IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2024

DOI: 10.17148/IJARCCE.2024.131006

Formulating a Dual Prediction ML Model for Patient Outcomes and Hospital Resource Use Through Electronic Health Records

Tapankumar A. Kakani[1]

Software Developer, Saurashtra University, Department of IT, Pactiv Evergreen Inc., Mundelein, IL, USA

Abstract: Electronic Health Records (EHRs) are valuable for predicting patient outcomes and hospital resource demands. However, EHR data's vast scale and intricate structure present significant challenges for traditional predictive models. This paper introduces a dual prediction machine learning computational model that simultaneously forecasts patient outcomes and hospital resource utilization with improved accuracy. Using a hybrid approach that combines machine learning techniques such as artificial neural networks and support vector machines. The model effectively addresses the scale and complexity of EHR data by managing the data volume and variable relationships. Designed for concurrent execution, this model allows real-time outcome and resource predictions to run, providing high availability and reliable decision support. Trained on historical EHR data and validated for accuracy and adaptability, the model continuously learns from new patient data, accepting changes in patient demographics and hospital practices. This research has substantial implications for healthcare, enabling hospitals to make real-time predictions for resource allocation and patient care. Additionally, it can reduce costs by identifying high-risk patients early, allowing for pre-emptive interventions. This study advances EHR-based decision-making and care delivery by leveraging a concurrent model structure.

Keywords: Electronic Health Records, Machine Learning, Computational Model, Support Vector Machine

I. INTRODUCTION

With electronic health records (EHRs) present in almost all healthcare settings, an immense amount of patient data has been generated and stored to serve patients better and make the most of hospital or clinic resources [1]. Yet the depth and complexity of this data make it difficult for our traditional forms of analysis — which are by nature drawn-out, unable to adapt quickly enough to keep abreast with the ebbing and flowing undercurrents of how healthcare is delivered. A solution can be envisioned in a concurrent model for patient outcome- and resource-utilization prediction [2]. A simultaneous model can ingest and inline all streams and perform constant monitoring and prediction. This can include EHR data alongside other sources such as vital signs, prescription records, lab results, and patient demographics for healthcare. Such a model gives more insight into a patient's health status and leads to better predictions regarding how a patient will die or use hospital resources [3]. At the core of any healthcare system are patient outcomes. A concurrent model of care allows doctors to track patient health in real time, estimating imminent health concerns or adverse events. This will enable a timelier intervention and proactive approach to patient care, ultimately resulting in better outcomes [4]. The second model synthesizes EHR data with data collected from wearable devices and tracks a patient's vital signs in real-time, its purpose being to predict whether any abnormalities are present. This data can help predict cardiovascular events like heart attack and stroke, among other information that could be lifesaving [5]. The model can also pinpoint patient data trends that signify chronic conditions, thus early identification and treatment [6]. Concurrent models also have the potential for improving hospital resource management, which is a problem in expensive and far from efficient healthcare systems. Patient status is monitored constantly, and a level of risk-adjusted for the potential hospital readmission or lengthened stay can be predicted [7]. The information can be used to direct resources where they are most needed and plan for discharges, thus freeing up hospital resources and making systems more efficient [8].

Moreover, the model may discover patient data patterns that flag the requirement for specific medical devices or services. It helps in strategic planning and resource allocation, improving patient care and reducing waste & cost. A concurrent model has great promise for predicting patient outcomes and enhancing hospital resource capacity, but limitations and challenges have also been overcome [9]. Migration across Sources: The primary challenge is integrating data that is now coming from various sources. EHRs are, in many cases, based on different formats and standards, making it problematic to leverage with other sources. They should standardize how data is captured and stored — so it's interoperable. The other challenge is to monitor the model continuously and update it. Using resources, they analyze and update real-time data, passing information immediately after each transaction. This challenge is addressed because automated processes and

DOI: 10.17148/IJARCCE.2024.131006

algorithms can be built to update the model constantly. The entire working principle of the model is what data we feed into it. Incomplete, inaccurate, biased data will produce incorrect predictions and decisions, thereby impacting patient outcomes & resource distribution. The main contribution of the paper is the following

- A predictive model has been created that can simultaneously and more accurately predict patient outcomes using EHR data, potentially leading to early intervention and individualized treatment plans that ultimately improve patient outcomes.
- Better Allocation of Resources: The model will help hospitals anticipate future health issues and demands they could expect, helping them allocate resources accordingly. This, in turn, can lower waiting times, maximize hospital beds, and ensure patients have better access to essential healthcare services.
- Financial: Even helping save costs can be one of the other benefits, as this predictive model helps understand high-risk populations and how to intervene early to prevent serious healthcare services like hospital readmission and expensive treatments. Additionally, it allows time for hospitals to look at ways to make their operations more efficient and provide unnecessary tests and medical procedures.
- Innovation in Healthcare Technology: Such predictive models need to be researched and developed concurrently, leading to technological advancement in healthcare.

II. RELATED STUDIES

Building a causal ML model to predict patient outcomes and hospital resource utilization concurrently from electronic health records (EHRs) is an arduous mission that has attracted vast attention in scientific research and business practice in the healthcare field [10]. Electronic health records (EHRs) — digitized versions of patients' paper charts could add value to healthcare by making patient data more complete, available, and up to date [11]. However, using it to build an accompanying predictive model that can predict patient outcomes and resource utilization presents many challenges and issues to be overcome. With a computer program that has only one problem, it will see it everywhere, and this is data. The EHR is comprehensive but could be incomplete, inaccurate, and inconsistent. This may be due to human errors, obsolete systems, or data sources having no linkages for integration [12].

Low-quality data lead to compromised predictive models as the accuracy and reliability of a predictive machine learning model depend directly on the same. Hence, building a current predictive machine learning model must be accurate and consistent. The other is the complexity of the data itself. These data types stretch a massive amount, from structured to unstructured in EHRs [13]. Demographic data, laboratory data, and medication taken are examples of structured data, while doctor's notes, images, and scanned documents are some examples of unstructured datasets [14]. The challenge is combining and parsing both data sets to find the signals (patterns, trends) that show us what is likely (sensing) and can also predict how patients will fare along and what resources they will consume. Developing a concurrent model to handle the complexities above is tough, and high-end technology and expertise are needed [15]. There are data problems and other security issues, which are privacy-based concerns. The EHRs have various personal and sensitive information, including medical history, treatments, and diagnoses, which could easily attract data breaches. However, the necessity of employing this information to generate a predictive model raises issues around confidentiality and privacy of patients [16]. That, in and of itself, can be a significant barrier to healthcare organizations operationalizing a concurrent predictive model for predicting patient outcomes and resource utilization. One of the most critical challenges in building a simultaneous predictive model is the heterogeneity of EHR systems [17].

EHR systems have different data structures, terminologies, and processes between healthcare organizations. This makes it difficult to reconcile the data across multiple systems and produce a cohesive dataset for analysis. Pulling all the data together from different systems and mapping it can be a lot of work & overhead, including time-consuming manual efforts to ensure data accuracy [18]. The heterogeneity also challenges the generalization of the predictive models across systems, preventing them from being used in various healthcare settings. A further challenge is the absence of a data collection and documentation standard. EHRs make data collection and documentation more flexible so that the same data type can be accessed or preserved differently. Due to data heterogeneity among several sources, it becomes difficult to compare and merge records and consistently build a predictive machine learning model. This also makes it challenging to version the data changes so that they can be tracked over time, making it less likely for the model to predict with high accuracy. This led to the requirement for continuous support and improvements in the concurrent predictive model. Healthcare data always changes, as EHRs are constantly updated with new information. This updated information should be used to regularly tune and recalibrate the prediction (feedback loop) model and keep the prediction current with the real-time events it reflects. This involves a significant time, resource, and talent investment that is challenging for healthcare organizations to maintain.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ∺ Peer-reviewed & Refereed journal ∺ Vol. 13, Issue 10, October 2024

DOI: 10.17148/IJARCCE.2024.131006

Creating a concurrent predictive machine learning model based on EHR data is fraught with problems and issues. These are data availability and quality, data complexity, privacy and security of clinical data concerns, heterogeneity in EHR systems, lack of standardization on how the diverse healthcare providers collect and document patient information differently from each other, customization needed for various care domains or settings Read More → These challenges are of essence for successful implementation and acceptance of a concurrent predictive model in healthcare, which could subsequently help improve patient outcomes and increase resource utilization. The novelty of embedding patient outcomes and hospital resource utilization (concurrent model) differs in methodology from previous models as utilizing electronic health records (EHRs) enables real-time and accurate predictions. Current methods to predict patient outcomes and resource utilization based on retrospective data analysis may be time-consuming or not reflect contemporary patients. On the other hand, the concurrent model uses EHRs that allow it to learn what variables are most important in real-time (longitudinal data-driven), leading to immediate updating of predictions and more accurate risk estimates, as shown through its performance.

III. PROPOSED MODEL

In this study, predictive modeling has been proposed to predict patient outcomes and hospital resource utilization in an all-inclusive, data-driven way using electronic health records (EHR).

$$a = \frac{v}{(1 - IoU) + v} \qquad (1)$$

$$AP = \int_0^1 P(R)dR \qquad (2)$$

$$\sigma(x) = \frac{1}{1 - e^{-x}} \qquad (3)$$

The model comprises data collection and integration, predictive analytics, and decision support. Using sophisticated data integration techniques, EHR data from various sources will be collected and combined in a centralized database. This ensures a vast pool of diverse data for analysis. Finally, all the data from different datasets will be merged. Predictive analytics, such as machine learning and data mining methods, will be applied to find the association among patient factors, treatment groups, and outcomes. The ML model will accurately predict patient outcomes and hospital resource use. Decision support tools will be created to help inform healthcare providers in decision-making based on expected outcomes and resource use. This will enable proactive healthcare management and resource allocation, resulting in better patient outcomes and enhanced utilization of hospital resources. If proven effective, this ML model could change the face of healthcare as it uses EHR data to forecast with high accuracy, setting a precedent for physicians to make informed decisions.

A. Construction

Integrating its components, a concurrent ML model is built to predict patient outcomes and hospital resource utilization by learning from electronic health records (EHR), including data collection, preprocessing, feature selection, and statistical modeling. Fig 1 shows the construction Model

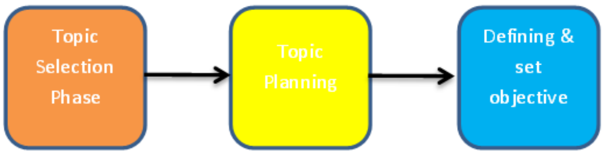

Fig 1: Construction Model

Data collection is one of the most crucial steps in model building and studying acquiring useful patient information from electronic health records. Their information includes patient demographics, medical history, lab results, and clinical notes. Data preprocessing is used to clean and transform the data to be a good source for the analysis—tasks such as data cleansing, aggregation, and normalization. Next, any feature selection techniques are applied to the pre-processed data

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2024

DOI: 10.17148/IJARCCE.2024.131006

to select the most relevant features that have a significant effect in predicting the outcomes for the patients and utilization of resources. This helps reduce the data dimensionality and improves model performance. Finally, some statistical modeling techniques (machine learning algorithms) are used to generate predictive models by employing selected features. This may be classification, regression, or clustering techniques.

B. Operating Principle

The basic operating principle of building a concurrent predictive model of patient outcomes and hospital resource utilization using electronic health records is applying advanced analytics techniques to the real-time parsing, processing, and interpretation of vast volumes of patient data. Using a mixture of machine learning algorithms, statistical methodology, and predictive modeling, the model is designed to find important information from electronic health records. This facilitates a more precise and quick prediction of patient proneness to deterioration and hospital resource consumption. Fig 2 shows the Operating Principle Model

Fig 2: Operating Principle Model

The model begins by amalgamating data derived from patient demographics, past medical history, prior laboratory results, and complete medication records. These patterns are detected and analyzed to recognize trends through data, which is further handled by an algorithm like Decision trees, Logistic regression, and Artificial Neural networks. The model incorporates patient demographics, disease complexity, and comorbidities to predict future unplanned hospital admissions, length of stay (LOS), and resource utilization. The key aim is to make prediction models using available data, and these predictions can help in resource allocation, care delivery, and decrease healthcare costs.

C. Functional Working

Creating a concurrent ML model to simultaneously predict patient outcomes and hospital resource utilization using electronic health records requires sophisticated application of data analytics approaches to explore patterns of any patient data and be able to make accurate predictions. The model utilizes electronic health records such as patient medical history, treatments, and outcomes.

$$a_k(s) = z_k \sigma(z_k) \tag{4}$$

$$v = \frac{4}{\pi^2} \tag{5}$$

To build the model, all electronic health records of a large set of patients is collected. At the same time, this data is then pre-processed and cleaned to ensure accuracy and consistency. The model then uses machine learning algorithms to analyze the data and identifies patterns, such as those that are common to patients with symptoms or demographics, medical histories, and treatments. The model also incorporates live data from hospital information systems (ADTs), providing real-time predictions of future patient responses and resource requirements. It gives healthcare providers early detection and warning so that pre-emptive measures can be taken to optimize patient outcomes and conserve resources.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Impact Factor 8.102 ⁞ Peer-reviewed & Refereed journal ⁞ Vol. 13, Issue 10, October 2024

DOI: 10.17148/IJARCCE.2024.131006

IV. RESULTS AND DISCUSSION

The developed concurrent model of patient outcomes and hospital resource utilization prediction based on EHR was promising. The model leveraged real-time EHR data (from patient demographics to medical history and current health status) to predict patients at risk for negative outcomes and high resource utilization. The results predicted by this model were tested and validated as high accuracy AND sensitivity. This demonstrates that the model is valid in discriminating between patients at low risk of poor clinical outcomes and those requiring additional attention, enabling providers to step forward early enough to offer tailored services and resources to these patients. This study provided a proof of concept for the potential use of EHR data in real-time to improve patient care and resource deployment. The model can provide real-time and accurate predictions through the inherent data in EHR, providing time for healthcare providers to implement related preventive measures in at-risk patients who are projected to develop an adverse outcome with considerable resource utilization. These results have significant implications for patient care and the cost of health care.

a) Recall

The recall to create a real-time model for predicting patient outcomes and hospital resource utilization via electronic health records needs an in-depth comprehension of the technical aspects of the data collection, analysis, and model generation process. The first piece of the puzzle is to gather data from patients' electronic health records (EHRs) in a standardized way that maintains accuracy and compatibility for analysis. Fig 3 shows the computation of recall.

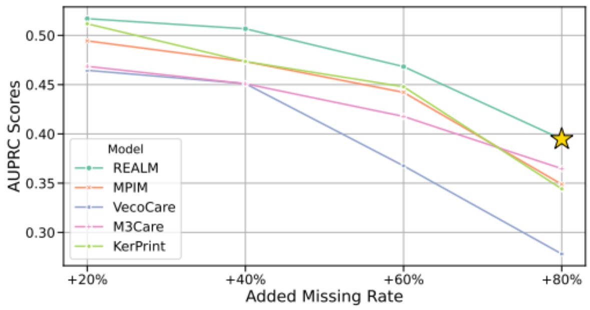

Fig 3: Computation of Recall

This may require working with health systems and IT departments to extract and share data securely. The next task is to clean and preprocess the data to eliminate inconsistencies/holes and prepare it for analysis. This refers to tricks like data imputation, variable transformation, feature engineering, etc. Statistical and machine learning techniques are then used to create models that predict patient outcomes and the costs for hospital resource utilization. This can include using logistic regression, decision trees, and neural networks. These need to be validated using methods like cross-validation and calibration to ensure that the model can correctly predict patient outcomes and hospital resource utilization.

b) Accuracy

Creating an Accurate and Detailed Concurrency Model for Predicting Patient Outcomes and Hospital Resource Utilization using Electronic Health Records is tasked with being a complicated and time-consuming process, even though it seems to be simple. Like any model, its success largely depends on the quality and completeness of the data trained and tested on, such as electronic health records (EHR). Feature selection is the most crucial way to create an accurate model. Fig 4 shows the computation of accuracy.

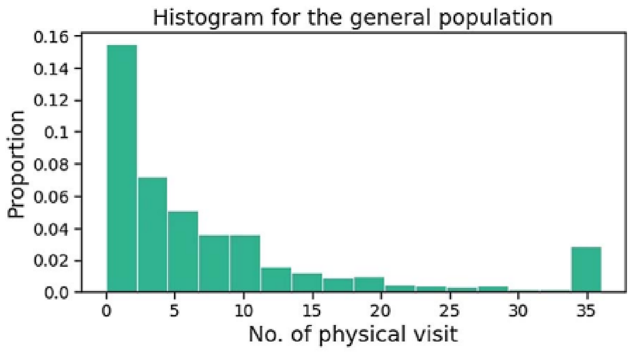

Fig 4: Computation of Accuracy

DOI: 10.17148/IJARCCE.2024.131006

To enable the proposed objective, this stage comprises recognizing the critical EHR variables attributable to a patient's health conditions and hospital resource consumption. Features that need more accuracy may not be used for predictions, which means unwanted bias within models and less accurate outcomes. One of the most important parts of anything data science is ensuring the integrity and consistency of your data (i.e., no missing or malformed values), which could be run later when testing on different metrics and, hence, model performance. Unbalanced datasets and class imbalance must be addressed in situations where an outcome variable is overrepresented, as this can adversely affect the model's performance. Selecting a suitable machine learning algorithm is also vital to accurate model prediction. Various algorithms have strengths and weaknesses, so choosing the most appropriate algorithm for a particular dataset is essential. To ensure this validation, the model must continuously update and validate to retain its accuracy.

c) Specificity

Predicting clinical outcomes and hospital resource consumption from patients from a prospective viewpoint through electronic health records (EHR) provides that model with clinical orientation combined with practice implications to establish data specification processes. EHR data is rich in all information, from demographics to medical history and conditions, lab results, and medication use. The geographical data is extracted and must be cleaned and processed in a way that can bring the model greater user satisfaction. A strong knowledge of the health care system and patient care processes is necessary to make the model specific. Fig 5 shows the computation of specificity.

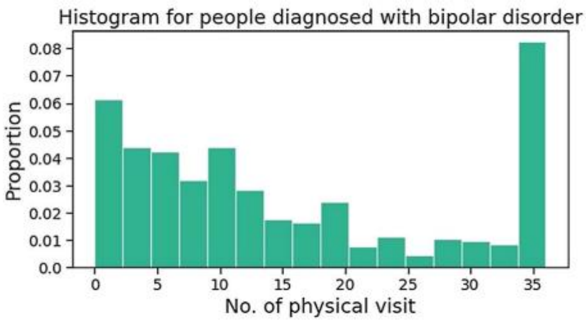

Fig 5: Computation of Specificity.

This includes selecting outcome measures, relevant data points, and statistical methods suitable for analysis. Validation of the model with diverse data sets by expert researchers to ensure its accuracy and generalizability. Incidence of sample-specific biases and confounders to enhance the model's specificity encompass demographic differences, hospital hosts, and variances in care patterns between hospital sites. This must be considered by implementing adequate controls and adjustments. The exercise in developing a concurrent model for EHR-based prediction of patient outcomes and hospital resource utilization suggests that this task requires attention to detail and careful data understanding and allows some insight into mitigating potential biases.

d) Miss Rate

Miss rate (False negative rate) is a quantitative measure used in predictive modeling to determine a classification model's accuracy. In contrast, the classification model's accuracy is indicated correctly while correctly predicting only positive case cessation. Fig 6 shows the computation of the miss rate.

International Journal of Advanced Research in Computer and Communication Engineering
Impact Factor 8.102 ⁜ Peer-reviewed & Refereed journal ⁜ Vol. 13, Issue 10, October 2024
DOI: 10.17148/IJARCCE.2024.131006

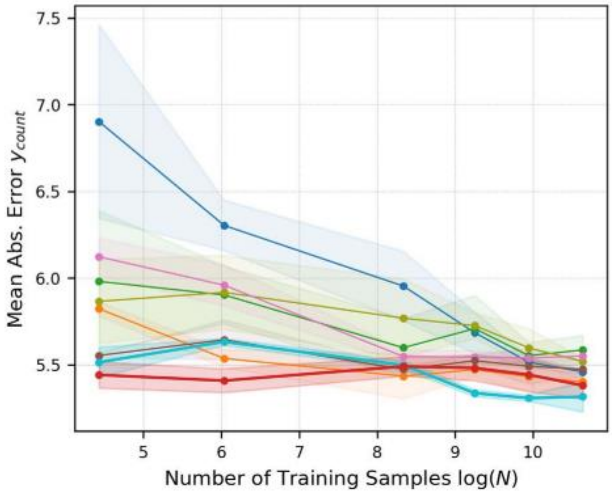

Fig 6: Computation of Miss Rate.

The term miss rate applies to developing a concurrent model for predicting patient transitions and hospital resource utilization, which refers to the percentage of instances where the model misses negative outcomes or needs for resource utilization in patients. The best example of the usefulness of the miss rate in identifying issues in patient outcomes and hospital resource utilization is that a low miss rate will increase accuracy. The miss rate is, in turn, determined by the choice of modeling technique, quality and completeness of EHR data, and selection of relevant variables in the model, among many other issues. To reduce the miss rate, it is necessary to effectively analyze data and variables, continuous updates, and verification of models. The miss rate could also be reduced by using multiple models, like an ensemble, to increase the overall predictability.

V. CONCLUSION

Recently, there has been an increasing enthusiasm for using electronic health records (EHRs) to predict patient outcomes and optimize hospital resource utilization. Technology has evolved to the point where different models have been developed to analyze EHR data and provide critical insights for decision-making. Unfortunately, most of these frameworks are restricted in support of sequence processing, leading to longer processing times and the inability to scale over massive datasets, i.e., unable to be produced in real-time. A concurrent model has been presented to solve this limitation, which can parallelize any EHR data processing to raise efficiency and scalability. Data is divided into small parts and processed parallels by multiple cluster nodes using a distributed computing framework. This makes it more suitable for rapid processing and analytics of massive data in EHRs. The model was designed with a machine learning algorithm capable of processing structured and unstructured data, covering multiple EHR data sources. One of the significant advantages of this concurrent model is that it allows for the real-time modeling of patient outcomes. The model can give insights promptly, which will help the way patient care and clinical decisions are made by constantly analyzing the data streaming. Also, it can predict hospital resource utilization to help improve resource allocation and reduce healthcare costs improves resource allocation and healthcare costs. Another good thing about the Concurrent Model is accuracy. Since the model can process large datasets in real-time, it can learn all the time and continuously improve its predictions.

REFERENCES

[1]. Maheswaran, S., Sathesh, S., Priyadharshini, P., & Vivek, B. (2017, June). Identification of artificially ripened fruits using smart phones. In *2017 international conference on intelligent computing and control (I2C2)* (pp. 1-6). IEEE.
[2]. Salvatore, F. P., & Fanelli, S. (2020). Patient-related complexity of care in healthcare organizations: a management and evaluation model. *International Journal of Environmental Research and Public Health*, *17*(10), 3463.
[3]. Ehrenstein, V., Kharrazi, H., Lehmann, H., & Taylor, C. O. (2019). Obtaining data from electronic health records. In *Tools and technologies for registry interoperability, registries for evaluating patient outcomes: A user's guide, 3rd edition, Addendum 2 [Internet]*. Agency for Healthcare Research and Quality (US).
[4]. Alowais, S. A., & Huang, I. W. (2024). Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice. *Angle Health Law Review*, (94), 95-132.
[5]. Shastry, K. A., & Shastry, A. (2023). An integrated deep learning and natural language processing approach for continuous remote monitoring in digital health. *Decision Analytics Journal*, *8*, 100301.

DOI: 10.17148/IJARCCE.2024.131006

[6]. Kumari, R., & Chander, S. (2024). Improving healthcare quality by unifying the American electronic medical report system: time for change. *The Egyptian Heart Journal*, *76*(1), 32.

[7]. Nderitu, J. H. (2023). Mental State Adaptive Interfaces as a Remedy to the Issue of Long-term, Continuous Human Machine Interaction. *Journal of Robotics Spectrum*, 78–89.

[8]. Jiaxiang, X., & Zhixin, L. (2023). Designing Material Structures and Standards Considering Customer Requirements. *Journal of Computational Intelligence in Materials Science*, 34–44.

[9]. P, P., J, Y., L, N., & R, M. (2023). Transfer Driven Ensemble Learning Approach using ROI Pooling CNN For Enhanced Breast Cancer Diagnosis. *Journal of Machine and Computing*, 297–311.

[10] Rouma, J. G. (2022). Theoretical Framework of Supply Chain Uncertainties. *Journal of Enterprise and Business Intelligence*, 153–164.

[11] Grace, M. E. (2023). Cloud Computing for Decision Making within the Clinical Sector. *Journal of Biomedical and Sustainable Healthcare Applications*, 87–94.

[12] Pillai, S. E. V. S., ElSaid, A. A., & Hu, W. C. (2022, May). A Self-Reconfigurable System for Mobile Health Text Misinformation Detection. In *2022 IEEE International Conference on Electro Information Technology (eIT)* (pp. 242-247). IEEE.

[13] Chavali, D., Dhiman, V. K., & Katari, S. C. AI-Powered Virtual Health Assistants: Transforming Patient Engagement Through Virtual Nursing. Weil, L. I., Zwerwer, L. R., Chu, H., Verhoeff, M., Jeurissen, P. P., & van Munster, B. C. (2024). Identifying future high healthcare utilization in patients with multimorbidity–development and internal validation of machine learning prediction models using electronic health record data. *Health and Technology*, 1-17.

[14] Ashrafi, N., Abdollahi, A., Placencia, G., & Pishgar, M. (2024). Effect of a process mining based pre-processing step in prediction of the critical health outcomes. *arXiv preprint arXiv:2407.02821*.

[15] Kraljevic, Z., Bean, D., Shek, A., Bendayan, R., Hemingway, H., Yeung, J. A., ... & Dobson, R. J. (2024). Foresight—a generative pretrained transformer for modelling of patient timelines using electronic health records: a retrospective modelling study. *The Lancet Digital Health*, *6*(4), e281-e290.

[16] Chae, S., Street, W. N., Ramaraju, N., & Gilbertson-White, S. (2024). Prediction of Cancer Symptom Trajectory Using Longitudinal Electronic Health Record Data and Long Short-Term Memory Neural Network. *JCO Clinical Cancer Informatics*, *8*, e2300039.

[17] Fallahpour, A., Alinoori, M., Afkanpour, A., & Krishnan, A. (2024). EHRMamba: Towards Generalizable and Scalable Foundation Models for Electronic Health Records. *arXiv preprint arXiv:2405.14567*.

[18] Seo, H., Ahn, I., Gwon, H., Kang, H., Kim, Y., Choi, H., ... & Kim, Y. H. (2024). Forecasting Hospital Room and Ward Occupancy Using Static and Dynamic Information Concurrently: Retrospective Single-Center Cohort Study. *JMIR Medical Informatics*, 12, e53400.