International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

A Multi-Label Text Classification with In-Ordinated Multi Classes for Predicting Suitable Practitioner Recommender Systems using Supervised Machine Learning Systems with I- FastText Method

G Venkata Ramana Reddy[1], Dr. T. Bhaskara Reddy[2]
[1]Research Scholar, Sri Krishnadevaraya University, Ananthapuramu, A.P, India
Department of CSE, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur Dist., A.P., India
[2]Professor, Sri Krishnadevaraya University, Ananthapuramu, A.P, India

Abstract
Practitioner Recommendation Systems (RS) addresses the issue of medical doctor referrals. However, these systems require access to patient feedback and medical records which may not always be available in real-world scenarios. This research focus on medical referrals and aims to predict recommendations in different specialties of physicians for both new patients and those with a consultation history. The traditional or existing mechanisms use limited Multilabel and multi classification employed for text-based classification tasks without any encoding features and cannot track the real time scenarios for predictions if any. Because of this it has become very much challenging for existing systems like XML methods are used through Predictive Model Mark Language (PMML) for multi label and multi class classification for recommendation of specialist doctor. If too many labels need to be classified with millions of classes or labels the traditional or the existing Extreme Multi-Label Classification (XML) methods may not be sufficient to handle. In this regard an updated method in machine learning I-FastText and I-BERT Methods are improvised to overcome the above said problem. As these methods provide a means for providing multi-labeling to multi classes, the task of handling the patient datacites from various sources of real time scenarios can be handled effectively. I-FastText classify the multi class text from different sources of patient datacites and I-BERT will find the relevant medical practitioners based on the query from large set of patient records. The information retrieval engine retrieves the suitable practitioner based on matching of the indexed and ranked documents.
The proposed system enables the early identification of the suitable doctor relevant to the patient datacites which enables the early start treatment of patients and early recovery. The system ensures minimal false predictions, maximizes the prediction of the specific doctor for appropriate recommendation to the patient and ensures maximal performance.
Keywords: Multilabel Classification, I-FastText, I-BERT, Recommender Systems, Collaborative Filtering, SVD, Pattern Processor(LightFM), XML methods.

1. Introduction

Selecting a suitable doctor for patients significantly affects patient health outcomes. During the referral process, a primary care physician (PC) should recommend and an appropriate physician according to the individual patient needs. Also, the system also enables the general medical practitioner in identifying the suitable practitioner inspite of limited consultation time, partial knowledge of all matching physicians, infrequent previous patient contact, and potential bias from the doctor’s social network.
Therefore, an additional tool is needed to facilitate the patient-centered recommendation of specialist physicians. Recommender systems make decision-making easier by suggesting relevant choices. Patients can use them to find doctors who fit their needs by checking previous patients’ experiences [1,2]. In this paper we recast a traditional recommender setting into a multilabel classification [3] problem that can be solved by current extreme classification methods. Also, we propose a unified model leveraging patient history across different specialties. Recommender

DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

systems (RS) [4] have been widely employed to facilitate these decision-making processes. These systems fall into three categories like Collaborative Filtering (CF) [5], Content-based (CB), and hybrid approaches. CF assumes that users who have rated items similarly in the past will continue to do so in the future. However, CF requires a large amount of data [6] and suffers from the "cold-start" problem — a challenge in RS where the system struggles to make accurate recommendations for new users due to a lack of historical interactions. This is particularly critical in healthcare and as inaccurate recommendations could harm patient care. On the other hand, CB systems recommend items similar to those the user has preferred in the past without relying on information about other users and yet these systems often struggle to expand users' interests. Hybrid approaches combine the strategies used in both CB and CF to take advantage of their different strengths. In the context of medical expert recommendations, specific requirements differentiate this field from other domains that employ recommender systems. Patients usually interact with significantly smaller set of possible physicians, unlike traditional settings with a large pool of interactions from which to learn. Therefore, while CF methods perform well in other domains, most healthcare methods are CB and Han. Furthermore, explainability is vital in healthcare applications since doctors and patients are less likely to trust black-box recommendations. This extends beyond user trust and the understanding the factors which contribute for recommendation could potentially provide the solution. The limited availability of patient metadata where privacy regulations presents unique challenges in developing effective health care recommending system.

The system is designed to tackle the cumbersome problem and handle limited patient metadata while respecting privacy concerns which are inherent in healthcare. Based on the constraints a multi-label classification based on XML methods have been implicated and metadata oriented health care recommendation is made. The decision making regarding the healthcare recommendation is done through decision making tree by implementing supervised machine learning model. The proposed methodology predicts a specialized doctor for a patient based on diagnosis which will lead to better results. Different instances of the patient have been considered where label space, instance and feature space has been considered as parental data sets. Each of these instances have been classified into different subsets of the instances and each of the subset is tagged with the doctor's(specialist) network. At this point I- FastText and I- BERT methods have been improvised to set up the meta data where multi label classification is done to predict the results of the patients based on the recommendation of the system. In this context it centralizes the different features based on the XML [7] applications where the actual data sets are defined and implemented. The repository contains the data sourced from different relevant resources of medical practitioners which overcomes the limitations of the traditional RS in comparison with state-of-the-art methods.

The researcher improvises methods like I- FastText and I- BERT methods and improves the response time of the doctor's(specialist) network based on the patient's features and there by patient may approach the doctor for better treatment with in the limited time.

2. Existing system

The existing module is based on web mining have some traditional drawbacks in finding appropriate doctor's specialization based on patient's history. The system is based on the data that is stored in computer database based on the patient's requirements and doctor's profile. The selection of the doctor or the recommendation to a doctor based on the features of the doctor with respect to the qualification. The system collects different sentiments of the patient and the feedback is collected over a period of time and based on which the data will be extracted and stored through multiple classes and classifications shown in Figure 1. This classification of the data is done based on patient's information where a system is built for relevant recommendation after proper training of the data set. This will really has given better idea to build a better health recommendation system.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

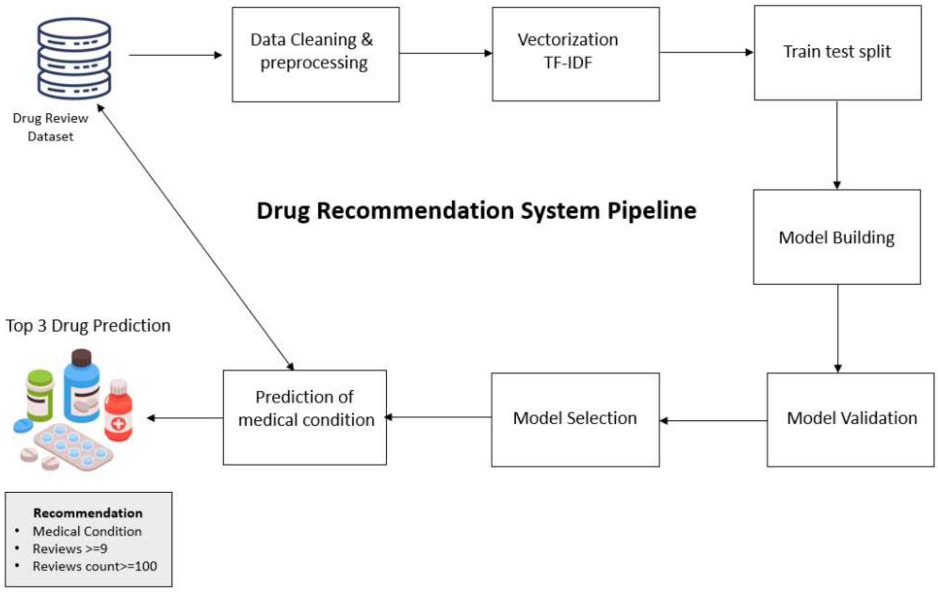

Fig.1: The health care recommendation system with feature extraction and with traditional approach

3. Proposed system:

The heuristics of patient data with doctor's history will allow the user a decision making very fast and reliable. This will rely on the cognitive aspects of the system and patient's record based on the scenario. This will really address the dual theory while decision making [8] with minimal vulnerability to the system. When a patient approach with a disease the system has to identify the nature of the disease and the symptoms of the disease. In this regard a kind of causal analysis has to be made to determine the exact cause of the disease along with remedy for that. In this regard classification of diseases, patients and doctors have to be personified to determine the exactness of the disease in recommending the patient to relevant doctor. Based on which doctor with specific specialization will be identified and realized after thorough diagnosis of disease. Classification of the patient has been done using Single Value Decomposition(SVD) [9] technique which implement the Principal Component Analysis(PCA). Feature selection will be made through dimensionality reduction of the patient object. An object having two or more features relevant to other class of data set will be encoded through Bi-Lateral Variational Encoder(BVE) and later on decoded to avoid the duplication among the feature data sets that are defined. For this an activation feature component is instantiated to recommend a doctor after classification. Supervised machine learning model [10] is implemented for better classification and labeling of the data is concerned. Once the patient approaches, the recommended model Activation Feature Derivative Model(AFDM) will realize the features of the data set and classify the disease and based on the bilateral variational encoding the actual disease and severity of the disease will be realized and analyzed. I-FastText is implemented for better text classification which obtain vector representations of the text or words. It allows the users for learning the text representation and text classification. This will help the system to classify the doctors and patients with regard to their nature of the responsibilities. Bidirectional Encoder Representations from Transformers(BERT) is used to understand the contextual domain. Both these methodologies where the data has been trained and classified for better recommendation of the doctor based on the patient's disease.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

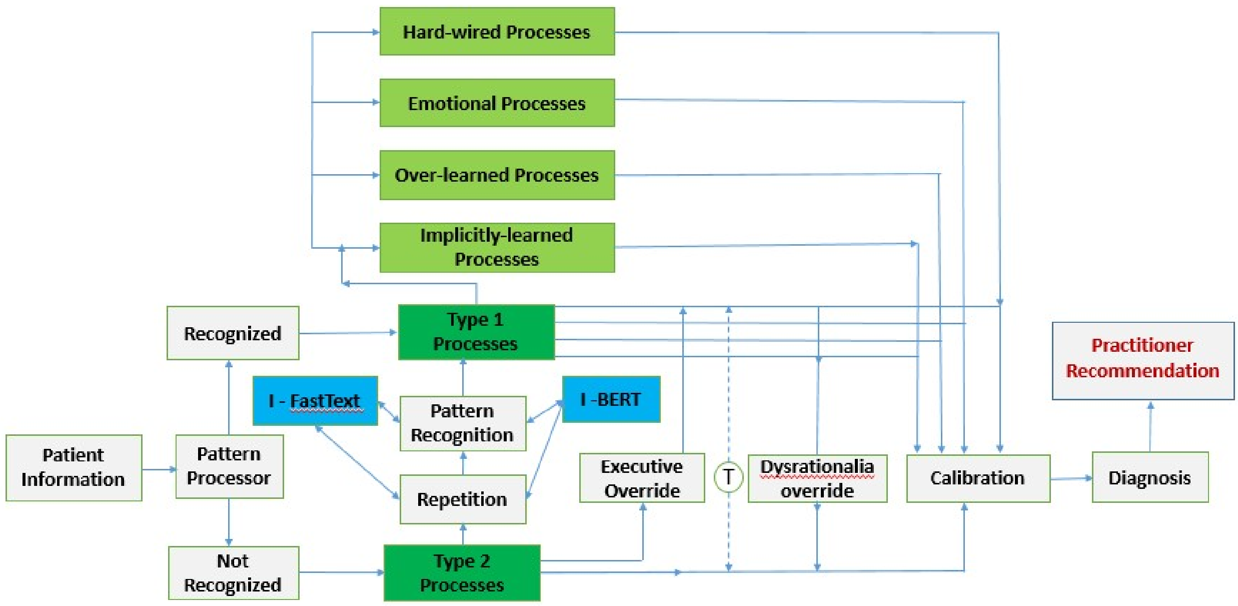

Fig.2: Architecture of proposed system with I-FastText and I-BERT

It is difficult to identify different patterns of the data which may lead to conflicts among classifiers in the data set. In Figure 2 a Type 1 and Type 2 processes are implemented based on the context of the transactional management. If the data is properly recognized, then a repetition of the data processing is done till the data is free from noise till the requirements are satisfied. If the data has got deviations if any a state is defined where some exception handling mechanisms are implemented till actual data is realized for better classification and labeling. A systematic review focusing on the medical profession showed that most studies found healthcare where I-FastText is used for textual data analysis [11] where it consists of library for text classification and word representation, is effective in capturing semantic relationships within medical texts. FastText can be used to classify medical texts, such as patient records, and historical data. After the data is classified successfully the data is sent to pattern processor [12] for matching criteria of the data text. I-BERT will encode the data and follow through transactional transformation of the text for better calibration of the data where precision of the doctor is made. This will be done based on a mechanism called Single Valued Decomposition(SVD) where the set of features will be decomposed and a unique feature is selected for better processing of the doctor's text object where a feedback mechanism called LightFM is used through Supervised Machine Leaning(SML). Under the limited feedback, sparse metadata, and the need for cross-specialty predictions have been made through feature selection of the patient and doctor specialization. The requirement to make predictions across different specialties [13] and for new users, as well as the limited feedback and metadata suggesting that I-FastText and I-BERT has the potential over the existing and traditional recommending system.

4. Dataset and Problem Setting

The data set forms a basic and predominant aspect feature for both doctor and patient records. Here the patient object is determined through P and L being the number of doctors. There will be interactions in the system where each interaction involves a unique patient *p* and a doctor *l*. This is based on the definite time and location that is assigned. The data set that is considered associated with various hospitals with respect to time and their locations. The temporal information [14] has not been addressed in the existing systems. But the proposed system has incorporated both Temporal and Ariel data in order to locate the doctor as well as the hospital he is working under.

DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

Every patient has to take the appointment before system constraints with respect to his credentials. The appoint time will play a crucial role in addressing better doctor based on the disease of the patient.

5. Recommendation Approach:

Let patient has made number of interactions with the doctor where LightFM will come to existence to provide the feedback mechanism for different doctors based on which a recommendation is made to the patient. The feedback mechanism [15] is properly trained based on the input data that is defined and classified through supervised machine learning model, However, when selecting a doctor, each patient does not consider all available doctors but all doctors of the specialty in need. Thus, we build the rating matrix R ∈ R P ×L with entries

$$R_{pl} = \frac{n_{ps}}{n_{pl}}$$

where n_{pl} is the number of visits of patient *p* to doctor *l* and n_{ps} is the number of times patient *p* visited doctors of specialty *s*, where *s* is the specialty of doctor *l*. Note that the absence of interactions is considered negative feedback, a usual assumption when handling implicit feedback. Traditional RS learn preferences from matrix R or directly from the interactions, if they accommodate implicit feedback.

XML Classification for Specialist Doctor Recommendation:

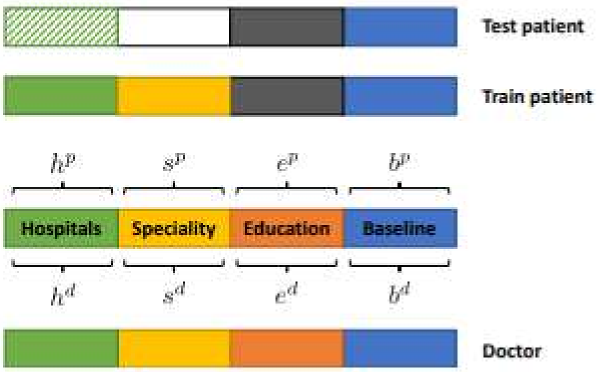

Fig.3: Four distinctive groups of consultations based on pre-defined data.

In Figure 3 Four distinctive groups of consultation of the pre-defined data is made and different patterns of the doctors and patients are realized. The following table and the above diagram will reflect the same. The table represent the number of visits that a patient is made based on the feature set [16]. The feature set is unique in nature processed through single valued decomposition method.

DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

	S-1		S-2		S-3		S-4		S-5	
	P	**D**	**P**	**D**	**P**	**D**	**P**	**D**	**P**	**D**
Baseline	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Specialization			✓	✓					✓	✓
Education		✓		✓		✓		✓		✓
Hospitals (visits)					✓	✓		✓	✓	✓
Hospitals (distances)							✓		✓	

Table1: Comparison of the data availability scenarios with respect to patients(P) and doctors(D).

	Train	Test seen	Test new
Interactions	314892	73730	12091
Patients	80960	32213	12498
Doctors	1044	992	699

Table 2: Patients and Doctors Datasets with characteristics.

In Table 2, the first column refers to the training dataset, the second to the seen patients in the test dataset, and the third to new patients in the test set.

6. Experimental Results:

Performance of methods I-FastText and I-BERT methods for recasting a traditional recommender[17] setting into a multilabel classification problem to be addressed by extreme classification. We show empirically the superiority of unified model [18] leveraging patient history across different specialties, with respect to SOTA baselines. We aim to determine which scenario (combination of features) is most suitable and how they impact XML predictions [19]. Additionally, we compare I-FastText and I-BERT with traditional XML recommenders explore the difference between observed and new patients, and analyze performance across different specializations by communicating either through teleconsultation [20] or feedback consultation.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

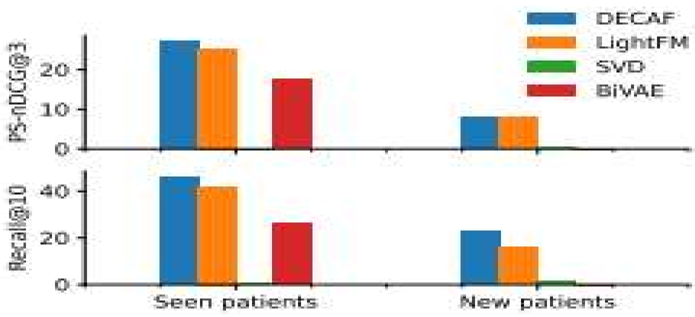

Fig. 4: Classification of Specialist Doctor Recommendation based on DECAF and LightFM.
* A data factorization is considered almost negligible factor, i.e. slope zero in SVD

As shown in Figure 4, DECAF (Definite Causal Activation Feature) will identify the required feature on patient and doctor data sets for better recommendation through LightFM(Hybrid recommender model)

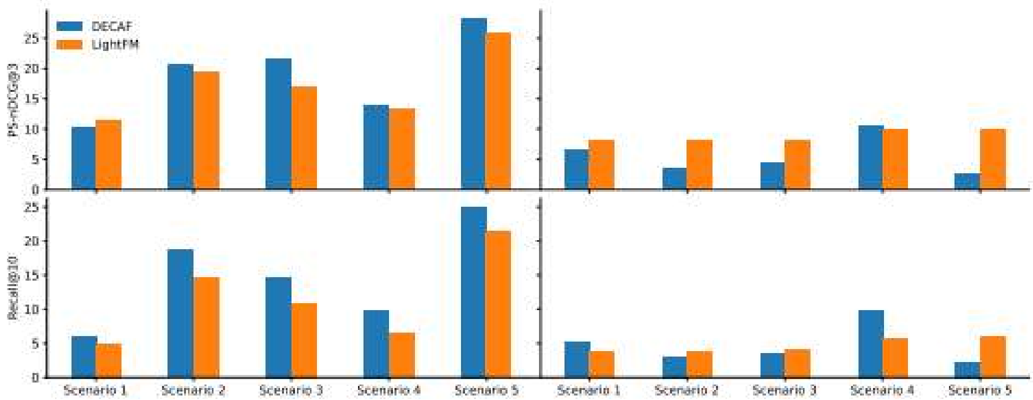

Fig. 5: Scenarios of Patients and Doctors interactions during different intervals of time.

LightFM for seen patients (left) and new patients (right). For each specialty (x-axis), we depict the score for each different scenario for LightFM in solid colors, and for XML in the same-color hatched bars presented in Figure 5. The absence of a bar indicates that XML does not provide at least 3 predictions for enough patients to have significant results. PS-nDCG(Probabilistic-Normalized Discounted Cumulative Gain (nDCG)) by incorporating probability estimates from a machine learning classifier to refine the ranking and discount functions.

CONCLUSION:

This study tackles the physician referral problem in healthcare environments where explicit feedback is sparse and patient metadata is limited, conditions that significantly constrain the effectiveness of traditional recommender systems. The presence of multiple overlapping medical specializations further exacerbates the problem by introducing high label cardinality, redundancy, and severe class imbalance. To address these challenges, we reformulate the referral task as an Extreme Multilabel Classification (XML) problem, enabling scalable learning over a large and complex label space. The proposed models, I-FASTTEXT and I-BERT, leverage textual representations of longitudinal patient histories across diverse specialties to capture nuanced clinical patterns and inter-specialty dependencies. This formulation allows the models to jointly predict multiple relevant physician specializations for a given patient context, rather than relying solely on user–item interaction matrices. Unlike conventional recommender systems, the XML-based

DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

approach is inherently suited to handling inordinate multi-class, multi-label scenarios with sparse observations. Experimental results, benchmarked against implicit and explicit feedback-based methods such as DECAF and LightFM, demonstrate that the proposed models consistently outperform baseline recommender approaches in top-□ recommendation accuracy and relevance. These findings highlight the effectiveness of extreme multilabel text classification in generating clinically meaningful physician referrals and establish the proposed framework as a robust alternative for complex healthcare recommendation tasks

References:

1. J. J. LIU, J. J. MATELSKI and C. M. BELL, Scope, breadth, and differences in online physician ratings related to geography, specialty, and year: observational retrospective study, Journal of Medical Internet Research 20(3), 2018, pp. 1–18
2. M. CHE, L. ZHAO and B. JIN, A doctor recommendation framework for online medical platforms using multi-source heterogeneous data, Proceedings of 7th International Conference on Computing and Artificial Intelligence, Tianjin, China, 2021, pp. 326–331.
3. Shakir Farhan, KanaparthiNayan, NancherlaKoushik, Ms. D. Swetha, BERT-Based Emotion Detection and Real-Time Content Recommendation System for Digital Mental Health Support, Interantional Journal of Scientific Research in Engineering and Management, 09(02):1-6, 2024, DOI: 10.55041/IJSREM38647.
4. Himanshu Singh, Moirangthem, Biken Singh, Ranju Sharma, Jayesh Gat, Ayush Kumar Agrawal, “Optimized Doctor Recommendation System using Supervised Machine Learning”, ICDCN '23: Proceedings of the 24th International Conference on Distributed Computing and Networking, 2023,Pages 360 – 365. https://doi.org/10.1145/3571306.3571372.
5. A. Sae-Ang, S. Chairat, N. Tansuebchueasai, O. Fumaneeshoat, T. Ingviya and S. Chaichulee, "Drug recommendation from diagnosis codes: classification vs. collaborative filtering approaches", International Journal of Environmental Research and Public Health, vol. 20, no. 1, pp. 309, 2023.
6. N. Kumar, K.H.S. Sai, V. Hordiichuk, R. Menon, C.J. Aarthy, G.C. Saha and K. Balaji, "Harnessing the power of big data: challenges and opportunities in analytics", TuijinJishu/Journal of Propulsion Technology, vol. 44, no. 2, pp. 363-371, 2023. DOI: https://doi.org/10.52783/tjjpt.v44.i2.193
7. P. Keikhosrokiani and G.M. Fye, "A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings", Multimedia Tools and Applications, vol. 83, pp. 45315–45344, 2023.
8. Y.H. Chang, Y.T. Guo, L.C. Fu, M.J. Chiu, H.M. Chiu and H.J. Lin, "Interactive healthcare robot using attention-based question-answer retrieval and medical entity extraction models", IEEE Journal of Biomedical and Health Informatics, vol. 27, no. 12, pp. 6039-6050, 2023.
9. A.D. Loiotile, D. Veneto, A. Agrimi, G. Semeraro and N. Amoroso, "An AI-based approach for the improvement of university technology transfer processes in healthcare, information systems and technologies," Lecture Notes in Networks and Systems, Springer, Cham, vol. 802, pp 311–320, 2023.
10. S.R. Sripathi, N.V.S. Pradyumna, A. Dhanush and R. Subramani, "Drug recommendation system using LDA", International Conference on Futuristic Technologies, pp. 1-7, 2022.
11. Sudhanshu, N.S. Punn, S.K. Sonbhadra and S Agarwal, "Recommending best course of treatment based on similarities of prognostic markers", International Conference on Neural Information Processing, Springer International Publishing, pp. 393-404, 2021.
12. Massaro, M. Digital transformation in the healthcare sector through blockchain technology. Insights from academic research and business developments. Technovation 2021, 2021, 102386. [Google Scholar] [CrossRef]
13. Kraus, S.; Schiavone, F.; Pluzhnikova, A.; Invernizzi, A.C. Digital transformation in healthcare: Analyzing the current state-of-research. J. Bus. Res. 2021, 123, 557–567. [Google Scholar] [CrossRef]

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.026 ISSN: 2581-4621

14. Shahed, K.S.; Azeem, A.; Ali, S.M.; Moktadir, M. A supply chain disruption risk mitigation model to manage COVID-19 pandemic risk. Environ. Sci. Pollut. Res. 2021, 1–16. [Google Scholar]
15. Pinar, U.; Anract, J.; Perrot, O.; Tabourin, T.; Chartier-Kastler, E.; Parra, J.; Vaessen, C.; de La Taille, A.; Roupret, M. Preliminary assessment of patient and physician satisfaction with the use of teleconsultation in urology during the COVID-19 pandemic. World J. Urol. 2021, 39, 1991–1996. [Google Scholar] [CrossRef]
16. Lancet, T. COVID-19: Learning from experience. Lancet 2020, 395, 1011. [Google Scholar] [CrossRef]
17. Moazzami, B.; Razavi-Khorasani, N.; Moghadam, A.D.; Farokhi, E.; Rezaei, N. COVID-19 and telemedicine: Immediate action required for maintaining healthcare providers well-being. J. Clin. Virol. 2020, 126, 104345. [Google Scholar] [CrossRef]
18. Katz, J.N.; Sinha, S.S.; Alviar, C.L.; Dudzinski, D.M.; Gage, A.; Brusca, S.B.; Flanagan, M.C.; Welch, T.; Geller, B.J.; Miller, P.E.; et al. COVID-19 and disruptive modifications to cardiac critical care delivery: JACC review topic of the week. J. Am. Coll. Cardiol. 2020, 76, 72–84. [Google Scholar] [CrossRef]
19. Sousa, M.J.; Pesqueira, A.M.; Lemos, C.; Sousa, M.; Rocha, Á. Decision-making based on big data analytics for people management in healthcare organizations. J. Med. Syst. 2019, 43, 290. [Google Scholar] [CrossRef]
20. Banerjee, M.; Chakraborty, S.; Pal, R. Teleconsultation and diabetes care amid COVID-19 pandemic in India: Scopes and challenges. J. Diabetes Sci. Technol. 2020, 14, 714–715. [Google Scholar] [CrossRef]