Journal of Cardiovascular Disease Research
ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024

OPTIMIZING PATIENT RECRUITMENT FOR CLINICAL TRIALS: A HYBRID CLASSIFICATION MODEL AND GAME-THEORETIC APPROACH FOR STRATEGIC INTERACTION

DR. RANJITH[1], E. MANVITHA[2], CH. LIPI VITHAL[3], CARIS HARSHITHA PAYARDA[4]
ASSOCIATE PROFESSOR 1, UG STUDENT 2,3,4
DEPARTMENT OF IT, MALLA REDDY ENGINEERIGNG COLLEGE FOR WOMEN (UGC-AUTONOMOUS), MAISAMMGUDA, HYDERABAD, TELANGANA-500100

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. Patient recruitment for clinical trials is a critical aspect of advancing medical research but has traditionally relied on manual or semi-automated methods, such as advertisements, referrals, and eligibility screening through databases. These approaches often suffer from inefficiencies, including high costs, prolonged timelines, and suboptimal participant matching, leading to delays in trial completion. These challenges highlight the limitations of traditional methods in meeting the growing demand for efficient and effective recruitment. Motivated by these shortcomings, this study proposes an advanced AI-driven system leveraging a Random Forest Classifier within a hybrid framework to address these issues. The proposed model integrates Logistic Regression, Support Vector Machines, Decision Trees, and Random Forests through a stacking classifier and employs Autoencoders for feature extraction to enhance predictive accuracy. Additionally, a game-theoretic approach involving Patients, Clinical Investigators, and Research Firms optimizes recruitment strategies through strategic interaction, providing equilibrium-seeking solutions for balanced payoffs among stakeholders. This AI-based approach offers a transformative solution, addressing inefficiencies in traditional recruitment methods and significantly advancing the field of healthcare research by streamlining the recruitment process.

INTRODUCTION

Clinical trials are integral to advancing medical science, yet patient recruitment remains a significant bottleneck, particularly in India. According to reports, over **40% of clinical trials globally are delayed** due to recruitment issues, with India facing unique challenges such as limited awareness and accessibility. In India, the **clinical trial market** is projected to grow at a **CAGR of 8.7%**, reaching **$3.15 billion by 2025**, but inefficiencies in patient recruitment hamper this growth. Traditional methods often fail to consider India's diverse demographics, cultural barriers, and the low digitalization of healthcare, resulting in an average recruitment delay of **4-6 months per trial**, escalating costs and slowing innovations in medicine. Efficient patient recruitment in clinical trials enhances medical research, reduces costs, and improves healthcare outcomes. This project applies AI to streamline recruitment, leveraging machine learning for predictive accuracy and game theory for strategic decision-making. The hybrid approach ensures optimized recruitment and faster trial completion, advancing healthcare delivery and drug development. Before machine learning, patient recruitment relied on **manual processes** like advertisements, physician referrals, and screening through databases, which were time-consuming and error-prone. Challenges included **limited outreach**, **mismatched eligibility**, and **data mismanagement**, leading to high costs and significant delays. Recruitment inefficiencies caused **40% of trials to fail to meet enrollment goals** and delayed trials by an average of **6 months**, stalling innovations in healthcare. These traditional methods lacked scalability and adaptability to modern medical needs. The inefficiencies of traditional recruitment methods highlight the need for **advanced, scalable solutions**. The growing clinical trial market demands **data-driven systems** to address recruitment delays and high costs. AI-driven approaches can leverage diverse patient data for **precise matching**, ensuring inclusion across demographics. The motivation stems from the potential to significantly **reduce trial timelines**, cut costs, and enhance trial outcomes, ultimately benefiting patients and accelerating healthcare innovations. Manual recruitment methods like advertisements, referrals, and basic database screening dominate the process, often missing eligible candidates or including ineligible ones. These methods are **costly**, **slow**, and **prone to errors**, with no mechanism for strategic optimization among stakeholders. The lack of robust predictive tools limits their ability to handle large-scale data and diverse patient pools, leading to recruitment failures and trial delays.

LITERATURE REVIEW

Research on Design and Application of Online English Education Platform Based on Web

- Yijuan Huang
- Published 22 October 2021

Aiming at the problems of low user satisfaction and long response time in traditional online English education platforms, this paper designs a web-based online English education platform and carries out relevant application tests. With the support of the web, the hardware and software of the platform are designed. The hardware of the platform is composed of student module, teacher module, administrator module, and database module. On this basis, the K-means clustering method is used to cluster the learner data in the web, to determine their English level, and the collaborative filtering algorithm is combined to recommend relevant course materials for learners so as to complete the platform software design. The application test results show that the platform has the advantages of superior function and security, high user satisfaction, and short response time and has certain practical application value.

Development of species recognition models using Google teachable machine on shorebirds and waterbirds

- Jenny Jenn Ney Wong, N. Fadzly
- Published in Journal of Taibah University… 11 November 2022

Species identification is an essential ability in every conservation initiative. An efficient and robust computer vision method was attested with an available online tool with Google's Teachable Machine. This pilot study on developing a species recognition app was to create and evaluate the usability and accuracy of using Teachable Machine for species identification at Teluk Air Tawar, Kuala Muda (TAT-KM), Malaysia. The accuracy of the created models was evaluated and compared with training images based on the web-mining (Google Images Repository) compared to actual photos taken at the same site. Model A (Google Image) had an average accuracy of 55.30%, while Model B (actual photos) was 99.42%. Regarding success rate at accuracy over 77%, 27 out of 49 test images (55.10%) were reported in Model A, while Model B had a 100% success rate. This approach can replace traditional methods of bird species recognition to handle large amounts of data.

BMRI-NET: A Deep Stacked Ensemble Model for Multi-class Brain Tumor Classification from MRI Images

- S. Asif, Mingde Zhao, +1 author Yusen Zhu
- Published in Interdisciplinary Sciences… 12 May 2023

Brain tumors are one of the most dangerous health problems for adults and children in many countries. Any failure in the diagnosis of brain tumors may lead to shortening of human life. Accurate and timely diagnosis of brain tumors provides appropriate treatment to increase the patient's chances of survival. Due to the different characteristics of tumors, one of the challenging problems is the classification of three types of brain tumors. With the advent of deep learning (DL) models, three classes of brain tumor classification have been addressed. However, the accuracy of these methods requires significant improvements in brain image classification. The main goal of this article is to design a new method for classifying the three types of brain tumors with extremely high accuracy. In this paper, we propose a novel deep stacked ensemble model called "BMRI-NET" that can detect brain tumors from MR images with high accuracy and recall. The stacked ensemble proposed in this article adapts three pre-trained models, namely DenseNe201, ResNet152V2, and InceptionResNetV2, to improve the generalization capability. We combine decisions from the three models using the stacking technique to obtain final results that are much more accurate than individual models for detecting brain tumors. The efficacy of the proposed model is evaluated on the Figshare brain MRI dataset of three types of brain tumors consisting of 3064 images. The experimental results clearly highlight the robustness of the proposed BMRI-NET model by achieving an overall classification of 98.69% and an average recall, F1-score and MCC of 98.33%, 98.40, and 97.95%, respectively. The results indicate that the proposed BMRI-NET model is superior to existing methods and can assist healthcare professionals in the diagnosis of brain tumors. Graphical Abstract.

EXISTING SYSTEM

Patient recruitment for clinical trials has traditionally been a labor-intensive and resource-intensive process. The primary methods include advertisements through print, radio, and online media to reach potential participants. Additionally, referrals from healthcare providers and direct outreach in clinics or hospitals are common practices. Eligibility screening is conducted manually or through semi-automated systems using patient databases, medical records, or health registries to identify suitable candidates based on trial requirements. These traditional methods often involve recruitment agencies or dedicated staff who screen, contact, and engage participants. While these systems rely on personal interaction and professional networks, they are constrained by their manual nature. Tools like spreadsheets or basic software for organizing data provide limited support for scaling recruitment efforts. The recruitment process is often prolonged as investigators navigate numerous applications to find eligible participants. Furthermore, patient enrollment monitoring typically involves manual follow-ups and periodic updates, which are time-consuming and prone to errors. Recruitment campaigns are also limited by geographical constraints, as outreach efforts are often localized, excluding a significant pool of potential participants. Despite incremental improvements with semi-automated databases, these systems struggle to handle large-scale recruitment efficiently.

Disadvantages:

- Time-Consuming and Labor-Intensive

The manual screening, contacting, and follow-up processes require significant time and resources, leading to prolonged recruitment timelines.

- Geographical Limitations

Outreach efforts are often localized, which excludes potential participants who reside outside the immediate area, limiting the diversity and size of the candidate pool.

ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024

- Inefficiency in Handling Large-Scale Recruitment

Traditional methods and semi-automated systems struggle to manage large volumes of data and participants, making scaling recruitment efforts difficult.

- Prone to Errors

Manual monitoring and follow-ups increase the risk of errors in data handling, participant tracking, and communication, potentially affecting trial accuracy and reliability.

- Limited Use of Advanced Technology

Reliance on basic tools like spreadsheets or simple software does not fully leverage modern technology, reducing efficiency and adaptability in addressing dynamic recruitment challenges.

PROPOSED SYSTEM

Random Forest Classifier (RFC)

What is RFC? Random Forest Classifier (RFC) is an ensemble learning method that combines multiple decision trees to improve classification accuracy. Each tree in the forest is built using a subset of the data and a subset of features. The final prediction is made by averaging the results of all individual trees.

How it Works: RFC operates by constructing several decision trees, each trained on a random subset of the training data. For classification tasks, each tree produces a class prediction, and the majority vote from all trees is taken as the final result. This process reduces overfitting and improves accuracy compared to individual decision trees.

Architecture:

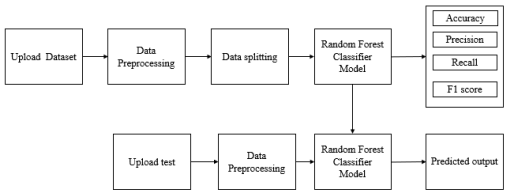
1. The dataset is divided into multiple random subsets.
2. Each subset is used to train an individual decision tree.
3. Features are randomly selected at each split to build diverse trees.
4. Each tree independently makes predictions.
5. The final prediction is based on the majority vote of all trees.

Advantages :

1. Reduces overfitting compared to individual decision trees.
2. Can handle both classification and regression tasks.
3. Effective for large datasets and can handle high-dimensional feature spaces.
4. Robust to noise and outliers in the data.
5. Parallelizable, making it scalable and faster on large datasets.

IMPLEMENTATION

SYSTEM ARCHITECTURE



MODULES

NumPy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

- A powerful N-dimensional array object
- Sophisticated (broadcasting) functions
- Tools for integrating C/C++ and Fortran code
- Useful linear algebra, Fourier transform, and random number capabilities

Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data. Arbitrary datatypes can be defined using NumPy which allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

Pandas

Pandas is an open-source Python Library providing high-performance data manipulation and analysis tool using its powerful data structures. Python was majorly used for data munging and preparation. It had very little contribution towards data analysis. Pandas solved this problem. Using Pandas, we can accomplish five typical steps in the processing and analysis of data, regardless of the origin of data load, prepare, manipulate, model, and analyze. Python with Pandas is used in a wide range of fields including academic and commercial domains including finance, economics, Statistics, analytics, etc.

Matplotlib

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and

ISSN: 0975-3583, 0976-2833 VOL15, ISSUE12, 2024

interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code. For examples, see the sample plots and thumbnail gallery.

For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined with IPython. For the power user, you have full control of line styles, font properties, axes properties, etc, via an object oriented interface or via a set of functions familiar to MATLAB users.

Scikit – learn

Scikit-learn provides a range of supervised and unsupervised learning algorithms via a consistent interface in Python. It is licensed under a permissive simplified BSD license and is distributed under many Linux distributions, encouraging academic and commercial use. Python

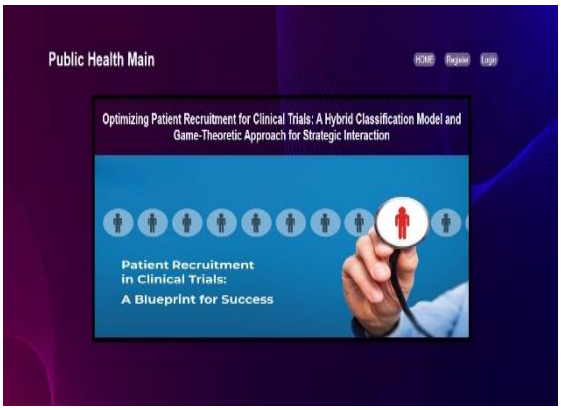
Python is an interpreted high-level programming language for general-purpose programming. Created by Guido van Rossum and first released in 1991, Python has a design philosophy that emphasizes code readability, notably using significant whitespace.

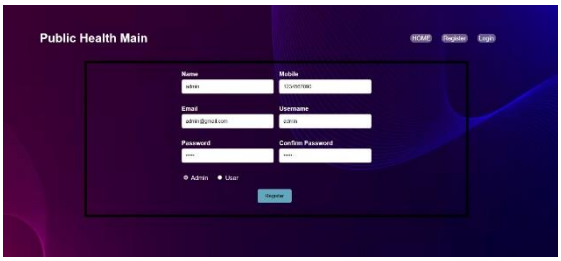
Python features a dynamic type system and automatic memory management. It supports multiple programming paradigms, including object-oriented, imperative, functional and procedural, and has a large and comprehensive standard library.

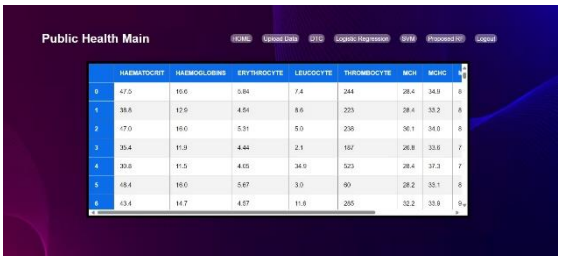
- Python is Interpreted − Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
- Python is Interactive − you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.

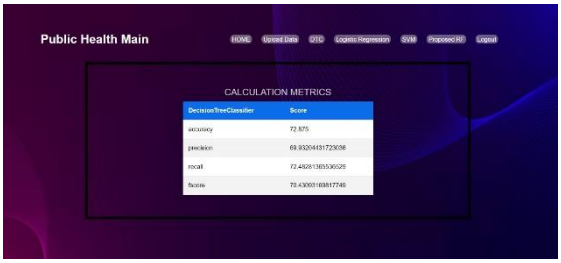
Python also acknowledges that speed of development is important. Readable and terse code is part of this, and so is access to powerful constructs that avoid tedious repetition of code. Maintainability also ties into this may be an all but useless metric, but it does say something about how much code you have to scan, read and/or understand to troubleshoot problems or tweak behaviors. This speed of development, the ease with which a programmer of other languages can pick up basic Python skills and the huge standard library is key to another area where Python excels. All its tools have been quick to implement, saved a lot of time, and several of them have later been patched and updated by people with no Python background - without breaking.

RESULT









CONCLUSION

This study demonstrates the significant potential of leveraging advanced AI methodologies, particularly a hybrid classification model combined with game-theoretic strategies, to revolutionize patient recruitment in clinical trials. Traditional recruitment methods, while foundational, are marred by inefficiencies, high costs, and lengthy timelines. By adopting a hybrid framework that incorporates machine learning techniques like Random Forest Classifier, Logistic Regression, Support Vector Machines, and Decision Trees, the proposed system enhances predictive accuracy in participant matching. The inclusion of autoencoders for feature extraction further optimizes the recruitment process, ensuring a higher likelihood of trial completion within stipulated timelines. The integration of game theory introduces a novel perspective to strategic interaction among stakeholders—patients, clinical investigators, and research firms—facilitating

balanced payoffs and equitable recruitment strategies. This holistic approach not only improves efficiency but also aligns the incentives of all parties involved, fostering collaboration and mutual benefit. By addressing the challenges of traditional systems, this research contributes to advancing healthcare innovation and reducing delays in life-saving medical research.

The findings underscore the transformative impact of AI in healthcare, offering a scalable and efficient solution to recruitment challenges. Ultimately, this hybrid model paves the way for more robust, timely, and cost-effective clinical trials, enhancing patient outcomes and accelerating the development of medical breakthroughs.

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 processingdriven 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 5Gbased 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.