no. 2, p. 185, May 2023.

[20] J. C. Brehautet al., "Using behavioral theory and shared decision-making to understand clinical trial recruitment: Interviews with trial recruiters," *Trials*, vol. 22, p. 298, 2021, doi: 10.1186/S13063-021-05257-X.

[21] Z. Tian, W. Han, and W. B. Powell, "Adaptive learning of drug quality and optimization of patient recruitment for clinical trials with dropouts," *Manuf. Service Oper. Manage.*, vol. 24, no. 1, pp. 580–599, Jan. 2022, doi: 10.1287/MSOM.2020.0936.

[22] W. Huang, H. Suominen, T. Liu, G. Rice, C. Salomon, and A. S. Barnard, "Explainable discovery of disease biomarkers: The case of ovarian cancer to illustrate the best practice in machine learning and Shapley analysis," *J. Biomed. Informat.*, vol. 141, May 2023, Art. no. 104365, doi: 10.1016/j.jbi.2023.104365.

[23] D.-K. Nguyen, C.-H. Lan, and C.-L. Chan, "Deep ensemble learning approaches in healthcare to enhance the prediction and diagnosing performance: The workflows, deployments, and surveys on the statistical, image-based, and sequential datasets," *Int. J. Environ. Res. Public Health*, vol. 18, no. 20, p. 10811, Oct. 2021, doi: 10.3390/IJERPH182010811.

[24] Y. Gu, A. Zalkikar, M. Liu, L. Kelly, A. Hall, K. Daly, and T. Ward, "Predicting medication adherence using ensemble learning and deep learning models with large scale healthcare data," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 18961.

[25] V. K. Daliya and T. K. Ramesh, "A machine learning based ensemble approach for predictive analysis of healthcare data," in *Proc. 2nd PhD Colloq. Ethically Driven Innov. Technol. Soc. (PhD EDITS)*, Nov. 2020, pp. 1–2, doi: 10.1109/PhDEDITS51180.2020.9315300.

[26] Y. Zhang, X. Wang, N. Han, and R. Zhao, "Ensemble learning based postpartum hemorrhage diagnosis for 5G remote healthcare," *IEEE Access*, vol. 9, pp. 18538–18548, 2021, doi: 10.1109/ACCESS.2021.3051215.

[27] P. Bhowal, S. Sen, J. D. Velasquez, and R. Sarkar, "Fuzzy ensemble of deep learning models using choquet fuzzy integral, coalition game and information theory for breast cancer histology classification," *Expert Syst. Appl.*, vol. 190, Mar. 2022, Art. no. 116167, doi: 10.1016/J.ESWA.2021.116167.

[28] W. Pannakkong, V. T. Vinh, N. N. M. Tuyen, and J. Buddhakulsomsiri, "A reinforcement learning approach for ensemble machine learning models in peak electricity forecasting," *Energies*, vol. 16, no. 13, p. 5099, Jul. 2023, doi: 10.3390/en16135099.

[29] A. Onan, "GTR-GA: Harnessing the power of graph-based neural networks and genetic algorithms for text augmentation," *Expert Syst. Appl.*, vol. 232, Dec. 2023, Art. no. 120908, doi: 10.1016/j.eswa.2023.120908.

[30] A. Onan, "Hierarchical graph-based text classification framework with contextual node embedding and BERT-based dynamic fusion," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 7, Jul. 2023, Art. no. 101610, doi: 10.1016/j.jksuci.2023.101610.

[31] A. Onan, "Bidirectional convolutional recurrent neural network architecture with group-wise enhancement mechanism for text sentiment classification," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 34, no. 5, pp. 2098–2117, May 2022, doi: 10.1016/j.jksuci.2022.02.025.

[32] A. Onan, "SRL-ACO: A text augmentation framework based on semantic role labeling and ant colony optimization," *J. King Saud Univ.-Comput. Inf. Sci.*, vol. 35, no. 7, Jul. 2023, Art. no. 101611, doi: 10.1016/j.jksuci.2023.101611.

[33] A. Onan, "Two-stage topic extraction model for bibliometric data analysis based on word embeddings and clustering," *IEEE Access*, vol. 7, pp. 145614–145633, 2019, doi: 10.1109/ACCESS.2019.2945911.

[34] D. S. Watson and L. Floridi, "The explanation game: A formal framework for interpretable machine learning," *Synthese*, vol. 198, no. 10, pp. 9211–9242, Oct. 2021, doi: 10.1007/S11229-020-02629-9.

[35] D. B. Rawat, A. Alshaikhi, A. Alshammari, C. Bajracharya, and M. Song, "Payoff optimization through wireless network virtualization for IoT applications: A three layer game approach," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2797–2805, Apr. 2019, doi: 10.1109/JIOT.2018.2874884.

[36] G. Kumar, J. P. Singh, and A. K. Singh, "Autoencoder-based feature extraction for identifying hate speech spreaders in social media," *IEEE Trans. Computat. Social Syst.*, doi: 10.1109/TCSS.2023.3240098.

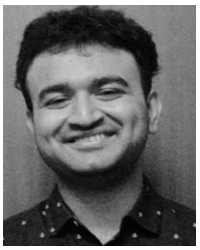
[37] Y. Zhang, Y. Zhou, S. Xu, D. Hong, H. Gao, C. Li, Q. Zhong, and B. Zhang, "Depthwise separable convolutional autoencoders for hyperspectral image change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, pp. 1–5, 2023.

[38] B. Yan and G. Han, "Effective feature extraction via stacked sparse autoencoder to improve intrusion detection system," *IEEE Access*, vol. 6, pp. 41238–41248, 2018.

[39] M. Chen, X. Shi, Y. Zhang, D. Wu, and M. Guizani, "Deep feature learning for medical image analysis with convolutional autoencoder neural network," *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 750–758, Oct. 2021.

[40] M. Qaraad, S. Amjad, I. I. M. Manhrawy, H. Fathi, B. A. Hassan, and P. E. Kafrawy, "A hybrid feature selection optimization model for high dimension data classification," *IEEE Access*, vol. 9, pp. 42884–42895, 2021.
[41] S. Nag, A. T. K. Baidya, A. Mandal, A. T. Mathew, B. Das, B. Devi, and R. Kumar, "Deep learning tools for advancing drug discovery and development," *3 Biotech*, vol. 12, no. 110, pp. 1–21, 2022, doi: 10.1007/s13205-022-03165-8.
[42] B. Perry, C. Dombeck, J. B. Smalley, B. Levitan, D. Leventhal, B. Patrick-Lake, L. Brennan, K. McKenna, Z. Hallinan, and A. Corneli, "Development and application of a patient group engagement prioritization tool for use in medical product development," *Therapeutic Innov. Regulatory Sci.*, vol. 55, no. 2, pp. 324–335, Mar. 2021, doi: 10.1007/S43441-020-00217-0.
[43] K. I. Mohammed, J. Jaafar, A. A. Zaidan, O. S. Albahri, B. B. Zaidan, K. H. Abdulkareem, A. N. Jasim, Ali. H. Shareef, M. J. Baqer, A. S. Albahri, M. A. Alsalem, and A. H. Alamoodi, "A uniform intelligent prioritisation for solving diverse and big data generated from multiple chronic diseases patients based on hybrid decision-making and voting method," *IEEE Access*, vol. 8, pp. 91521–91530, 2020.
[44] D. Aubel, A. Ciupek, S. Olshan, Z. Pemberton-Whiteley, A. Stunt, K. Davis, M. Piggin, M. C. Saxton, A. Olshan, A. Salamakha, F. Cardoso, and G. Mason, "Creating patient-friendly informed consent form (ICF): Insights from global oncology patient insights panels (GOPIPs)," *J. Clin. Oncol.*, vol. 41, no. 16, Jun. 2023, Art. no. e24199, doi: 10.1200/jco.2023.41.16_suppl.e24199.
[45] K. Kroenke, T. A. Miksch, A. C. Spaulding, G. L. Mazza, C. C. DeStephano, S. K. Niazi, A. J. C. Illies, M. Bydon, P. J. Novotny, A. Goyal, and M. K. Lee, "Choosing and using patient-reported outcome measures in clinical practice," *Arch. Phys. Med. Rehabil.*, vol. 103, no. 5, pp. S108–S117, May 2022, doi: 10.1016/J.APMR.2020.12.033.
[46] A. Heath, M. G. Myriam Hunink, E. Krijkamp, and P. Pechlivanoglou, "Prioritisation and design of clinical trials," *Eur. J. Epidemiol.*, vol. 36, no. 11, pp. 1111–1121, Nov. 2021, doi: 10.1007/S10654-021-00761-5.
[47] T. M. Baetu, "The virtues and limitations of randomized experiments," *Acta Analytica*, vol. 37, no. 4, pp. 453–470, Dec. 2022, doi: 10.1007/S12136-021-00497-7.
[48] I. S. Khayal and A. M. Farid, "Designing patient-oriented healthcare services as a system of systems," in *Proc. 13th Annu. Conf. Syst. Syst. Eng. (SoSE)*, Jun. 2018, pp. 233–239.
[49] T. Zwiep, S. Ahn, J. Brehaut, F. Balaa, D. I. McIsaac, S. Rich, T. Wallace, and H. Moloo, "Group practice impacts on patients, physicians and healthcare systems: A scoping review," *BMJ Open*, vol. 11, no. 1, Jan. 2021, Art. no. e041579, doi: 10.1136/BMJOPEN-2020-041579.
[50] B. M. Buckley, "The patient as sub-investigator," in *Innovation in Clinical Trial Methodologies*. Academic, 2021, doi: 10.1016/B978-0-12-824490-6.00005-0.
[51] A. Rosca, T. Krones, and N. Biller-Andorno, "Shared decision making: Patients have a right to be informed about possible treatment options and their risks and benefits," *Swiss Med. Weekly*, vol. 150, no. 2728, Jul. 2020, Art. no. w20268, doi: 10.4414/SMW.2020.20268.
[52] K. Gillies and M. K. Campbell, "Development and evaluation of decision aids for people considering taking part in a clinical trial: A conceptual framework," *Trials*, vol. 20, no. 1, pp. 1–11, 2019.
[53] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[54] M. Sadikin, "EHR dataset for patient treatment classification," Mendeley Data, V1, 2020, doi: 10.17632/7kv3rctx7m.1.
[55] Q. Li, F. Gama, A. Ribeiro, and A. Prorok, "Graph neural networks for decentralized multi-robot path planning," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Oct. 2020, pp. 11785–11792.
[56] E. S. da Silva, H. Pedrini, and A. Santos, "Applying graph neural networks to support decision making on collective intelligent transportation systems," *IEEE Trans. Netw. Service Manage.*, vol. 20, no. 4, pp. 4085–4096, Dec. 2023, doi: 10.1109/TNSM.2023.3257993.
[57] A. Perperoglou, Y. Zhang, and D. Kipourou, "Modeling time-varying recruitment rates in multicenter clinical trials," *Biometrical J.*, vol. 65, no. 6, 2023, Art. no. 2100377.
[58] M. Cuggia, P. Besana, and D. Glasspool, "Comparing semi-automatic systems for recruitment of patients to clinical trials," *Int. J. Med. Informat.*, vol. 80, no. 6, pp. 371–388, Jun. 2011.
[59] M. Zhu and H. Zhu, "Learning a diagnostic strategy on medical data with deep reinforcement learning," *IEEE Access*, vol. 9, pp. 84122–84133, 2021.
[60] C. Hettiarachchi, N. Malagutti, C. Nolan, E. Daskalaki, and H. Suominen, "A reinforcement learning based system for blood glucose control without carbohydrate estimation in type 1 diabetes: In silico validation," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2022, pp. 950–956.
[61] P. V. Astillo, G. Choudhary, D. G. Duguma, J. Kim, and I. You, "TrMAps: Trust management in specification-based misbehavior detection system for IMD-enabled artificial pancreas system," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 10, pp. 3763–3775, Oct. 2021.
[62] A. M. Khedr, Z. A. Aghbari, A. A. Ali, and M. Eljamil, "An efficient association rule mining from distributed medical databases for predicting heart diseases," *IEEE Access*, vol. 9, pp. 15320–15333, 2021.
[63] Q. Wu, X. Chen, Z. Zhou, and J. Zhang, "FedHome: Cloud-edge based personalized federated learning for in-home health monitoring," *IEEE Trans. Mobile Comput.*, vol. 21, no. 8, pp. 2818–2832, Aug. 2022.

C. U. OM KUMAR received the B.Tech. and M.Tech. degrees (Hons.) in computer science and engineering from Jawaharlal Nehru Technological University Anantapur, in 2006 and 2013, respectively, and the Ph.D. degree in cyber security from Anna University, Chennai, in 2020. From 2013 to 2016, he embarked on a teaching career with the SRM Easwari Engineering College, Chennai, where he discovered his passion for teaching and honed his pedagogical skills. He is currently an Assistant Professor (Senior Grade) with the School of Computer Science Engineering, Vellore Institute of Technology, Chennai Campus, where he has a distinguished academic and professional journey. His research interests include the IoT and deep learning, leading to the filing of an international patent and the publication of 25 research articles in reputable journals, establishing an H-Index value of 6.



ISSHAAN SINGH was born in Delhi, India, in 2003. He is currently pursuing the B.Tech. degree in computer science, specialization in artificial intelligence and robotics, with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai Campus, Chennai, Tamil Nadu, India. Since 2023, he has been involve in research work, in the domain of machine learning, game theory, and graph theory. His research interests include healthcare analytics and optimization of systems or real-life scenarios.

M. SUGUNA received the B.E. degree in computer science and engineering (CSE) from the Kumaraguru College of Technology, Coimbatore, the M.E. degree in CSE from the Government College of Technology, Coimbatore, and the Ph.D. degree in information and communication engineering from Anna University, Chennai.

She is currently an Assistant Professor (Senior Grade II) with the School of Computer Science and Engineering, Vellore Institute of Technology, Chennai. She has published more than 20 research papers in international journals and 15 international conferences. Her research interests include data analytics, health care analytics, cloud computing, and Agile project management. She is a member of ISTE and IAENG.