






## Article

# Improving Diagnostics with Deep Forest Applied to Electronic Health Records

Atieh Khodadadi <sup>1,†</sup> , Nima Ghanbari Bousejin <sup>2,†</sup> , Soheila Molaei <sup>3,\*</sup> , Vinod Kumar Chauhan <sup>3</sup> ,  
Tingting Zhu <sup>3</sup>  and David A. Clifton <sup>3,4</sup>

<sup>1</sup> Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology, 76133 Karlsruhe, Germany

<sup>2</sup> Independent Researcher, Tehran 009821, Iran; bousejin@gmail.com

<sup>3</sup> Department of Engineering Science, University of Oxford, Oxford OX1 3AZ, UK; vinod.kumar@eng.ox.ac.uk (V.K.C.); tingting.zhu@eng.ox.ac.uk (T.Z.); david.clifton@eng.ox.ac.uk (D.A.C.)

<sup>4</sup> Oxford-Suzhou Centre for Advanced Research (OSCAR), Suzhou 215123, China

\* Correspondence: soheila.molaei@eng.ox.ac.uk

† These authors contributed equally to this work.

**Abstract:** An electronic health record (EHR) is a vital high-dimensional part of medical concepts. Discovering implicit correlations in the information of this data set and the research and informative aspects can improve the treatment and management process. The challenge of concern is the data sources' limitations in finding a stable model to relate medical concepts and use these existing connections. This paper presents Patient Forest, a novel end-to-end approach for learning patient representations from tree-structured data for readmission and mortality prediction tasks. By leveraging statistical features, the proposed model is able to provide an accurate and reliable classifier for predicting readmission and mortality. Experiments on MIMIC-III and eICU datasets demonstrate Patient Forest outperforms existing machine learning models, especially when the training data are limited. Additionally, a qualitative evaluation of Patient Forest is conducted by visualising the learnt representations in 2D space using the t-SNE, which further confirms the effectiveness of the proposed model in learning EHR representations.

**Keywords:** electronic health record; deep learning; intensive care unit; deep random forest; representation learning



**Citation:** Khodadadi, A.; Ghanbari Bousejin, N.; Molaei, S.; Kumar Chauhan, V.; Zhu, T.; Clifton, D.A. Improving Diagnostics with Deep Random Forest Applied to Electronic Health Records. *Sensors* **2023**, *23*, 6571. <https://doi.org/10.3390/s23146571>

Academic Editors: Wenyang Jia, Yi Gao, Zhi-Hong Mao and Mingui Sun

Received: 30 May 2023

Revised: 8 July 2023

Accepted: 14 July 2023

Published: 21 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Medical and therapeutic techniques have substantially benefited from the collection of health data and the use of such data in the field of data science [1–3]. EHRs are one of these enormous sources of data, helpful for a variety of predictive tasks in medical applications [4]. EHRs hold a patient's demographics, medical history, vital signs, laboratory tests, recommended medicine, diagnosis, and clinical outcomes during an interaction [5,6]. EHR databases may contain several patient visits, establishing a longitudinal patient record that can be used to aim the treatment process, such as disease prediction, mortality prediction, and enhancing the efficacy of the therapeutic process.

Initially, EHR systems were intended to manage the basic administrative functions of hospitals, permitting the use of regulated terminology and labelling schemes. Numerous labelling schemes exist, including ICD (International Statistical Classification of Diseases) codes for diagnostic [7–10], CPT (Current Procedural Terminology) codes for procedures [11–13], and LOINC (Logical Observation Identifiers Names and Codes) for laboratories [14,15], ATC (Anatomical Therapeutic Chemical) for drug [16,17], and RxNorm for medication [12]. The various labelling techniques produce standard datasets for varied specialisations. As the EHR system develops, the volume of EHR data increases annually, and several studies have been conducted on the secondary use of these data.

EHRs offer numerous benefits, including improved patient care, increased efficiency, and reduced healthcare costs [18]. Regardless of the potential for EHRs in various applications, their effective usage is hindered by data-specific restrictions [6], such as high missingness and irregular sampling [19–21], as well as imbalanced classes due to uneven prevalence of illnesses [22]. Therefore, it is important to address these limitations in order to fully realise the potential of EHRs.

Previous work on learning EHRs representations has mainly focused on developing methods for data integration, such as ontology-based mapping or semantic matching, which often require significant manual effort and are limited in their effectiveness in combining different data types. To address these limitations, our proposed approach leverages the latest advancements in deep learning and representation learning to create a more unified representation of EHR data. Our method utilises neural networks to learn a compact and meaningful representation of EHR data, which can be used for a variety of disease prediction tasks. Additionally, our approach can combine different data types effectively, such as demographic information, laboratory test results, and imaging studies, into a single, unified representation. This enables us to capture the complex relationships between different data elements and provides a more complete picture of the patient's health status.

Classic machine learning approaches, such as Support Vector Machine (SVM) [23] and Random Forest (RF) [24], have been used previously to process large EHR datasets. However, these methods are limited in their ability to capture the complex patterns in the underlying data. In comparison, deep learning approaches based on deep data feature analysis are capable of producing efficient and reliable analytical outcomes, particularly in the real-world context of enormous data volumes. While deep learning-based models have shown promise in improving patient outcome prediction, they often require large amounts of data and computational power, which may not be available in all settings. Compared to deep learning models, Patient Forest offers several advantages. By using an ensemble of decision trees, Patient Forest is able to perform well even with smaller datasets and less computational power, which can be important in clinical settings.

In this work, we propose a novel approach, called Patient Forest, to learning EHRs representations based on the cascade deep forest method [25]. Our Patient Forest technique incorporates statistical features to create a more accurate classifier for predicting readmission and mortality. Our approach seeks to automate the process of data integration and generate representations that capture the complex patterns in the underlying data. Moreover, our technique performs a brief preprocessing in order to optimise the final performance. RF is a widely used technique for assessing data with a large dimension [26], and by utilising its random attribute, generating several forests, and getting multiple outcomes simultaneously in each layer of the deep network, our approach is able to enhance the performance of the predictions.

The main contributions of this paper include:

- We propose Patient Forest, a machine-learned approach for predicting patient outcomes, that incorporates statistical features to learn EHR representations.
- We conduct an evaluation of our proposed method on two large-scale EHR datasets and demonstrate its effectiveness in predicting readmission and mortality outcomes.
- We compare the performance of Patient Forest with strong machine learning baselines to reveal the effectiveness of our approach to improve patient care and reduce healthcare costs.

## 2. Related Work

In this section, we present a summary of related works in representation learning for EHR applications.

### 2.1. Vector-Based Methods

One of the learning models that represents patient information on this basis is a fully connected Deep Neural Network (DNN). Futoma et al. [27] evaluated various models' propensity to forecast hospital readmissions using data from a large EHR database. The outcome demonstrates DNN outperforms other approaches that have previously been used to solve this issue in terms of prediction performance. The study given in [28] employed a deep generative learning model to overcome the problem of insufficient data using MRI pictures efficiently by learning and categorising tumour locations from MRI images. The search by Zheng et al. [29] for suicide ideation, behaviour, or death prediction in the literature was based on the health records of patients who had visited a Berkshire Health System hospital. Multiple machine learning and deep learning methodologies are employed in EHRs to classify the severity of patients in [30]. The experimental findings indicate DNN performed exceptionally well. In the type II diabetes disease prediction [31], a deep learning neural network architecture model was adopted. All these studies demonstrated the DNN can be utilised for EHR data analysis and diagnosis. Despite this, the majority of recent research has considered this architecture to be the classic way [32].

Autoencoders are vector-based, unsupervised deep learning models, which are an efficient dimensionality reduction technique with promising performance for the deep representation of medical data [33]. Autoencoders have also been effectively applied to datasets comprising massive collections of electronic health records, where they are very adept at handling missing data [34]. A comparison study by Sadati et al. [35] emphasised the effectiveness of several types of autoencoders for electronic health record-based data sets. Combining a recurrent autoencoder with two GANs, Lee et al. [36] suggested sequential electronic health records with a dual adversarial autoencoder (DAAE). Biswal et al. [37] synthesised sequences of discrete EHR encounters and encounter features using a variational autoencoder. Very recently, in [38], for adverse drug event preventability, a model of dual autoencoders was explored in EHRs. Wang et al. [39] compared the model with autoencoder features to traditional models, which could show a reasonable result.

Convolutional Neural Networks (CNNs) are a further vector-based technique. EHR research [40] focuses on capturing the local temporal dependence of these data, which are then used to predict multiple diseases and for other related tasks. Wang et al. [41] adopted a CNN learning with 1929 features for the classification of 1099 international diseases. Researchers in [42] aimed to develop a convolutional neural network model for the prediction of the risk of advanced nonmelanoma skin cancer (NMSC) in Taiwanese adults. In an intriguing study [43], CNN was applied over electronic health records to determine the top 20 lung-cancer-related indicators in order to avoid radiation exposure and costs. CNN has shown its superior ability to measure patient similarity. However, the traditional CNN architecture could not properly exploit the temporal and contextual information of EHRs for disease prediction. Consequently, it is increasingly difficult to represent the timing and substance of EHR data concurrently [44].

Natural language processing was the original inspiration for word2vec [45], which was developed to learn word embeddings from large-scale text resources. In [46], the authors pursue the word2vec technique to train a two-layer neural network to improve clinical application prediction accuracy relative to baselines. Choi et al. [47] applied skip-gram to longitudinal EHR data to learn low-dimensional representations of medical concepts. To improve the performance of a convolutional neural network for patient phenotyping, Yang et al. [48] explored a model that combines token-level and sentence-level inputs. Similarly, in [49], clinical text was employed to expect clinical notions. Steinberg et al. [50] proposed a novel analogy of language modelling on discretised clinical time-series data. However, these techniques do not explicitly model dynamic temporal information or address the challenges of heterogeneous data sources [51].

## 2.2. Temporal Matrix-Based Methods

Lee and Seu [52] presented Non-Negative Matrix Factorisation (NMF) as a method for discovering a collection of basic functions for expressing non-negative data. This matrix pertains to electronic health records, which generate a matrix with a time dimension and a clinical event dimension. Bioinformatics has extensively used NMF for clustering sources of variation [53–55]. There are other efforts to use NMF or its variants in the depiction of patient data in EHRs. In [56], disease trajectories are analysed using NMF to extract multi-morbidity patterns from a huge data collection of electronic health records. Zhao et al. [57] suggested that the NMF identifies relationships between genetic variants and disease phenotypes. In a recent study [58], NMF was used to examine the symptoms of covid and predict long-term infection. Controlling the degree to which the representation is sparse is difficult since sparseness is a side effect of the NMF algorithm [59]. The huge number of various diagnosis codes is an additional obstacle that results in a combinatorial explosion of the number of possible diseases, many of which are unique to a single patient [60].

## 2.3. Graph-Based Methods

The graph technique can be expressed using the EHR by using nodes to represent medical events and edges between the nodes to highlight the temporal links among clinical events. One emerging method of deep learning on graph-structured data is Graph Neural Networks (GNNs) [61]. GNNs can infer the missing information, leading to a representation that is more explicable [62]. The hierarchical relationships in EHRs were captured using GNN, as described in reference [63,64]. In [65], GNN reflected the links between drugs, side effects, diagnosis, associated treatments, and test results. For instance, Park et al. [66] suggested a knowledge graph-based question answering with EHR. Research [67] introduced an EHR-oriented knowledge graph system to efficiently utilise non-used information buried in EHRs. In EHRs, it is typical for spurious edges to be included and for other edges to be absent. Even though the observed graph is clean, it may contravene the properties of GNNs because it is not jointly optimised with them. These flaws in the observed graph may precipitously degrade the performance of GNNs [68].

## 2.4. Sequence-Based Methods

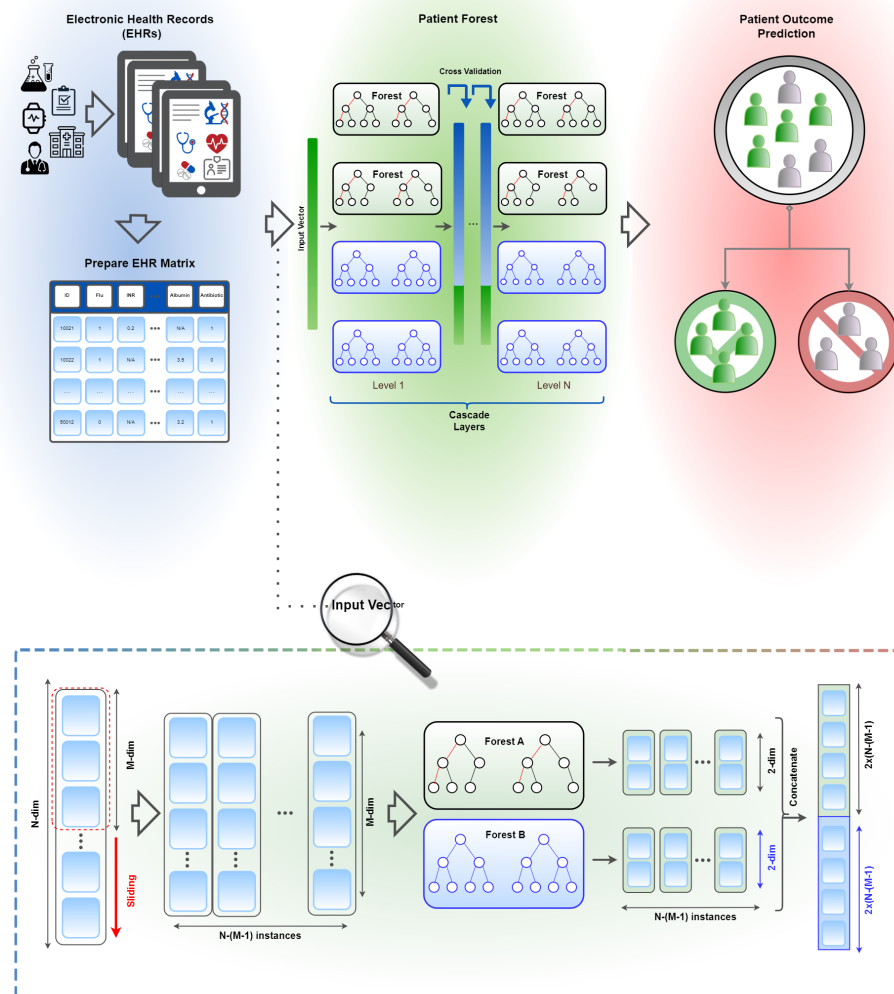
Sequence-based patient representation turns EHR data into a temporally ordered sequence of clinical events for use in prediction. A recurrent neural network (RNN) is a neural network that includes the GRU and LSTM networks as specific cases, according to Sherstinsky's study [69]. RNNs are widely used in patient representation research that focuses on combinations or sequences of clinical codes [62]. The research included aid in early diagnosis [70,71] and disease prediction [72–79]. Recently, Gupta et al. [80] adopted a general LSTM network architecture to make improved predictions of BMI and obesity. Ref. [81] examined the performance of various deep neural network architectures, including LSTM, in scenarios involving clinical factors and chest X-ray radiology reports, revealing that the recommended BiLSTM model outperforms other DNN baseline models. RNN is frequently stated without context or rationale. In addition, training equations are frequently removed entirely; therefore, partial descriptions or missing formulas in RNN may result in its inefficiency [69].

## 2.5. Tensor-Based Methods

Tensor-based methods apply an n-dimensional tensor to represent patient information. The multi-dimensional and high level of tensor factors in EHR data make complex relationships understandable and interpretable [82]. Zhao et al. [83] identified previously unknown cardiovascular characteristics using a modified non-negative tensor-factorisation technique. Afshar et al. [84] implemented temporal and static tensor factorisation to extract clinically significant characteristics. Hernandez et al. [85] used a novel tensor-based dimensionality reduction method to predict the onset of haemodynamic decompensation.

### 3. Methodology

In the following, the specifics of the proposed approach will be presented. We frame the patient outcome prediction task as a classification problem. Patient Forest uses a prepared EHR matrix to learn EHR representations by using an ensemble of decision trees. These representations then may be retrieved and used for downstream patient outcome prediction tasks. Figure 1 depicts a summary of the model. In the subsequent sections, we explain the implementation details and experimental setup.



**Figure 1.** Schematic representation of Patient Forest, illustrating the process of patient outcome prediction based on learnt EHR representations from a prepared EHR matrix.

#### 3.1. Patient Forest

We are provided with a set of  $N$  encounters,  $X = \{\vec{x}_1, \vec{x}_2, \dots, \vec{x}_N\}$ , as input, where  $\vec{x}_i \in R^F$  represents encounter  $i$  features. We feed encounters to the gcForest model [25] in order to learn EHR representations. The learnt EHR representations are then used for downstream patient outcome prediction tasks.

The gcForest model is made up of two distinct modules: multi-grained scanning and cascade forest. The multi-grained scanning module is responsible for generating a set of diversified features from input data by using multiple layers of sliding windows (convolutional filters) of different sizes, which results in a set of sub-sampled feature maps, each capturing different aspects of the input data at different granularities. The output of this module is then fed to the cascade forest module.

The cascade forest module is made up of multiple levels of random forests. Each level takes as input the output of the previous level and further refines the extracted features. The final output of the cascade forest is a set of predictions for the input data.

During training, the gcForest model learns the input features by optimising the parameters of the convolutional filters and random forests using a backpropagation algorithm. These learned features are then used to make predictions on new, unseen data. For the training objective, we use a standard binary cross-entropy (BCE) loss between the target and predicted labels.

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N (\hat{y}_i \log(y_i) + (1 - \hat{y}_i) \log(1 - y_i)) \quad (1)$$

where  $\hat{y}_i$  is the network's predicted label, and  $y_i$  is the ground-truth label.

### 3.2. Datasets and Preprocessing

#### 3.2.1. Datasets

Our proposed model examined, using the eICU [86] and MIMIC-III [87] datasets, both of which are large-scale electronic health record (EHR) dataset collections and are accessible through the PhysioNet repository [88].

**eICU.** Philips Healthcare has created the eICU Program, a telemedicine system that utilises these data to aid in the treatment of critically sick patients. The eICU Collaborative Research Database is a multi-centre intensive care unit (ICU) database providing high-resolution data for over 200,000 admissions between 2014 and 2015, to one of 335 units at 208 US hospital institutions. The de-identified database contains information such as vital sign readings, care plan paperwork, sickness severity measurements, diagnostics, and treatments [86].

**MIMIC-III.** Medical Information Mart for Intensive Care III (MIMIC-III) is a big, single-centre database including information on Beth Israel Deaconess Medical Center (BIDMC) in the United States from 2001 to 2012. Data comprises vital signs, medicines, laboratory measures, observations and comments documented by care providers, fluid balance, procedure codes, diagnostic codes, imaging reports, hospital length of stay, and survival data [87].

Readmission rates and death rates are the primary outcomes that catch our attention for the purpose of this article.

#### 3.2.2. Preprocessing

The EHR representations on MIMIC and eICU datasets are learned by following the preprocessing suggested in [89]. To prep the datasets, we removed encounters that were shorter than 24 h and eliminated duplicate codes (e.g., repeated medication administration). Lab results were also excluded as their values can change during ICU stays (e.g., blood pH level). These steps resulted in 50,391 and 41,026 patients in the MIMIC and eICU datasets, respectively. Throughout the paper, we focus on a single encounter and do not consider the time-series nature of EHRs. See Table 1 for the statistical details of the two datasets.

### 3.3. Baselines and Tasks

#### 3.3.1. Baselines

To assess our model's performance in prediction tasks, we evaluate it against different baselines, including a Random Forest, Multi-Layer Perceptron, Logistic Regression, Support Vector Machine, Naïve Bayes, Classification and Regression Trees, and K-Nearest Neighbours.

- **Random Forest (RF):** Random Forest is a classifier comprising decision trees, where each tree at input  $x$  casts one vote for the most popular class [90].
- **Multi-Layer Perceptron (MLP):** Multi-Layer Perceptron is a type of artificial neural network consisting of multiple layers of neurons that are connected to each other and used for supervised learning tasks such as classification [91].

- Logistic Regression (LR): Logistic Regression is a machine learning algorithm used for predicting binary outcomes given a set of features [92].
- Support Vector Machines (SVM): Support Vector Machines are a type of supervised learning algorithm used for finding the line that maximises the minimum distance to the line [93].
- Naive Bayes (NB): Naive Bayes is a probabilistic classifier based on the assumption that all features are conditionally independent of each other given a class label [94].
- Classification and Regression Trees (CART): Classification and Regression Trees are a type of decision tree algorithm used for classification and regression problems that employ past data to generate decision trees, which are then used to categorise fresh data [95].
- K-Nearest Neighbours (K-NN): K-Nearest Neighbours is a non-parametric algorithm used for classification by identifying the closest k-neighbours to an observation and then assigning it a class label [96].

**Table 1.** Statistical characteristics of the datasets used to train and evaluate the model for both the mortality and readmission tasks. N/A stands for Not Applicable.

Dataset	MIMIC	eICU
Avg. # of diagnosis per visit	11.5	6.5
Avg. # of treatment per visit	4.5	5.0
# of Positives (Readmission)	N/A	7051
# of Positives (Mortality)	5377	2983
# of Patients	50,391	41,026

### 3.3.2. Tasks

The purpose of this study was to compare Patient Forest and baseline models in their ability to predict two primary tasks: readmission prediction and mortality prediction.

- Readmission Prediction: The models were trained to extract visit representations from an encounter record in order to predict whether a patient will be readmitted to the ICU within the same hospital stay. This task was evaluated using only the eICU dataset.
- Mortality Prediction: The models were trained to extract visit representations from an encounter record in order to forecast patient mortality upon ICU admission. This task was assessed using both the MIMIC and eICU datasets.

### 3.4. Training and Evaluation

The gcForest model used in our experiments had a cascade structure comprising 4 completely random tree forests and 4 random forests, with 500 trees in each forest. Class vector generation was achieved using three-fold cross-validation. The number of cascade levels was automatically determined by splitting the training set into a growing set and an estimating set. The cascade was retrained after determining the estimated number of levels. Multi-grained scanning was performed using three different window sizes for feature windows. Moreover, 80% of the training data was used for the growing set and 20% for the estimating set.

We used two metrics to measure how accurately our patient outcome prediction tasks were performed: the Area Under the Precision-Recall Curve (AUPRC) [97] and the Area Under the Receiver Operating Characteristic Curve (AUROC) [98]. These metrics were calculated on the test set, which had the same class distribution as the actual data. AUPRC is sensitive to the proportion of positive outcomes, so the lowest possible value and the value of a random classifier would depend on the positive class rate [99]. AUROC is a measure of how well a classifier can separate positive and negative outcomes, regardless of the class distribution. The AUROC value of a random classifier is always 0.5.

We conducted 20 runs of training across three different train–test set splits (75%:25%, 50%:50%, and 30%:70%) and evaluated the performance of patient outcome prediction

tasks across our baselines by measuring the AUPRC and AUROC on the test nodes of our method. To prevent data contamination, we used a patient-level split of the data, ensuring that each patient was included only in one split.

#### 4. Experiments

We compared the results of our Patient Forest model to several baseline models such as RF, MLP, LR, VM, NB, CART, and K-NN. The performance was evaluated using the AUPRC and AUROC metrics, and the results are reported in Tables 2 and 3 for the 70%:30% data split and Figure 2 for the 50%:50% and 30%:70% data splits.

**Table 2.** Readmission and mortality prediction performance on eICU and MIMIC in terms of AUPRC using 75%:25% data split.

Models	MIMIC Mortality	eICU Mortality	eICU Readmission
RF [90]	0.5166	0.4843	0.4528
MLP [100]	0.5676	0.4549	0.4519
LR [92]	0.5863	0.4903	0.4628
SVM [101]	0.5602	0.5193	0.4612
NB [101]	0.4506	0.3532	0.3491
K-NN [96]	0.4575	0.4196	0.4016
XGB [96]	0.5187	0.4599	0.4481
Patient-Forest (ours)	0.619	0.5732	0.5952

**Table 3.** Readmission and mortality prediction performance on eICU and MIMIC in terms of AUROC using 75%:25% data split.

Models	MIMIC Mortality	eICU Mortality	eICU Readmission
RF [90]	0.7628	0.8392	0.8136
MLP [100]	0.7492	0.8163	0.8121
LR [92]	0.7732	0.8328	0.8192
SVM [101]	0.7463	0.8473	0.8213
NB [101]	0.6583	0.7628	0.7605
K-NN [96]	0.6538	0.7927	0.8001
XGB [96]	0.7719	0.8129	0.8083
Patient-Forest (ours)	0.8010	0.8637	0.8690

#### Results

- Predictive Performance:** We compared the performance of Patient Forest with other baseline models, such as RF, MLP, LR, SVM, NB, CART, and K-NN. We used two metrics to evaluate the models: AUPRC and AUROC. AUPRC measures how well the model can identify the positive class (i.e., patients who died or were readmitted), while AUROC measures how well the model can separate positive and negative classes (i.e., patients who survived or were not readmitted). Higher values of both metrics mean better performance. We used two datasets: MIMIC and eICU, and three data splits: 75%:25%, 50%:50%, and 30%:70%. Tables 2 and 3 show the results for AUPRC and AUROC using the 75%:25% split. Figure 2 shows the results for AUPRC using the other two splits. The results show that Patient Forest consistently outperformed all the baseline models on both metrics and both datasets.

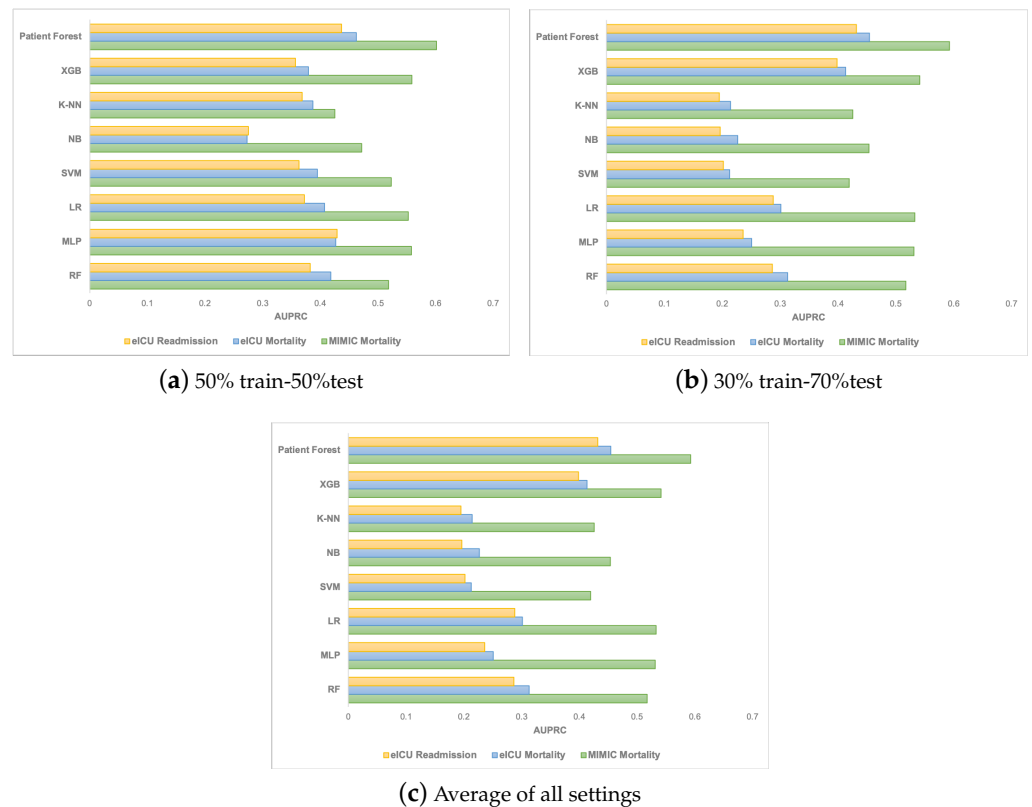
Using the 75%:25% train–test split, the Patient Forest model outperformed all the other models on both AUPRC and AUROC. It achieved 0.619 AUPRC and 0.8010 AUROC for MIMIC Mortality, 0.5732 AUPRC and 0.8637 AUROC for eICU Mortality, and 0.5952 AUPRC and 0.8690 AUROC for eICU Readmission. Using the 50%:50% split, the model still demonstrated superior performance with AUPRC of 0.6019, 0.4626, and 0.4372 in MIMIC Mortality, eICU Mortality, and eICU Readmission, in order. Even



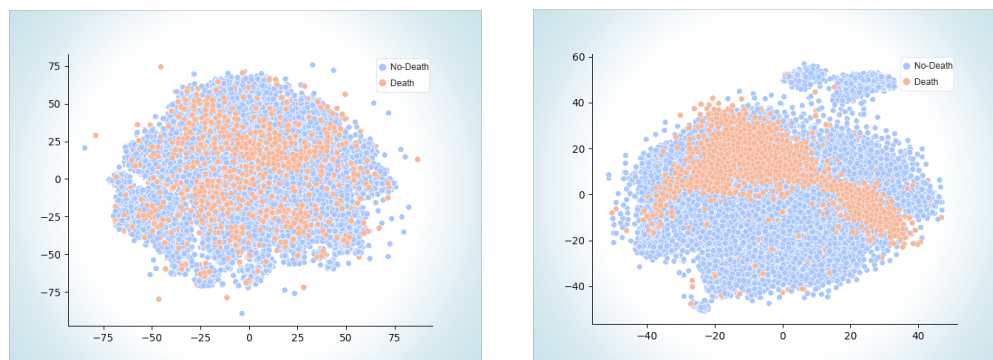
with a reduced percentage of training data (30%), our model still outperformed the other baselines with AUPRC of 0.5927, 0.4544, and 0.4317 in MIMIC Mortality, eICU Mortality, and eICU Readmission, respectively.

These results demonstrate the robustness and generalizability of our model across different datasets and data splits. The superior performance of Patient Forest can be attributed to its ability to capture the heterogeneity and complexity of the patient data using a forest of patient-specific decision trees. By learning from the patient's own history and features, our model can better predict the patient's future outcomes than the models that use a single global classifier for all patients. Moreover, by aggregating the predictions of multiple trees, our model can reduce the variance and improve the stability of the results.

- **Comparative Evaluation:** To assess the effectiveness of Patient Forest and its learned representations, we plotted t-distributed stochastic neighbour embedding (t-SNE) plots [102] of the generated representations (Figure 3) for the MIMIC mortality prediction task. The different colours denote different patient classes. Our qualitative results demonstrate Patient Forest is able to learn representations that place patients with similar outcomes close to each other, indicating the model's capability of accurate prediction of outcomes.



**Figure 2.** Readmission and mortality prediction performance on eICU and MIMIC in terms of AUPRC with different data splits: 50%:50% (a), 30%:70% (b), and average of all settings (c).



**Figure 3.** t-SNE embeddings of the patients in the MIMIC dataset based on raw (left) and learnt Patient Forest features (right).

## 5. Discussion and Conclusions

We presented Patient Forest, a model which can learn EHR representations to predict mortality and readmission rates. Our proposed approach was trained and tested on two extensive EHR datasets and three benchmark tasks. We compared its performance with other prominent models and highlighted the benefits of our model. Furthermore, we did a qualitative assessment of Patient Forest by plotting the t-SNE of the embeddings for the targeted outcomes across two studied datasets. Results indicated that the learnt representations provide a 2-D projection that clearly reveals clustering.

This study has notable merits. Patient Forest can accurately learn EHR representations and surpass other strong models, especially when the training data is scarce, which is a common difficulty in healthcare domains [103]. We also validated our proposed model on two expansive EHR datasets. Our method can provide valuable insights and guidance for clinicians and patients to improve the quality of care and health outcomes.

Conversely, there are certain limitations to this study. We did not factor in the temporal properties of EHRs. Additionally, we only incorporated three primary tables from MIMIC and eICU datasets, such as diagnosis, laboratory results, and treatment tables. We intend to explore strategies to extend our work to time-series EHRs, which could be beneficial for learning representations of patient deterioration tasks over time and include more tables like demographics and procedures in upcoming works.

To summarise, our study reveals that Patient Forest is a viable model for predicting mortality and readmission rates. Furthermore, it should be explored for its potential in other areas such as disease stratification, diagnostics, and prognosis. Additionally, research should be done to determine the optimal hyperparameters of Patient Forest, as well as explore its integration with other AI models to enhance accuracy and performance. Last, the application of this model in a clinical setting must be evaluated to assess its utility for healthcare professionals.

**Author Contributions:** Conceptualization, A.K., N.G.B., S.M. and D.A.C.; methodology, S.M., T.Z. and D.A.C.; software, S.M.; validation, S.M. and T.Z.; formal analysis, N.G.B.; investigation, A.K., N.G.B., S.M. and V.K.C.; resources, A.K.; data curation, S.M.; writing—original draft preparation, N.G.B.; writing—review and editing, N.G.B., S.M., V.K.C., T.Z. and D.A.C.; visualization, A.K. and S.M.; supervision, D.A.C.; project administration, D.A.C.; funding acquisition, D.A.C. All authors have read and agreed to the published version of the manuscript.

**Funding:** D.A.C. was supported in part by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC); the InnoHK Hong Kong Centre for Cerebro-cardiovascular Health Engineering; the Pandemic Sciences Institute at the University of Oxford; an NIHR Research Professorship and a Royal Academy of Engineering Research Chair. T.Z. was supported by a Royal Academy of Engineering Research Fellowship.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Datasets are accessible through the PhysioNet repository [88] at <https://www.physionet.org/>.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Kouchaki, S.; Ding, X.R.; Sanei, S.; Zhang, Z.; Liu, Y.; Zhang, J.; Zhang, Y.; Shen, D.; Zhang, J. Artificial Intelligence-Based Applications in Medical Imaging: A Review of Recent Advances and Future Directions. *Sensors* **2021**, *21*, 4800. [[CrossRef](#)]
2. Bieberle, A.; Windisch, D.; Iskander, K.; Bieberle, M.; Hampel, U. Artificial Intelligence in Medical Sensors. *Sensors* **2020**, *20*, 5174. [[CrossRef](#)]
3. Rajpurkar, P.; Chen, E.; Banerjee, O.; Topol, E.J. AI in health and medicine. *Nat. Med.* **2022**, *28*, 31–38. [[CrossRef](#)]
4. Nordo, A.H.; Levieux, H.P.; Becnel, L.B.; Galvez, J.; Rao, P.; Stem, K.; Prakash, E.; Kush, R.D. Use of EHRs data for clinical research: Historical progress and current applications. *Learn. Health Syst.* **2019**, *3*, e10076. [[CrossRef](#)] [[PubMed](#)]
5. Birkhead, G.S.; Klompas, M.; Shah, N.R. Uses of electronic health records for public health surveillance to advance public health. *Annu. Rev. Public Health* **2015**, *36*, 345–359. [[CrossRef](#)] [[PubMed](#)]
6. Ghosheh, G.; Li, J.; Zhu, T. A review of Generative Adversarial Networks for Electronic Health Records: Applications, evaluation measures and data sources. *arXiv* **2022**, arXiv:2203.07018.
7. Cloitre, M. ICD-11 complex post-traumatic stress disorder: Simplifying diagnosis in trauma populations. *Br. J. Psychiatry* **2020**, *216*, 129–131. [[CrossRef](#)] [[PubMed](#)]
8. Harrison, J.E.; Weber, S.; Jakob, R.; Chute, C.G. ICD-11: An international classification of diseases for the twenty-first century. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 206. [[CrossRef](#)]
9. Jacobs, J.P.; Franklin, R.C.; Béland, M.J.; Spicer, D.E.; Colan, S.D.; Walters, H.L.; Bailliard, F.; Houyel, L.; Louis, J.D.S.; Lopez, L.; et al. Nomenclature for pediatric and congenital cardiac care: Unification of clinical and administrative nomenclature—The 2021 international paediatric and congenital cardiac code (IPCCC) and the eleventh revision of the International classification of diseases (ICD-11). *Cardiol. Young* **2021**, *31*, 1057–1188.
10. Maercker, A. Development of the new CPTSD diagnosis for ICD-11. *Borderline Personal. Disord. Emot. Dysregulation* **2021**, *8*, 1–4. [[CrossRef](#)]
11. Joo, H.; Burns, M.; Kalidaikurichi Lakshmanan, S.S.; Hu, Y.; Vydiswaran, V.V. Neural machine translation-based automated current procedural terminology classification system using procedure text: Development and validation study. *JMIR Form. Res.* **2021**, *5*, e22461. [[CrossRef](#)]
12. Bowie, M.J. *Understanding Current Procedural Terminology and HCPCS Coding Systems*; Cengage Learning: Boston, MA, USA, 2021.
13. Levy, J.; Vattikonda, N.; Haudenschild, C.; Christensen, B.; Vaickus, L. Comparison of machine-learning algorithms for the prediction of current procedural terminology (CPT) codes from pathology reports. *J. Pathol. Inform.* **2022**, *13*, 100165. [[CrossRef](#)] [[PubMed](#)]
14. Stram, M.; Gigliotti, T.; Hartman, D.; Pitkus, A.; Huff, S.M.; Riben, M.; Henricks, W.H.; Farahani, N.; Pantanowitz, L. Logical observation identifiers names and codes for laboratorians: potential solutions and challenges for interoperability. *Arch. Pathol. Lab. Med.* **2020**, *144*, 229–239. [[CrossRef](#)]
15. Yeh, C.Y.; Peng, S.J.; Yang, H.C.; Islam, M.; Poly, T.N.; Hsu, C.Y.; Huff, S.M.; Chen, H.C.; Lin, M.C. Logical observation identifiers names and codes (Loinc<sup>®</sup>) applied to microbiology: A national laboratory mapping experience in Taiwan. *Diagnostics* **2021**, *11*, 1564. [[CrossRef](#)]
16. Tayebati, S.K.; Nittari, G.; Mahdi, S.S.; Ioannidis, N.; Sibilio, F.; Amenta, F. Identification of World Health Organisation ship's medicine chest contents by Anatomical Therapeutic Chemical (ATC) classification codes. *Int. Marit. Health* **2017**, *68*, 39–45. [[CrossRef](#)]
17. Tang, S.; Chen, L. iATC-NFMLP: Identifying Classes of Anatomical Therapeutic Chemicals Based on Drug Networks, Fingerprints, and Multilayer Perceptron. *Curr. Bioinform.* **2022**, *17*, 814–824.
18. Kataria, S.; Ravindran, V. Electronic health records: A critical appraisal of strengths and limitations. *J. R. Coll. Physicians Edinb.* **2020**, *50*, 262–268. [[CrossRef](#)] [[PubMed](#)]
19. Conway, M.; Berg, R.L.; Carrell, D.; Denny, J.C.; Kho, A.N.; Kullo, I.J.; Linneman, J.G.; Pacheco, J.A.; Peissig, P.; Rasmussen, L.; et al. Analyzing the heterogeneity and complexity of Electronic Health Record oriented phenotyping algorithms. *AMIA Annu. Symp. Proc.* **2011**, *2011*, 274–283. [[PubMed](#)]
20. Madden, J.M.; Lakoma, M.D.; Rusinak, D.; Lu, C.Y.; Soumerai, S.B. Missing clinical and behavioral health data in a large electronic health record (EHR) system. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 1143–1149. [[CrossRef](#)]
21. Chauhan, V.K.; Thakur, A.; O'Donoghue, O.; Clifton, D.A. COPER: Continuous patient state perceiver. In Proceedings of the 2022 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Ioannina, Greece, 27–30 September 2022; pp. 1–4.
22. Wu, J.; Roy, J.; Stewart, W.F. Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. *Med. Care* **2010**, *48*, S106–S113. [[CrossRef](#)]
23. Huang, S.; Cai, N.; Pacheco, P.P.; Narrandes, S.; Wang, Y.; Xu, W. Applications of support vector machine (SVM) learning in cancer genomics. *Cancer Genom. Proteom.* **2018**, *15*, 41–51.

24. Svetnik, V.; Liaw, A.; Tong, C.; Culberson, J.C.; Sheridan, R.P.; Feuston, B.P. Random forest: A classification and regression tool for compound classification and QSAR modeling. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1947–1958. [[CrossRef](#)] [[PubMed](#)]
25. Zhou, Z.H.; Feng, J. Deep forest. *Natl. Sci. Rev.* **2019**, *6*, 74–86. [[CrossRef](#)] [[PubMed](#)]
26. Borup, D.; Christensen, B.J.; Mühlbach, N.S.; Nielsen, M.S. Targeting predictors in random forest regression. *Int. J. Forecast.* **2022**, *39*, 841–868. [[CrossRef](#)]
27. Futoma, J.; Morris, J.; Lucas, J. A comparison of models for predicting early hospital readmissions. *J. Biomed. Inform.* **2015**, *56*, 229–238. [[CrossRef](#)]
28. Solares, J.R.A.; Raimondi, F.E.D.; Zhu, Y.; Rahimian, F.; Canoy, D.; Tran, J.; Gomes, A.C.P.; Payberah, A.H.; Zottoli, M.; Nazarzadeh, M.; et al. Deep learning for electronic health records: A comparative review of multiple deep neural architectures. *J. Biomed. Inform.* **2020**, *101*, 103337. [[CrossRef](#)]
29. Zheng, L.; Wang, O.; Hao, S.; Ye, C.; Liu, M.; Xia, M.; Sabo, A.N.; Markovic, L.; Stearns, F.; Kanov, L.; et al. Development of an early-warning system for high-risk patients for suicide attempt using deep learning and electronic health records. *Transl. Psychiatry* **2020**, *10*, 1–10. [[CrossRef](#)]
30. Alam, A.; Reza, R.; Abrar, A.; Ahmed, T.; Ahmed, S.; Sharar, S.; Rasel, A.A. Patients' Severity States Classification based on Electronic Health Record (EHR) Data using Multiple Machine Learning and Deep Learning Approaches. *arXiv* **2022**, arXiv:2209.14907.
31. Nguyen, B.P.; Pham, H.N.; Tran, H.; Nghiem, N.; Nguyen, Q.H.; Do, T.T.; Tran, C.T.; Simpson, C.R. Predicting the onset of type 2 diabetes using wide and deep learning with electronic health records. *Comput. Methods Progr. Biomed.* **2019**, *182*, 105055. [[CrossRef](#)]
32. Ma, F.; You, Q.; Xiao, H.; Chitta, R.; Zhou, J.; Gao, J. Kame: Knowledge-based attention model for diagnosis prediction in healthcare. In Proceedings of the 27th ACM International Conference on Information and Knowledge Management, Torino, Italy, 22–26 October 2018; pp. 743–752.
33. Koumakis, L. Deep learning models in genomics; are we there yet? *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1466–1473. [[CrossRef](#)]
34. Beaulieu-Jones, B.K.; Moore, J.H.; CONSORTIUM, P.R.O.A.A.C.T. Missing data imputation in the electronic health record using deeply learned autoencoders. In Proceedings of the Pacific Symposium on Biocomputing 2017, Fairmont Orchid, HI, USA, 3–7 January 2017; pp. 207–218.
35. Sadati, N.; Nezhad, M.Z.; Chinnam, R.B.; Zhu, D. Representation learning with autoencoders for electronic health records: A comparative study. *arXiv* **2018**, arXiv:1801.02961.
36. Lee, D.; Yu, H.; Jiang, X.; Rogith, D.; Gudala, M.; Tejani, M.; Zhang, Q.; Xiong, L. Generating sequential electronic health records using dual adversarial autoencoder. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1411–1419. [[CrossRef](#)] [[PubMed](#)]
37. Biswal, S.; Ghosh, S.; Duke, J.; Malin, B.; Stewart, W.; Xiao, C.; Sun, J. EVA: Generating longitudinal electronic health records using conditional variational autoencoders. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Virtual, 6–7 August 2021; pp. 260–282.
38. Liao, W.; Derijks, H.J.; Blencke, A.A.; de Vries, E.; van Seyen, M.; van Marum, R.J. Dual autoencoders modeling of electronic health records for adverse drug event preventability prediction. *Intell.-Based Med.* **2022**, *6*, 100077. [[CrossRef](#)]
39. Wang, L.; Tong, L.; Davis, D.; Arnold, T.; Esposito, T. The application of unsupervised deep learning in predictive models using electronic health records. *BMC Med. Res. Methodol.* **2020**, *20*, 1–9. [[CrossRef](#)] [[PubMed](#)]
40. Suo, Q.; Ma, F.; Yuan, Y.; Huai, M.; Zhong, W.; Zhang, A.; Gao, J. Personalized disease prediction using a CNN-based similarity learning method. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Kansas City, MO, USA, 13–16 November 2017; pp. 811–816.
41. Wang, Y.H.; Nguyen, P.A.; Islam, M.M.; Li, Y.C.; Yang, H.C. Development of Deep Learning Algorithm for Detection of Colorectal Cancer in EHR Data. *MedInfo* **2019**, *264*, 438–441.
42. Kreinovich, V.; Phuong, N.H. *Soft Computing for Biomedical Applications and Related Topics*; Springer: Berlin/Heidelberg, Germany, 2021.
43. Yeh, M.C.H.; Wang, Y.H.; Yang, H.C.; Bai, K.J.; Wang, H.H.; Li, Y.C.J. Artificial intelligence-based prediction of lung cancer risk using nonimaging electronic medical records: Deep learning approach. *J. Med. Internet Res.* **2021**, *23*, e26256. [[CrossRef](#)]
44. Zhu, Z.; Yin, C.; Qian, B.; Cheng, Y.; Wei, J.; Wang, F. Measuring patient similarities via a deep architecture with medical concept embedding. In Proceedings of the 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, 12–15 December 2016; pp. 749–758.
45. Mikolov, T.; Sutskever, I.; Chen, K.; Corrado, G.S.; Dean, J. Distributed representations of words and phrases and their compositionality. *Adv. Neural Inf. Process. Syst.* **2013**, *26*, 3111–3119. [[CrossRef](#)]
46. Choi, E.; Bahadori, M.T.; Searles, E.; Coffey, C.; Thompson, M.; Bost, J.; Tejedor-Sojo, J.; Sun, J. Multi-layer representation learning for medical concepts. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1495–1504.
47. Choi, Y.; Chiu, C.Y.I.; Sontag, D. Learning low-dimensional representations of medical concepts. *AMIA Summits Transl. Sci. Proc.* **2016**, *2016*, 41.
48. Yang, Z.; Dehmer, M.; Yli-Harja, O.; Emmert-Streib, F. Combining deep learning with token selection for patient phenotyping from electronic health records. *Sci. Rep.* **2020**, *10*, 1432. [[CrossRef](#)]

49. Bai, T.; Chanda, A.K.; Egleston, B.L.; Vucetic, S. EHR phenotyping via jointly embedding medical concepts and words into a unified vector space. *BMC Med. Inform. Decis. Mak.* **2018**, *18*, 15–25. [[CrossRef](#)]
50. Steinberg, E.; Jung, K.; Fries, J.A.; Corbin, C.K.; Pfohl, S.R.; Shah, N.H. Language models are an effective representation learning technique for electronic health record data. *J. Biomed. Inform.* **2021**, *113*, 103637. [[CrossRef](#)] [[PubMed](#)]
51. Che, C.; Xiao, C.; Liang, J.; Jin, B.; Zho, J.; Wang, F. An rnn architecture with dynamic temporal matching for personalized predictions of parkinson’s disease. In Proceedings of the 2017 SIAM International Conference on Data Mining, Houston, TX, USA, 27–29 April 2017; pp. 198–206.
52. Lee, D.D.; Seung, H.S. Learning the parts of objects by non-negative matrix factorization. *Nature* **1999**, *401*, 788–791. [[CrossRef](#)] [[PubMed](#)]
53. Stein-O’Brien, G.L.; Arora, R.; Culhane, A.C.; Favorov, A.V.; Garmire, L.X.; Greene, C.S.; Goff, L.A.; Li, Y.; Ngom, A.; Ochs, M.F.; et al. Enter the matrix: Factorization uncovers knowledge from omics. *Trends Genet.* **2018**, *34*, 790–805. [[CrossRef](#)] [[PubMed](#)]
54. Esposito, F.; Boccarelli, A.; Del Buono, N. An NMF-Based methodology for selecting biomarkers in the landscape of genes of heterogeneous cancer-associated fibroblast Populations. *Bioinform. Biol. Insights* **2020**, *14*, 1177932220906827. [[CrossRef](#)]
55. Quintero, A.; Hübschmann, D.; Kurzawa, N.; Steinhäuser, S.; Rentzsch, P.; Krämer, S.; Andresen, C.; Park, J.; Eils, R.; Schlesner, M.; et al. ShinyButchR: Interactive NMF-based decomposition workflow of genome-scale datasets. *Biol. Methods Protoc.* **2020**, *5*, bpa022. [[CrossRef](#)]
56. Hassaine, A.; Canoy, D.; Solares, J.R.A.; Zhu, Y.; Rao, S.; Li, Y.; Zottoli, M.; Rahimi, K.; Salimi-Khorshidi, G. Learning multimorbidity patterns from electronic health records using non-negative matrix factorisation. *J. Biomed. Inform.* **2020**, *112*, 103606. [[CrossRef](#)] [[PubMed](#)]
57. Zhao, J.; Feng, Q.; Wu, P.; Warner, J.L.; Denny, J.C.; Wei, W.Q. Using topic modeling via non-negative matrix factorization to identify relationships between genetic variants and disease phenotypes: A case study of Lipoprotein (a)(LPA). *PLoS ONE* **2019**, *14*, e0212112. [[CrossRef](#)]
58. Huang, Y.; Pinto, M.D.; Borelli, J.L.; Mehrabadi, M.A.; Abrihim, H.; Dutt, N.; Lambert, N.; Nurmi, E.L.; Chakraborty, R.; Rahmani, A.M.; et al. COVID symptoms, symptom clusters, and predictors for becoming a long-hauler: Looking for clarity in the haze of the pandemic. *MedRxiv* **2021**. [[CrossRef](#)]
59. Hoyer, P.O. Non-negative matrix factorization with sparseness constraints. *J. Mach. Learn. Res.* **2004**, *5*, 1–13.
60. Haug, N.; Deischinger, C.; Gyimesi, M.; Kautzky-Willer, A.; Thurner, S.; Klimek, P. High-risk multimorbidity patterns on the road to cardiovascular mortality. *BMC Med.* **2020**, *18*, 1–12. [[CrossRef](#)]
61. Molaei, S.; Bousejin, N.G.; Zare, H.; Jalili, M.; Pan, S. Learning graph representations with maximal cliques. *IEEE Trans. Neural Networks Learn. Syst.* **2021**, *34*, 1089–1096. [[CrossRef](#)] [[PubMed](#)]
62. Si, Y.; Du, J.; Li, Z.; Jiang, X.; Miller, T.; Wang, F.; Zheng, W.J.; Roberts, K. Deep representation learning of patient data from Electronic Health Records (EHR): A systematic review. *J. Biomed. Inform.* **2021**, *115*, 103671. [[CrossRef](#)]
63. Xie, X.; Xiong, Y.; Yu, P.S.; Zhu, Y. Ehr coding with multi-scale feature attention and structured knowledge graph propagation. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management, Beijing, China, 3–7 November 2019; pp. 649–658.
64. Lu, C.; Reddy, C.K.; Ning, Y. Self-Supervised Graph Learning with Hyperbolic Embedding for Temporal Health Event Prediction. *IEEE Trans. Cybern.* **2021**, *51*, 1–14. [[CrossRef](#)] [[PubMed](#)]
65. Choi, E.; Xu, Z.; Li, Y.; Dusenberry, M.; Flores, G.; Xue, E.; Dai, A. Learning the graphical structure of electronic health records with graph convolutional transformer. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 606–613.
66. Park, J.; Cho, Y.; Lee, H.; Choo, J.; Choi, E. Knowledge graph-based question answering with electronic health records. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Virtual, 6–7 August 2021; pp. 36–53.
67. Shang, Y.; Tian, Y.; Zhou, M.; Zhou, T.; Lyu, K.; Wang, Z.; Xin, R.; Liang, T.; Zhu, S.; Li, J. EHR-Oriented Knowledge Graph System: Toward Efficient Utilization of Non-Used Information Buried in Routine Clinical Practice. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2463–2475. [[CrossRef](#)] [[PubMed](#)]
68. Wang, R.; Mou, S.; Wang, X.; Xiao, W.; Ju, Q.; Shi, C.; Xie, X. Graph structure estimation neural networks. In Proceedings of the Web Conference 2021, Ljubljana, Slovenia, 19–23 April 2021; pp. 342–353.
69. Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. *Phys. Nonlinear Phenom.* **2020**, *404*, 132306. [[CrossRef](#)]
70. Lipton, Z.C.; Kale, D.C.; Elkan, C.; Wetzell, R. Learning to diagnose with LSTM recurrent neural networks. *arXiv* **2015**, arXiv:1511.03677.
71. Ma, F.; Chitta, R.; Zhou, J.; You, Q.; Sun, T.; Gao, J. Dipole: Diagnosis prediction in healthcare via attention-based bidirectional recurrent neural networks. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 1903–1911.
72. Choi, E.; Schuetz, A.; Stewart, W.F.; Sun, J. Using recurrent neural network models for early detection of heart failure onset. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 361–370. [[CrossRef](#)]
73. Choi, E.; Bahadori, M.T.; Schuetz, A.; Stewart, W.F.; Sun, J. Doctor ai: Predicting clinical events via recurrent neural networks. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Los Angeles, CA, USA, 19–20 August 2016; pp. 301–318.
74. Choi, E.; Bahadori, M.T.; Sun, J.; Kulas, J.; Schuetz, A.; Stewart, W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Adv. Neural Inf. Process. Syst.* **2016**, *29*. [[CrossRef](#)]

75. Esteban, C.; Staeck, O.; Baier, S.; Yang, Y.; Tresp, V. Predicting clinical events by combining static and dynamic information using recurrent neural networks. In Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI), Chicago, IL, USA, 4–7 October 2016; pp. 93–101.
76. Liu, J.; Zhang, Z.; Razavian, N. Deep ehr: Chronic disease prediction using medical notes. In Proceedings of the Machine Learning for Healthcare Conference, PMLR, Palo Alto, CA, USA, 17–18 August 2018; pp. 440–464.
77. Ashfaq, A.; Sant’Anna, A.; Lingman, M.; Nowaczyk, S. Readmission prediction using deep learning on electronic health records. *J. Biomed. Inform.* **2019**, *97*, 103256. [[CrossRef](#)]
78. Gao, C.; Osmundson, S.; Edwards, D.R.V.; Jackson, G.P.; Malin, B.A.; Chen, Y. Deep learning predicts extreme preterm birth from electronic health records. *J. Biomed. Inform.* **2019**, *100*, 103334. [[CrossRef](#)]
79. Wang, Z.; Li, H.; Liu, L.; Wu, H.; Zhang, M. Predictive multi-level patient representations from electronic health records. In Proceedings of the 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), San Diego, CA, USA, 18–21 November 2019; pp. 987–990.
80. Gupta, M.; Phan, T.L.T.; Bunnell, H.T.; Beheshti, R. Obesity Prediction with EHR Data: A deep learning approach with interpretable elements. *ACM Trans. Comput. Healthc. Health* **2022**, *3*, 1–19. [[CrossRef](#)]
81. Bagheri, A.; Groenhof, T.K.J.; Veldhuis, W.B.; de Jong, P.A.; Asselbergs, F.W.; Oberski, D.L. Multimodal learning for cardiovascular risk prediction using EHR data. *arXiv* **2020**, arXiv:2008.11979.
82. He, H.; Henderson, J.; Ho, J.C. Distributed tensor decomposition for large scale health analytics. In Proceedings of the World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 659–669.
83. Zhao, J.; Zhang, Y.; Schlueter, D.J.; Wu, P.; Kerchberger, V.E.; Rosenbloom, S.T.; Wells, Q.S.; Feng, Q.; Denny, J.C.; Wei, W.Q. Detecting time-evolving phenotypic topics via tensor factorization on electronic health records: Cardiovascular disease case study. *J. Biomed. Inform.* **2019**, *98*, 103270. [[CrossRef](#)]
84. Afshar, A.; Perros, I.; Park, H.; Defilippi, C.; Yan, X.; Stewart, W.; Ho, J.; Sun, J. Taste: Temporal and static tensor factorization for phenotyping electronic health records. In Proceedings of the ACM Conference on Health, Inference, and Learning, Toronto, ON, Canada, 2–4 April 2020; pp. 193–203.
85. Hernandez, L.; Kim, R.; Tokcan, N.; Derksen, H.; Biesterveld, B.E.; Croteau, A.; Williams, A.M.; Mathis, M.; Najarian, K.; Gryak, J. Multimodal tensor-based method for integrative and continuous patient monitoring during postoperative cardiac care. *Artif. Intell. Med.* **2021**, *113*, 102032. [[CrossRef](#)] [[PubMed](#)]
86. Pollard, T.J.; Johnson, A.E.; Raffa, J.D.; Celi, L.A.; Mark, R.G.; Badawi, O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Sci. Data* **2018**, *5*, 1–13. [[CrossRef](#)] [[PubMed](#)]
87. Johnson, A.E.; Pollard, T.J.; Shen, L.; Lehman, L.W.H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; Mark, R.G. MIMIC-III, a freely accessible critical care database. *Sci. Data* **2016**, *3*, 1–9. [[CrossRef](#)] [[PubMed](#)]
88. Moody, G.B.; Mark, R.G.; Goldberger, A.L. PhysioNet: A web-based resource for the study of physiologic signals. *IEEE Eng. Med. Biol. Mag.* **2001**, *20*, 70–75. [[CrossRef](#)]
89. Choi, E.; Xu, Z.; Li, Y.; Dusenberry, M.W.; Flores, G.; Xue, Y.; Dai, A.M. Graph convolutional transformer: Learning the graphical structure of electronic health records. *arXiv* **2019**, arXiv:1906.04716.
90. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
91. Gardner, J.; Craven, M.; Dow, C.; Hines, E. The prediction of bacteria type and culture growth phase by an electronic nose with a multi-layer perceptron network. *Meas. Sci. Technol.* **1998**, *9*, 120. [[CrossRef](#)]
92. DeMaris, A. A tutorial in logistic regression. *J. Marriage Fam.* **1995**, *57*, 956–968. [[CrossRef](#)]
93. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
94. Daphne, K.; Nir, F.; Francis, B. Probabilistic graphical models: Principles and techniques. In *Adaptive Computation and Machine Learning*; MIT Press: Cambridge, MA, USA, 2009.
95. De’ath, G.; Fabricius, K.E. Classification and regression trees: A powerful yet simple technique for ecological data analysis. *Ecology* **2000**, *81*, 3178–3192. [[CrossRef](#)]
96. Laaksonen, J.; Oja, E. Classification with learning k-nearest neighbors. In Proceedings of the International Conference on Neural Networks (ICNN’96), Washington, DC, USA, 3–6 June 1996; Volume 3, pp. 1480–1483.
97. Boyd, K.; Eng, K.H.; Page, C.D. Area under the precision-recall curve: Point estimates and confidence intervals. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Grenoble, France, 19–23 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 451–466.
98. Hanley, J.A.; McNeil, B.J. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **1982**, *143*, 29–36. [[CrossRef](#)]
99. Davis, J.; Goadrich, M. Relationship between Precision-Recall and ROC Curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; pp. 233–240.
100. Tangri, N.; Ansell, D.; Naimark, D. Predicting technique survival in peritoneal dialysis patients: comparing artificial neural networks and logistic regression. *Nephrol. Dial. Transplant.* **2008**, *23*, 2972–2981. [[CrossRef](#)] [[PubMed](#)]
101. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Scholkopf, B. Support vector machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [[CrossRef](#)]

102. Van der Maaten, L.; Hinton, G. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 1–27.
103. Char, D.S.; Shah, N.H.; Magnus, D. Implementing machine learning in health care—Addressing ethical challenges. *N. Engl. J. Med.* **2018**, *378*, 981. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.