60th International Scientific Conference on Information, Communication and Energy Systems and Technologies, Ohrid, North Macedonia, June 26–28, 2025

Transforming health data into knowledge predictive model for patient health risk

Natasha Blazheska-Tabakovska[1], Snezana Savoska[2] and Mimoza Bogdanoska Jovanovska[3]

Abstract – **The rapid digital transformation of the healthcare sector has opened new avenues for patient-centred care. This paper proposes a data-driven healthcare framework using Personal Health Records (PHRs), collaborative filtering, and predictive modelling. By transforming diverse patient data into structured vectors and calculating similarities, the system enables personalised recommendations, early disease detection, and risk assessment, contributing to more efficient, proactive, and patient-centred healthcare delivery.**

Keywords – **Personal Health Records (PHR), Predictive Model, Collaborative Filtering, Healthcare Recommendation Systems, Digital Health Analytics.**

I. Introduction

The healthcare sector is a vital and complex domain that plays a crucial role in improving the quality of life and overall well-being of individuals. In recent years, digitalisation has transformed the healthcare industry, introducing innovative solutions that enhance efficiency, accuracy, and accessibility. The integration of information and communication technology (ICT) and the Internet of Medical Things (IoMT) has led to a paradigm shift in healthcare delivery, enabling real-time monitoring, remote diagnostics, and enhanced patient engagement. Designing the hospital of the future requires seamlessly integrating advanced hospital services with cutting-edge technologies [1]. A smart and dynamic healthcare ecosystem enhances both clinical and administrative processes, optimising workflows, improving decision-making, and ultimately elevating patient care [2]. By fostering efficiency and collaboration, this ecosystem not only enhances the working environment for healthcare professionals but also significantly contributes to patient health and well-being. The European Commission's eHealth Action Plan 2012-2020 highlights the significance of digital health technologies in fostering citizen-centric healthcare [3]. By promoting eHealth solutions, this initiative aims to improve healthcare accessibility, efficiency, and quality while reducing hospitalisation times and medical errors [2]. Additionally, the World Health Organisation (WHO) underscores the importance of eHealth in its efforts to enhance health systems through the effective use of ICT. WHO emphasises improving healthcare for everyone, everywhere by accelerating the development and adoption of appropriate, accessible, affordable, scalable and sustainable person-centric digital health solutions to prevent, detect and respond to epidemics and pandemics, developing infrastructure and applications that enable countries to use health data to promote health and well-being WHO defines eHealth as the secure and cost-effective utilization of digital technologies for healthcare, encompassing electronic health records (EHRs), telemedicine, mobile health (mHealth), and digital health surveillance systems [4].

A key component of digital healthcare transformation is the adoption of Electronic Health Records (EHRs) and Personal Health Records (PHRs) [5]. The EHRs, also called electronic medical records, refer to a structure in digital format of patients' health data that is maintained throughout their lives and is stored accurately in a repository [6]. EHRs provide a structured digital repository of patient health data, including medical history, diagnoses, medications, and laboratory results, ensuring comprehensive and accessible patient information. Electronic Health Records (EHRs) originate from diverse sources, including healthcare providers, prescription data, laboratory results, bio-monitoring data, and referral records. They can also be collected directly from sensors measuring vital signs or non-clinical sources such as exercise habits, dietary patterns, screening results, and exposome data [7]. The widespread use of EHRs, wearable devices, and other digital health technologies has resulted in an unprecedented accumulation of healthcare data. The Internet of Medical Things (IoMT) infrastructure plays a vital role in generating and integrating healthcare data by connecting medical devices and software applications that seamlessly interact with Personal Health Records (PHR) and other healthcare systems [8]. The data generated from different sources can be structured, semi-structured or unstructured. This vast amount of information, generated from diverse sources such as clinical notes, lab tests, sensors, and social media, presents both opportunities and challenges in healthcare data management.

Data analytics has emerged as a critical component in revolutionising healthcare delivery by harnessing the power of vast amounts of data generated within the healthcare ecosystem [9]. Effective analysis of healthcare data is essential for deriving meaningful insights that support decision-making, improve patient care, and optimise healthcare operations. With the application of advanced analytics techniques such as predictive modelling, machine learning, and artificial intelligence, healthcare providers can identify trends, correlations, and potential health risks. These innovations facilitate the shift from reactive to proactive healthcare models, enabling early disease detection, personalised treatment plans, and preventive interventions.

As healthcare continues to evolve in the digital age, the integration of ICT and IoMT, alongside robust data analytics

[1]Blazheska-Tabakovska Natasha, University St. Kliment Ohridski Bitola, Faculty of ICT, Partizanska bb, 7000 Bitola, R.of N.Macedonia, E-mail: natasa.tabakovska@uklo.edu.mk
[2]Savoska Snezana, University St. Kliment Ohridski Bitola, Faculty of ICT, Partizanska bb, 7000 Bitola, R.of N.Macedonia, E-mail: snezana.savoska@uklo.edu.mk
[3]Bogdanoska Jovanovska Mimoza, University St. Kliment Ohridski Bitola, Faculty of ICT, Partizanska bb, 7000 Bitola, R.of N.Macedonia:a, E-mail: mimoza.jovanovska@uklo.edu.mk

979-8-3315-2655-9/25/$31.00 ©2025 IEEE

frameworks, will play a pivotal role in shaping the future of medical services. Leveraging these technologies will not only enhance patient-centred care but also contribute to the overall efficiency and sustainability of healthcare systems.

II. Integrating Recommendation Systems and Predictive Modelling in Healthcare

The main focus of this paper is knowledge extraction from collected data originating from various sources and stored in Personal Health Records (PHR). The extracted knowledge is utilised within subsystems, particularly a recommendation system, which aids patients by enabling early disease detection, personalised treatment plans, and preventive interventions, ultimately enhancing healthcare outcomes. Recommendation systems are typically categorized into three main types: Collaborative Filtering – These systems analyze user behaviour along with patterns from users with similar interests to generate personalized recommendations; Content-Based Filtering – These systems utilize detailed item descriptions and a user's past preferences to suggest relevant options; Hybrid Filtering – These systems integrate both collaborative and content-based filtering methods, enhancing recommendation accuracy and diversity [10]. This paper addresses the research problem of identifying effective techniques for organising data for knowledge discovery. To enhance recommendation accuracy, predictive modelling plays a crucial role by leveraging historical data and advanced analytics to forecast patient outcomes. These models identify risk factors, stratify patient populations, and guide clinical decision-making, improving care and optimising resource allocation [11]. Common applications include disease prediction, hospital readmission prevention, chronic disease management, and treatment optimisation [12]. With advancements in artificial intelligence and machine learning, integrating predictive modelling with recommendation systems enhances healthcare efficiency, improves patient outcomes, and supports data-driven clinical decision-making.

Predictive modelling follows a structured process that includes data preprocessing, feature selection, model training, validation, and deployment. The first step, data preprocessing, involves cleaning, transforming, and standardising raw data to ensure its quality for analysis. Feature selection then identifies the most relevant variables that enhance the model's predictive accuracy. During model training, historical data is used to develop the predictive model through various algorithms and techniques. Finally, model validation assesses the model's performance by testing it on separate datasets before deployment in real-world applications.

III. Predictive Model for Patient Risk Assessment

We propose a predictive model based on collaborative filtering techniques, designed to analyse data from Personal Health Records (PHRs) in order to deliver personalised healthcare recommendations. The model is composed of several sequential steps that transform raw, heterogeneous health data into meaningful predictions and tailored interventions. Each step plays a crucial role in structuring, analysing, and interpreting the PHR data to uncover patterns, assess risks, and generate individualised recommendations.

The diagram below outlines the main components and flow of the predictive model.

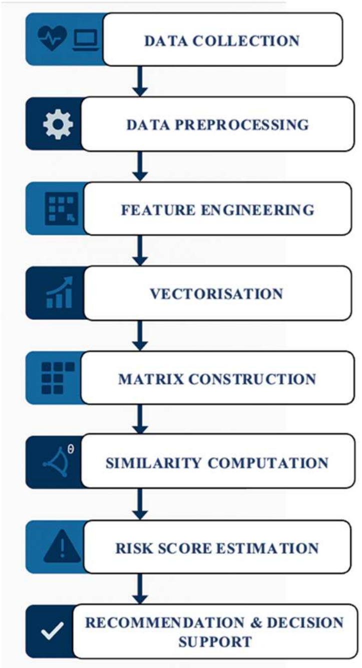

Fig. 1. Patient Algorithm Steps: Patient Risk Prediction

An enhanced approach to developing a Personal Health Record (PHR) extends beyond traditional medical records generated within healthcare institutions, enabling more accurate predictions and tailored recommendations for preventive care, early diagnosis, and timely treatment. According to Blazheska-Tabakovska et al., instead of restricting PHRs to disease-related data, this approach incorporates a broader range of health-related factors, including lifestyle habits, living conditions, and behavioural patterns [8]. By adopting this comprehensive perspective, the system constructs a more holistic representation of an individual's health lifecycle, encompassing physical, psychological, and social dimensions. Integrating diverse data sources enhances personalised healthcare interventions, optimises patient outcomes, and promotes long-term well-being.

We propose using collaborative filtering as the next approach for analysing patient data. Collaborative filtering leverages the matrix of distances generated by cosine similarity to identify patterns in patient preferences and behaviours. By comparing the similarities between patients, collaborative filtering can recommend personalised interventions, treatments, or preventive measures based on the experiences and preferences of similar patients.

Data preprocessing and feature engineering are the first and crucial steps in preparing raw data for analysis, ensuring it is in an optimal format for predictive modelling. Data transformation techniques such as normalisation, standardisation, and log transformation are applied to maintain a consistent scale and approximate a normal distribution. These techniques enhance model performance and facilitate convergence [13]. Feature selection focuses on identifying the most relevant variables that contribute to the model's predictive accuracy [14]. By reducing dimensionality, it enhances interpretability, improves computational efficiency, and minimises the risk of overfitting, ultimately leading to more robust and reliable predictions.

The data stored in a Personal Health Record (PHR) exists in various formats, including text, images, audio, and other multimedia types. To enable effective learning and analysis, this data must first be converted into a numerical format. Various transformation techniques are used to convert different data types into vectors, which are structured mathematical objects of real-world data. These techniques ensure that the data is structured appropriately for machine learning models, enabling accurate predictions and meaningful insights. By organising data into vectors, the model is prepared to perform various operations, such as clustering, classification, and regression. These vectorised representations allow the model to calculate distances and similarities, enabling it to make predictions and uncover patterns within the data. This structured approach enhances the model's ability to learn and apply insights for more accurate and meaningful outcomes. The result of this first step is the creation of patient PHR vectors, which represent the data in a structured numerical format.

$$PHR_{(patient^{(i)})} = \left(r_1^{(i)}, r_2^{(i)}, r_3^{(i)}, r_4^{(i)}, r_5^{(i)}, r_6^{(i)}, r_7^{(i)}, r_8^{(i)}, r_9^{(i)}\right) \quad (1)$$

where: $r_1^{(i)}$ - allergies $r_2^{(i)}$ - disability; $r_3^{(i)}$ - family history; $r_4^{(i)}$ - surgeries; $r_5^{(i)}$ - hospitalizations; $r_6^{(i)}$ - immunizations; $r_7^{(i)}$ – diagnoses; $r_8^{(i)}$ -medication; $r_9^{(i)}$ - lifestyle (Fig.2).

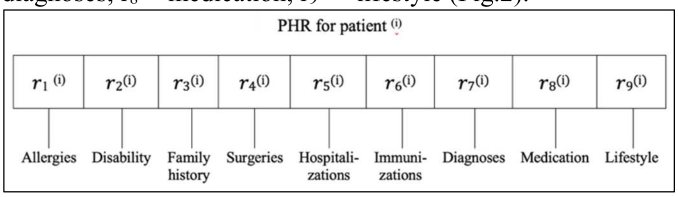

Fig. 2. Patient i PHR vector

The next step is the construction of a patient matrix, where the patient PHR vectors are organised into a matrix format. Each row in the matrix corresponds to an individual patient, while each column represents a specific feature or variable from the PHR data. This matrix structure allows the model to efficiently analyse relationships between patients and their health data, facilitating operations such as clustering, classification, and prediction.

$$PHR = \begin{bmatrix} r_1^{(1)} & \cdots & r_9^{(1)} \\ \vdots & \ddots & \vdots \\ r_1^{(m)} & \cdots & r_9^{(m)} \end{bmatrix}_{mx9} \quad (2)$$

The following process involves calculating the cosine similarity between patient vectors. Cosine similarity measures the similarity of the obtained knowledge vectors by calculating the angle between two vectors.

$$s\left(p^{(i)}, p^{(j)}\right) = \frac{r_{p^{(i)}}\, r_{p^{(j)}}}{\left|r_{p^{(i)}}\right|\left|r_{p^{(j)}}\right|} \quad (3)$$

The resulting value ranges from -1 to 1, where a lower value indicates lower similarity in preferences, and a higher value indicates higher similarity. This operation results in a matrix of distances, where each entry represents the similarity between two patients based on their PHR data. This matrix of distances can then be used to group similar patients or make personalised predictions and recommendations.

The recommendation process based on similar profiles involves determining the risk level by assessing the similarity between a given patient and others with comparable profiles. The system analyses the diagnosis and health outcomes of patients with similar characteristics to predict potential risks for the target patient. The normalised resulting score gives a value $\widehat{risk}_{p^{(i)}r_k}$, which indicates for patient i the level of risk for $r_{(k)}$ disease and recommended intervention (τ' is a sample).

$$\widehat{risk}_{p^{(i)}r_k} = \frac{\sum_{\substack{(p^{(j)},r_k \in \tau \\ p^{(j)} \neq p^{(i)}}} s\left(p^{(i)}, p^{(j)}\right) risk_{p^{(j)}r_k}}{\sum_{\substack{p^{(j)},r_k) \in \tau \\ p^{(j)} \neq p^{(i)}}} s\left(p^{(i)}, p^{(j)}\right)} \quad (4)$$

IV. Conclusion

The rapid digital transformation of the healthcare sector has opened new avenues for patient-centred care through the integration of Information and Communication Technology (ICT), the Internet of Medical Things (IoMT), and data analytics. This paper demonstrates how the integration of digital technologies—specifically Personal Health Records (PHR), collaborative filtering, and predictive modelling—can significantly improve healthcare delivery. By transforming diverse patient data into structured numerical formats and applying similarity measures, we enable personalized health recommendations, early disease risk detection, and timely interventions. The proposed framework supports a shift toward proactive, data-driven, and patient-centric care, enhancing medical outcomes, optimising clinical workflows, and improving resource utilisation.

In addition to technical innovation, this approach aligns with global healthcare strategies that emphasise preventive care, digital inclusion, and the use of predictive analytics in medicine. It has the potential to support physicians in clinical decision-making, reduce hospitalisation rates, and enable remote monitoring through intelligent, interoperable systems. The use of collaborative filtering based on cosine similarity allows the model to draw insights from shared patient characteristics, offering robust and scalable solutions applicable across diverse healthcare environments.

Compared to traditional predictive models in healthcare, the proposed model offers distinct advantages through its use of collaborative filtering techniques. While many existing approaches rely heavily on predefined clinical variables and labelled datasets, our model leverages similarity between patient profiles based on Personal Health Records (PHRs), enabling it to uncover patterns without requiring extensive manual feature engineering or large-scale annotation. This makes it particularly well-suited for personalised recommendations and risk assessment in heterogeneous health data. Furthermore, our model supports the adaptability of different data sources, including multimodal inputs such as text, images, and sensor data from IoMT devices.

As healthcare systems continue to embrace digital transformation, the proposed model offers a framework for intelligent decision support with scalability, adaptability, and future-proofing in mind. It is designed to accommodate increasing volumes of Personal Health Record (PHR) data, maintaining performance and effectiveness as data availability, integration and system demands grow.

The computational complexity of the similarity computation step, which involves pairwise cosine similarity calculations, is $O(n^2)$ in the worst case. However, this can be mitigated through dimensionality reduction techniques (e.g., PCA) and approximate nearest neighbour search algorithms, which significantly reduce latency without compromising accuracy. Future work will explore integration with real-time IoMT infrastructures, validation with larger datasets, and deployment in clinical practice. Additionally, ethical considerations such as data privacy, security, and patient consent will be addressed to ensure trustworthy and responsible use of health data.

References

[1] J. Lopes, T. Guimarães and M. F. Santos, “Predictive and Prescriptive Analytics in Healthcare: A Survey,” in *Procedia Computer Science 170*, Warsaw, Poland, 2020.

[2] S. Savoska, N. Blazheska-Tabakovska, I. Jolevski, A. Bocevska, B. Ristevski and V. Trajkovik, “Integration of IoMT Sensors’ Data from Mobile Applications into Cloud Based Personal Health Record.,” in *13th International Conference, ICT Innovations*, Ohrid, Macedonia, 2021.

[3] European Commission, “eHealth action plan 2012–2020: Innovative healthcare for the 21st century.,” Publications Office of the European Union. https://ec.europa.eu/digital-strategy/ehealth-action-plan-2012-2020, 2012.

[4] World Health Organization., “Global strategy on digital health 2020–2025.,” WHO. https://www.who.int/publications/i/item/9789240020924, 2021.

[5] A. Bocevska, Savoska, Snezana, I. Jolevski, N. Blazheska-Tabakovska and B. Ristevski, “mplementation of Innovative e-Health Services and Digital Healthcare Ecosystem – Cross4all Project Summary,” in *The 15-th conference on Information Systems and Grid Technologies*, Sofia, Bulgaria, 2022.

[6] T. Gunter and N. Terry, “The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions.,” *J Med Internet Res 7(1):e3 doi: 10.2196/jmir.7.1.e3. PMID: 15829475; PMCID: PMC1550638.,* 2005 Mar 14.

[7] S. Savoska, B. Ristevski, N. Blazheska-Tabakovska and I. Jolevski, “Towards integration exposome data and personal health records in the age of IoT, I,” in *ICT innovations 2019*, Ohrid, 2019.

[8] N. Blazheska-Tabakovska, A. Bocevska, I. Jolevski, B. Ristevski, N. Beredimas, V. Kilintzis, N. Maglaveras and S. Savoska, “Implementation of Cloud-Based Personal Health Record Integrated with IoMT,” in *The 14-th conference on Information Systems and Grid Technologies*, Sofia, Bulgari, 2021.

[9] A. Rehman, S. Naz and I. Razzak, “Leveraging big data analytics in healthcare enhancement: trends, challenges and opportunities.,” *Multimedia Systems, 28(4), ,* pp. 1339-1371, 2022.

[10] N. Blazheska-Tabakovska, Knowledge-based systems, Bitola: Macedonian scientific society, 2023.

[11] I. G. Duncan, Healthcare risk adjustment and predictive modeling., Actex Publications, 2011.

[12] E. Okpokoro, M. Lesosky, C. Osa-Afiana, F. Bada, U. Okwor, G. Odonye, V. Igbinomwanhia, S. Abdurrahman, N. Medugu, B. Kagina and A. Abimiku, “Prevalence and Risk Factors for Mycobacterium tuberculosis Infection among Health Workers in HIV Treatment Centers in North Central, Nigeria.,” *The American Journal of Tropical Medicine and Hygiene, 109(1),* pp. 60-68, 2023.

[13] C. S. Nwaimo, A. E. Adegbola and M. D. Adegbola, “Transforming healthcare with data analytics: Predictive models for patient outcomes,” *GSC Biological and Pharmaceutical Science27(03),* p. 25–35, 2024.

[14] M. E. Shipe, S. A. Deppen, F. Farjah and E. L. Grogan, “Developing prediction models for clinical use using logistic regression: an overview.,” *Journal of thoracic disease,* vol. 11(Suppl 4), p. S574, 2019.