A Transformer-Based Model Trained on Large Scale Claims Data for Prediction of Severe COVID-19 Disease Progression

Manuel Lentzen, Thomas Linden, Sai Veeranki, Sumit Madan, Diether Kramer, Werner Leodolter, and Holger Fröhlich

***Abstract*—In situations like the COVID-19 pandemic, healthcare systems are under enormous pressure as they can rapidly collapse under the burden of the crisis. Machine learning (ML) based risk models could lift the burden by identifying patients with a high risk of severe disease progression. Electronic Health Records (EHRs) provide crucial sources of information to develop these models because they rely on routinely collected healthcare data. However, EHR data is challenging for training ML models because it contains irregularly timestamped diagnosis, prescription, and procedure codes. For such data, transformer-based models are promising. We extended the previously published Med-BERT model by including age, sex, medications, quantitative clinical measures, and state information. After pre-training on approximately 988 million EHRs from 3.5 million patients, we developed models to predict Acute Respiratory Manifestations (ARM) risk using the medical history of 80,211 COVID-19 patients. Compared to Random Forests, XGBoost, and RETAIN, our transformer-based models more accurately forecast the risk of developing ARM after COVID-19 infection. We used Integrated Gradients and Bayesian networks to understand the link between the essential features of our model. Finally, we evaluated adapting our model to Austrian in-patient data. Our study highlights the promise of predictive transformer-based models for precision medicine.**

***Index Terms*—COVID-19, precision medicine, transformer-based models.**

Manuscript received 12 December 2022; revised 17 May 2023; accepted 15 June 2023. Date of publication 22 June 2023; date of current version 6 September 2023. This work was supported by the Fraunhofer 'Internal Programs' through 'COPERIMOplus' initiative under Grant Anti-Corona 840266. *(Corresponding author: Holger Fröhlich.)*

Manuel Lentzen, Thomas Linden, and Holger Fröhlich are with the Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53757 Sankt Augustin, Germany, and also with the Bonn-Aachen International Center for IT (b-it), University of Bonn, 53115 Bonn, Germany (e-mail: manuel.lentzen@scai.fraunhofer.de; thomas.linden@scai.fraunhofer.de; holger.froehlich@scai.fraunhofer.de).

Sai Veeranki is with the Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), 8010 Graz, Austria, also with the Institute of Neural Engineering, Technical University, 8010 Graz, Austria, and also with the AIT Austrian Institute of Technology, 8010 Graz, Austria (e-mail: sai.veeranki@kages.at).

Sumit Madan is with the Department of Bioinformatics, Fraunhofer Institute for Algorithms and Scientific Computing (SCAI), Schloss Birlinghoven, 53757 Sankt Augustin, Germany, and also with the Institute of Computer Science, University of Bonn, 53115 Bonn, Germany (e-mail: sumit.madan@scai.fraunhofer.de).

Diether Kramer and Werner Leodolter are with the Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes), 8010 Graz, Austria (e-mail: diether.kramer@kages.at; werner.leodolter@kages.at).

This article has supplementary downloadable material available at https://doi.org/10.1109/JBHI.2023.3288768, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2023.3288768

I. Introduction

CORONAVIRUS disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus type 2 (SARS-CoV-2) that arose in December 2019. Since its emergence, 628 million people have been infected, and 6.58 million have died (https://coronavirus.jhu.edu/map.html, accessed 25.10.2022). In such pandemic circumstances, healthcare systems face a tremendous challenge as they can quickly collapse under the burden of this unprecedented crisis. Despite taking countermeasures such as testing, lockdowns, and vaccinations, the pandemic temporarily put immense stress on global healthcare systems. The use of decision support systems such as patient-level risk models can assist with the critical tasks of quickly and efficiently identifying high-risk patients so that the existing resources are best distributed and vulnerable patient subgroups are effectively protected.

Structured Electronic Health Records (EHRs) offer great opportunities for the efficient development of such risk models as they are routinely collected in many healthcare systems in large quantities. They contain data on diagnoses, prescriptions, procedures, and quantitative clinical measurements, such as vital values from bedside monitoring. Additionally, demographic data such as age, gender, and geographical region may be included. Models trained on such data could be used to better understand risk factors, such as comorbidities and medications, in addition to predicting a patient's risk of severe disease development. However, these data present significant challenges due to their high dimensionality, heterogeneity, temporal dependence, sparsity, and irregularity, making them difficult to fully exploit [1].

Furthermore, the coding of diagnoses is frequently biased for economic reasons. Since there is no unique mapping of a physician's diagnosis to a coding scheme such as ICD, there is a tendency to select the code that delivers the greatest economic benefit from among several possible codes. Concerning medications, it is noteworthy that categorization often occurs at the product level as opposed to the chemical substance level and that several medications may contain the same chemical substance.

TABLE I
COMPARISON OF EXMED-BERT WITH OTHER MODELS

	BEHRT	G-BERT	Med-BERT	BRLTM	ExMed-BERT
Pre-training dataset	CPRD	MIMIC-III	Cerner Health Facts	Private EHRs	IBM Explorys Therapeutic
Pre-training patients (No.)	1.6M	20K	20M	44K	3.5M
Vocabulary size	301	< 4K	82K	9,285	2,480
Input codes	Caliber	ICD-9, ATC	ICD9, ICD10	ICD9, CPT, Med	PheWas, ATC
Input structure	code + visit + age	code embeddings	code + visit + serialization	code + age + gender + position + segment	code + gender + state + age + visit
Pre-training task	Masked LM	Modified Masked LM	Masked LM + PLOS	Masked LM	Masked LM + PLOS
Evaluation task	Code prediction	Medication code prediction	Disease prediction	Prediction of depression over different time frames	Prediction of ARM following COVID-19
Publicly available	✗	✗	✗	✗	✓

In the past, many ML approaches have been taken to work with structured EHR data. Simpler methods often limited the time information and just worked with a one-hot encoding (OHE) of diagnoses and prescriptions, which allowed the application of standard ML techniques, such as logistic regression, random forest (RF), XGBoost (XGB), and Bayesian methods [2]. Recently, more studies focused on the use of time-series information. Methods for such an approach include autoencoders, convolutional neural networks [3], or sequential models like recurrent neural networks (RNN) [4] or transformer-based models [5], [6], [7], [8], [9]. Transformer-based models originate from natural language processing (NLP) and have recently gained much attention since they have achieved excellent results in many areas [10], [11], [12], [13]. A principal advantage of transformer models is the ability to train them in a parallel fashion and to weigh different parts of a time series differently due to their in-built attention mechanism. Transformer-based models typically undergo two-stage training: pre-training for generic representation learning and transfer learning (fine-tuning) for application-specific prediction. This approach enables sharing pre-trained models, often based on large datasets like the entire Wikipedia or protein sequences, with a broader community. These models can then be fine-tuned for various unforeseen tasks, highlighting the transformer-based approach's versatility and strength.

Variants of the Bidirectional Encoder Representations from Transformers (BERT) [14] model have recently been applied to structured EHR data. For instance, Shang et al. developed a graph-augmented transformer model named G-BERT to encode the medical history of single medical appointments and used the generated embeddings for a medication recommendation task [9]. Later, Li et al. developed BERT for EHR (BEHRT), which generated a patient embedding based on the history of diagnoses and used it for disease prediction in different time windows [5]. Since BEHRT – like most transformer-based models – is limited with respect to the maximum sequence length, the authors later developed a hierarchical BEHRT variant (HI-BEHRT), which can process longer medical histories [6]. Meng et al. presented another model in 2021, the Bidirectional Representation Learning model with a Transformer architecture on Multimodal EHR (BRLTM), which employed a strategy similar to BEHRT but incorporated a larger vocabulary, including diagnoses, medications, and procedures [8]. Med-BERT, another transformer-based model for structured EHR data, is closely related to BRLTM, but it features an even larger vocabulary and slightly different training objectives [7]. A comprehensive comparison of these models is provided in Table I. Unfortunately, none of the above-mentioned models is publicly available in a pre-trained form and thus not usable for the broader community.

Our contribution is an extension of the Med-BERT approach by including information about prescribed medications and demographic information such as state of residence, gender, and age as well as quantitative clinical measurements. We pre-trained our model, named ExMed-BERT, on 987,846,612 EHRs collected between 2010 and 2021, stemming from 3.5 million US patients in the IBM Explorys Therapeutic dataset. As a showcase, we subsequently used data from 80,211 COVID-19 patients to develop ML models for predicting the risk of acute respiratory manifestation (ARM) within three weeks after a confirmed COVID-19 diagnosis. This time frame was chosen because, on the one hand, a COVID-19 infection typically lasts 10 to 14 days. On the other hand, the timestamp of the COVID-19 diagnosis provided in the data may only be accurate up to a weekly resolution. The aim was thus to capture a serious event that could be time-wise related to the previously reported infection.

We compared our ExMed-BERT models with the three baseline models, which included the RNN-based RETAIN model [15], as well as two models (RF [16] and XGBoost [17]) that ignore time information. We then used explainable AI methods to gain insights into the underlying mechanisms of our models. A specific contribution is the use of Bayesian networks (BNs) to disentangle the relationship between most predictive features. Finally, we explored how our ExMed-BERT models could be adapted to external data from an Austrian hospital group (KAGes) via transfer learning strategies. Opposed to previous work, we make our ExMed-BERT model available to the scientific community.

II. MATERIALS & METHODS

A. General Overview

The work in this article consists of four phases (Fig. 1):

1) Pre-Training of transformer-based model for structured EHR data: Initially, we prepared a dataset of large-scale

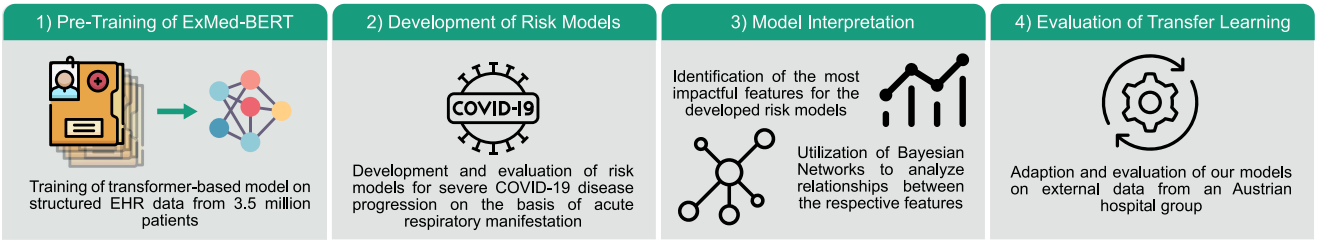


Fig. 1. Study overview. First, we pre-trained a transformer model on about 988 million EHR records from 3.5 million patients. Then, we developed patient-level risk models for COVID-19 disease progression. Next, we interpreted our developed risk models using Integrated Gradients in conjunction with Bayesian networks. Finally, we evaluated the possibility of adapting the models to external data.

claims data and pre-trained a transformer-based model called ExMed-BERT for structured EHR data.

2) Development of risk models for COVID-19 disease progression: Subsequently, we used our newly trained model to develop risk models for predicting severe COVID-19 disease progression – namely ARM – and compared their performances with RF, XGB, and RETAIN models.
3) Interpretation of developed risk models: Then, we used the Integrated Gradients approach in conjunction with Bayesian Networks to offer detailed explanations for model predictions.
4) Evaluation of the adaptation of our models to data obtained from an Austrian hospital group within a transfer learning approach.

In the following, we describe our approach in more detail.

B. Data Preprocessing

1) Preparation of Data for Modeling: This study used the IBM Explorys Therapeutic dataset (https://www.ibm.com/products/explorys-ehr-data-analysis-tools), which comprises EHRs and insurance claims from 4.5 million patients from all over the USA from 2010 until mid of 2021. Records consist of prescribed drugs, diagnoses, performed procedures, and a few quantitative clinical measures (e.g., blood pressure). We focused on demographic data and drugs, diagnoses, and available quantitative clinical measures. We excluded patients with fewer than five observations. This led to a reduced dataset of 3.5 million patients with 987,846,612 recorded diagnoses and drugs, which we used for pre-training a transformer model (details described later). The intent behind the pre-training of a transformer model is to learn a suitable vector representation of timestamped structured EHRs, irrespective of any later clinical use case. The fit of the model to a dedicated clinical endpoint is then performed within a subsequent fine-tuning/transfer learning step, for which we selected only patients with a confirmed COVID-19 diagnosis defined by the use of the International Classification of Diseases (ICD10) [18] code *U07.1* or a set of Logical Observation Identifier Names and Codes (LOINC) [19] codes (see Supplementary Section A) (n = 80,211). We corrected the diagnosis or observation dates of the records by subtracting seven days to get an approximation of the index date of infection. Then we focused on the ARM endpoint, which was defined if at least one of the following diagnoses appeared within three weeks after the COVID-19 infection was reported (n = 10,743):

- Pneumonia due to coronavirus disease 2019 (J12.82)
- Acute bronchitis due to other specified organisms (J20.8)
- Unspecified acute lower respiratory infection (J22)
- Bronchitis, not specified as acute or chronic (J40)
- Acute respiratory distress syndrome (J80)
- Respiratory failure, not elsewhere classified (J96)
- Other specified respiratory disorders (J98.8)

For fine-tuning, we used one year of medical history of the COVID-positive patients prior to their infection. Patients who fulfilled these criteria for the ARM endpoint were labeled as positives. Supplementary Fig. A.1 depicts the filtering process in further detail.

To identify negatives while adjusting for the potentially confounding effects of age and gender, we used the technique of Inverse Probability of Treatment Weighting (IPTW) [20], [21], [22]. We used the Python package *psmpy* [23] (version 0.2.8) to calculate propensity scores (PS), and subsequently, the IPTW weights for each patient sample were calculated by the following equation and used in the fine-tuning process.

$$\text{IPTW} = \begin{cases} \frac{1}{\text{PS}} & \text{if positive} \\ \frac{1}{1-\text{PS}} & \text{if negative} \end{cases} \quad (1)$$

2) Mapping of Drug and Diagnosis Codes: The IBM Explorys Therapeutic dataset includes information about diagnoses encoded as ICD9 and ICD10 codes and administered or prescribed drugs as RXNorm [24] identifiers. To harmonize the two versions of ICD diagnosis codes, we mapped them to Phecodes provided by the Phenome-wide association study (PheWAS) [25]. Due to the lower number of Phecodes, the problem of a non-unique mapping between a physician's diagnosis and the ICD coding scheme is reduced. Hence, we reduced potential coding biases and the feature space from 59,709 to 1,850 codes. Similarly, we mapped the provided RXNorm identifiers (RxCUI) to the fourth level of the Anatomical Therapeutic Chemical (ATC) [26] classification system for chemical compounds and thus addressed the sparse use of some RxCUIs by reducing the feature space from 23,801 to 630 codes.

3) Input Representation for the Models: For the Random Forest (RF) and XGBoost (XGB) models, we employed a one-hot encoding approach to represent all categorical features. In this scheme, a diagnosis or drug recorded in the one-year medical history was denoted as one and zero otherwise. We applied the same encoding technique to the state of residence and sex variables. However, the data formatting requirements for the RETAIN model and our transformer-based model significantly

differ from those of the RF and XGB baseline models. Owing to their design, which is tailored to handle sequential data, these models necessitate the representation of each patient's entire medical history as a sequence. As illustrated in the lower part of Fig. 2, we created separate sequences for each modality, with each element of the sequence being an integer corresponding to one vector of the embedding matrices. Quantitative clinical measures were only considered during the fine-tuning phase.

C. Model Training and Evaluation

In this section, we provide a comprehensive overview of the methods used in our study, detailing the pre-training and fine-tuning of various models across multiple experiments. We begin by outlining the structure and pre-training of our novel transformer-based model before describing the development of machine learning-based risk models using RF, XGBoost, RETAIN, and our new model. However, we note that recent models such as BEHRT or Med-BERT could not be included in our comparison as the pre-training data and models are not publicly available. Finally, we describe the integration of quantitative clinical measurements with the patient's medical history. While previous works have extensively described the model architectures, we refer interested readers to publications by Vaswani et al. and Devlin et al. for Transformer-based models [14], [27], Breiman for Random Forest [16], Chen and Guestrin for further details on the XGBoost approach [17], and Choi et al. for RETAIN [15].

1) Basic Model Structure and Pre-Training of ExMed-BERT: We followed a similar strategy as the Med-BERT article and focused on an extension of the BERT embedding layer. In addition to the diagnoses that were included in the Med-BERT model, we further extended Med-BERT by adding information on prescribed drugs, the patient's sex, state of residency, and age. We denote our model as Extended Med-BERT (ExMed-BERT). As shown in Fig. 2, we used different embeddings to accommodate the five feature modalities. Diagnoses and drugs were represented in one embedding via Phecodes and ATC codes. The sex and state embeddings contained static information. The age sequence contained the patients' age encoded in months, and lastly, the visit sequence was used to distinguish between each visit in a sequence. Since the order of drugs and diagnoses within one visit was random, we passed on a serialization embedding. Similar to Med-BERT, we did not use CLS and SEP tokens in our input sequences.

We used the same hyperparameters and training objectives as Med-BERT and pre-trained the model on the entire information of the 3.5 million patients in the pre-training cohort. If sequences exceeded the maximum sequence length of 512 diagnosis and drug codes, we split the sequences and processed the samples individually. We used the following joint training objectives to pre-train our model:

- *Masked language modeling (MLM):* This task is identical to the BERT approach and we followed the Med-BERT strategy in masking only one of the codes at a time. In 80% of the cases, the masked code was replaced with [MASK],

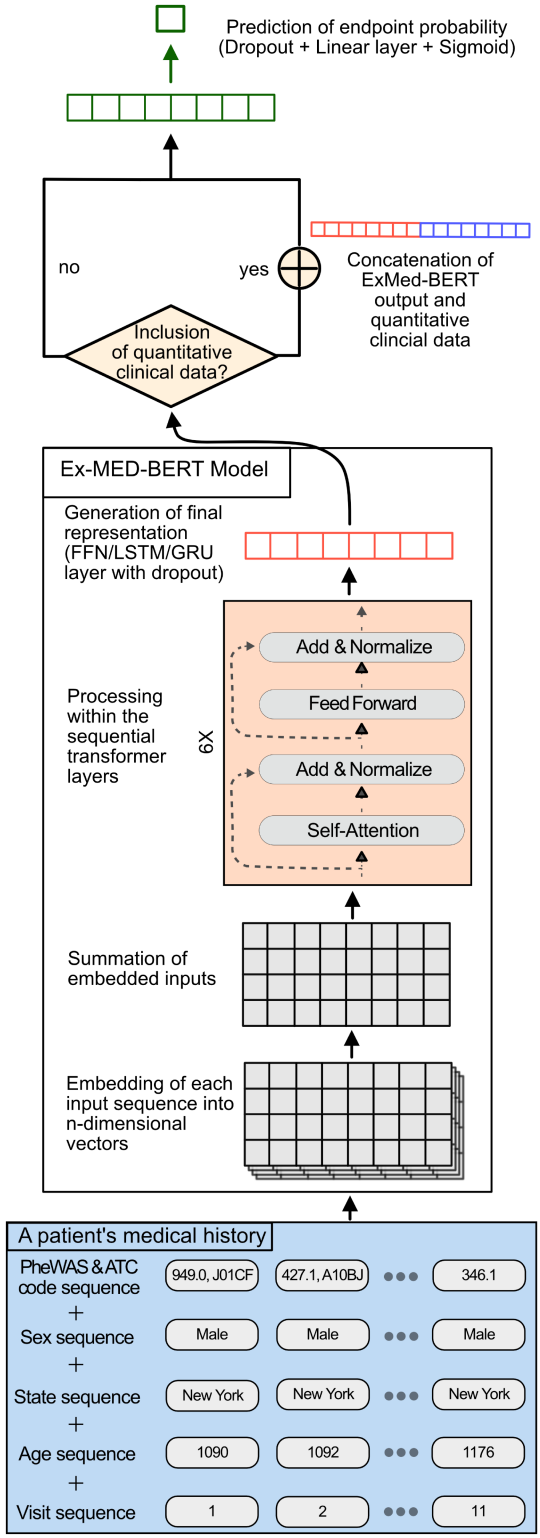


Fig. 2. Overview of the model structure. Compared to BERT or other transformer-based models, we employed a multimodal embedding layer for structured EHR data comprising drug, diagnosis and visit information and information about a patient's sex, state of residence, and age. After embedding, the input is passed through 6 transformer layers before a final representation of a patient's medical history is generated with an FFN, LSTM, or GRU head. Subsequently, these patient representations were either concatenated with the quantitative clinical data or directly passed through an FFN head for classification.

in 10% it was replaced with another code and in the remaining 10%, it remained unchanged. The model's task was to predict the correct code based on the information provided by the remaining sequence.

- *Prediction of prolonged length of stay (PLOS) in hospital:* As Rasmy et al. [7], we also predicted whether a patient had a prolonged stay in a hospital (>7 days) throughout his or her medical history. This task requires assessing the severity of a patient's health condition throughout their medical history.

2) Machine Learning-Based Risk Models: In this study, we aimed to predict severe COVID-19 disease progression by developing ML-based risk models to determine whether a patient would develop ARM within three weeks of their COVID-19 diagnosis. To achieve this, we utilized one year of the patient's medical history prior to diagnosis. To adjust for potential confounding effects of age and gender, we employed the IPTW approach described before.

During the fine-tuning of our ExMed-BERT model, we evaluated different classification head variants. We trained three different models using a feed-forward network (FFN), long short-term memory (LSTM), and gated recurrent unit (GRU) head. We split the data into training, validation, and testing sets in a stratified manner (70/10/20%) and used Bayesian hyperparameter optimization (*optuna* [28], version 2.10.0) to tune model parameters such as the learning rate, batch size, warmup ratio, weight decay, and in the case of RNNs, also the number of RNN layers. Similarly, we trained RF, XGB, and RETAIN classifiers and optimized several model hyperparameters. A detailed list of all optimized parameters can be found in Supplementary Table B.1.

For RETAIN, we used a Keras-based implementation[1] with slight modifications described in the next section. The input was similar to our ExMed-BERT model, using PheWAS and ATC codes and visit/date information but excluding a patient's age, sex, and state of residency, as these were not part of the original RETAIN approach.

3) Combination With Quantitative Clinical Measurements: In this study, we also assessed the integration of diagnosis and prescription codes with numerical clinical data, such as blood pressure readings. Our analysis focused on data documented in the two weeks leading up to the revised index date, excluding features with over 60% missing data. Given the considerable sparsity of this data, we were left with only eight features for our investigation: weight, body mass index (BMI), body surface area (BSA), height, body temperature, diastolic and systolic blood pressure, and heart rate. The number of patients with available numerical data for each feature is displayed in Table II (n = 23,949). To impute the numerical data for all patients, we employed a Random Forest (RF)-based approach, utilizing only the training data (*missingpy* [29], version 0.2.0).

The imputed numerical clinical features were combined with one-hot encoded (OHE) data for the RF experiments. In the case of the XGBoost (XGB) model, no prior imputation was carried out. Instead, we fused the numerical clinical features with the OHE data, relying on the imputation mechanism built into XGB.

[1][Online]. Available: https://github.com/Optum/retain-keras

TABLE II
CHARACTERISTICS OF THE PRE-TRAINING AND FINE-TUNING DATASETS

	Pre-training	Fine-tuning A	Fine-tuning B
Source	IBM Explorys Therapeutic Dataset		Austrian Hospital Group
Time span	2010 – 2021	2019 – 2021	2019 – 2021
No. of patients	3,478,438	80,211	6,335
ARM-positive	–	10,743	385
Patients with quantitative clinical data	–	23,949	–
Avg. patient age	41.2	49.8	47.9
Ratio of female patients (%)	56.3	62.1	51.8
Avg. number of records per patient	284.0	83.8	60.19
Total number of records	987,846,612	6,722,594	381,296

Note: ARM stands for acute respiratory manifestation.

In terms of model modifications, no alterations were required for the RF and XGB models. However, we needed to adjust the ExMed-BERT and RETAIN model architectures to accommodate numerical clinical features. Specifically, we utilized the same ExMed-BERT model as before to produce a patient's medical history embeddings. We then combined these vector-based representations with the numerical input before feeding it into a final classification head. The overall model architecture and the concatenation approach are detailed in Fig. 2. Similarly, we produced the latent encodings in the RETAIN model and combined them with the numerical clinical features before the classification layer.

D. Model Interpretation

1) Feature Importance: To better understand the ExMed-BERT models, we used the Integrated Gradients (IG) [30] approach to determine which drugs and diagnoses had the strongest influence on the model predictions. The IG method is an axiomatic model interpretability technique that awards, in the case of the ExMed-BERT models, an attribution score for each diagnosis or drug in the medical history. Next to an input sample ($x \in R^n$), the IG method requires a baseline input ($x' \in R^n$), which we constructed using a sequence of padding tokens. The IGs are then approximated by summing the gradients at points along the path from the specified baseline to the input using the following formula:

$$\mathrm{IG}_i(x)^{\mathrm{approx}} := (x_i - x_i') \times \int_0^1 \frac{\partial F(x' + \alpha \times (x - x'))}{\partial x_i} d\alpha \tag{2}$$

Here, F is a differentiable function ($F : R^n \Rightarrow [0, 1]$) that represents our ExMed-BERT model. We performed 50 steps to approximate the integrated gradients.

Initially, we computed IG attributions for all patients in the test dataset. Based on these, we calculated the mean absolute attribution for each diagnosis and drug that occurred at least ten percent of the time to identify the top features for each model. Subsequently, we calculated partial dependency scores using the top 20 features. To do so, we first calculated the probability for each patient for a specific endpoint using our fine-tuned ExMed-BERT models; we refer to this probability as p_r. The data for each of the top 20 features were then permuted individually by exchanging the respective diagnosis or drug codes with a *PAD* token. Subsequently, the modified data was used as input for our models to calculate the probability p_m. Finally, a fold change for each feature was calculated using the probabilities obtained for actual (p_r) and modified data (p_m) to estimate the effect (fold change; FC) of certain features on the model's prediction:

$$FC = \frac{p_r}{p_m} \tag{3}$$

2) Unraveling Feature Dependencies: To better comprehend the numerous interactions and dependencies between the most influential features, we developed BN models. BNs are probabilistic graphical models that can represent complex multivariate distributions with many variables. They can be graphically depicted with nodes representing random variables and edges expressing conditional statistical relationships. Let $G = (V, E)$ be a directed acyclic graph and $\{X_v | v \in V\}$ a set of random variables indexed over nodes in V. Then for any BN $B = (X, G)$:

$$p(X|G) = \prod_{v \in V} p\left(X_v | pa_v\right) \tag{4}$$

where pa_v denotes the parents of $v \in V$ according to the graph structure G. Because of their ability to model (potentially causal) relationships between variables, BNs are frequently employed in many areas of science, including system biology and medicine. In this work, we learned the graph structure G of a BN for the 100 most important features (according to the IG method) using the R package *bnlearn* [31] (version 4.7). We used a one-hot encoding for the respective features to indicate whether it was present in the one-year medical history, similar to the data preparation for the tree-based models. We also provided the patients' age, sex, and endpoint status. The tabu algorithm [32], [33] was used for BN structure learning. This was performed within a non-parametric bootstrap sampling scheme: We randomly subsampled n = 80,211 patients with replacement for 1000 times, and for each bootstrap sample we performed a complete network structure learning. We then focused on edges occurring in over half of the 1000 network architectures acquired from the non-parametric bootstrapped samples.

E. Transfer Learning on Austrian Hospital Data

1) Overview of the Data: Data from the Austrian hospital group consisted of pseudonymized in-patient records of 6,335 COVID-19-positive patients, out of which 385 suffered from ARM within a 3-week follow-up period after the initial visit to the hospital. The medication prescriptions were already encoded in ATC, but as ICD9/10 codes were used for diagnoses, these were mapped to Phecodes, akin to the procedure described earlier for the IBM Explorys dataset.

2) Transfer Learning of ExMed-BERT: We continued training the ExMed-BERT model for the ARM endpoint for only five epochs on the Austrian hospital data. This was done due to computational constraints. For the same reason, we did no substantial hyperparameter tuning but used the optimal hyperparameters discovered on the IBM Explorys data. We used 5-fold cross-validation to account for the small amount of available data. Alongside the ExMed-BERT model, we trained a new RF model for comparison.

III. Results

In this study, we predicted severe COVID-19 disease progression based on a patient's medical history. We begin by presenting the pre-training results of our newly created ExMed-BERT model. Then, we show the performances of the developed risk models, and lastly, we interpret our models using an explainable AI methodology.

A. Model Pre-Training

We utilized MLM and PLOS as training objectives for pre-training of the ExMed-BERT model. After 4.5 M steps (epoch 37), the MLM accuracy increased to around 51% and the PLOS F1 score to 70%. Following the inclusion of 61 missing ATC codes and the corresponding changes to the embedding, we began training for 750 K steps. Finally, we achieved an MLM accuracy of 67% and a PLOS F1-score of 66% (epoch 42, Supplementary Fig. B.1).

B. Evaluation of Risk Models

Following pre-training, we developed and evaluated risk models for predicting the ARM endpoint. Initially, we considered only the medical history without additional quantitative clinical measures. As shown in Table III, all ExMed-BERT models performed better than the RF, XGB, and RETAIN variants on unseen test data. Without quantitative clinical data, the ExMed-BERT models scored roughly 78% AUROC for the ARM endpoint, and the AUPR varied between 36.7% and 38.2%. The RF model, on the other hand, only achieved an AUROC of 73.4% and an AUPR of 29.1%. The XGB model had a slightly lower AUROC of 72.4% and AUPR of 28.2%. The RETAIN model achieved the lowest performance, with an AUROC of 68.5% and an AUPR of 26.8%.

The results of nearly all models improved when quantitative clinical measurements were integrated. The ExMed-BERT model with the GRU classification head integrating quantitative data gave the overall best result, with an AUROC of 79.8% and an AUPR of 38.7%, which is significantly higher than all other models.

When only patients with fully recorded quantitative clinical measurements were used, all models performed worse. That means the potential negative effect of imputing missing values was far less than the benefit of including additional data.

TABLE III
EVALUATION RESULTS OF THE RISK MODELS FOR PREDICTING ARM

Model	AUROC [%]	AUPR [%]
RF	73.4 [72.6, 74.3]	29.1 [27.6, 30.7]
+ Quant	77.7 [76.9, 78.6]	34.7 [33.0, 36.4]
subset w/o missingness	68.6 [67.1, 70.1]	40.1 [37.8, 42.6]
subset w/o missingness + Quant	70.2 [68.8, 71.6]	42.1 [39.7, 44.5]
XGB	72.4 [71.5, 73.3]	28.2 [26.7, 29.7]
+ Quant	77.7 [76.9, 78.5]	35.5 [33.8, 37.3]
subset w/o missingness	67.8 [66.3, 69.2]	38.8 [36.5, 41.2]
subset w/o missingness + Quant	70.6 [69.2, 72.0]	42.2 [39.7, 44.7]
RETAIN	68.5 [67.4, 69.5]	26.8 [25.3, 28.2]
+ Quant	66.3 [65.4, 67.4]	22.4 [21.3, 23.6]
subset w/o missingness	63.0 [61.5, 64.6]	35.8 [33.5, 38.1]
subset w/o missingness + Quant	63.3 [61.7, 64.8]	35.2 [33.0, 37.4]
ExMed-BERT-FFN	77.5 [76.7, 78.4]	38.2 [36.4, 40.0]
+ Quant	77.7 [76.8, 78.5]	38.1 [36.3, 39.8]
subset w/o missingness	67.6 [66.1, 69.1]	39.3 [36.9, 41.6]
subset w/o missingness + Quant	70.1 [68.7, 71.6]	41.9 [39.5, 44.4]
ExMed-BERT-GRU	77.7 [76.8, 78.6]	36.7 [35.0, 38.4]
+ Quant	**79.8 [78.9, 80.6]**	38.7 [36.9, 40.4]
subset w/o missingness	70.7 [69.3, 72.1]	42.8 [40.2, 45.4]
subset w/o missingness + Quant	72.0 [70.5, 73.5]	44.7 [42.1, 47.3]
ExMed-BERT-LSTM	77.7 [76.9, 78.6]	37.6 [35.9, 39.4]
+ Quant	78.4 [77.4, 79.3]	39.3 [37.4, 41.0]
subset w/o missingness	71.8 [70.5, 73.3]	**45.0 [42.4, 47.4]**
subset w/o missingness + Quant	71.4 [70.0, 72.8]	43.3 [40.8, 45.8]

Notes: Areas under the Receiver-Operator Characteristic Curve (AUROC) and the Precision-Recall Curve (AUPR) are shown for each model and feature modality. The best results per column are highlighted in bold. The suffix "+Quant" represents the additional use of quantitative clinical data during fine-tuning. The suffix "subset w/o missingness" indicates that we used a reduced subset (n = 23,949) for training and evaluation, where all quantitative clinical measures were available. The values in brackets indicate the 95% confidence interval, which we estimated by performing bootstrap resampling 1000 times.

C. Model Explanation

To better understand model predictions, we used an explainable AI methodology – namely IG – to calculate attribution scores for all features in the best-performing ExMed-BERT model. We calculated the IG attributions and used them to identify the 20 most important features by ranking them based on their mean absolute value. Fig. 3 shows all the IG attributions and FC scores, which are in agreement with each other. We found that the presence of diagnoses for chronic airway obstruction, congestive heart failure, cough, dementia, edema, obesity, shortness of breath, spondylosis, and type 2 diabetes in the medical history has a large impact on the prediction of a patient's risk for ARM. Similarly, the prescriptions of angiotensin II receptor blockers, biguanides, dihydropyridine derivatives, and thiazides have a substantial positive impact on our models' predictions.

Of course, these prescriptions and diagnoses could be correlated with each other, and thus, not all of them might have a direct impact on the ARM endpoint. Hence, we learned the graph structure of a BN to determine how the significant diagnoses or drugs could be related to one another. The overall network structure is provided as a *graphml* file, an XML-based data format for graph representation, as supplementary material to this article. Fig. 4 shows two excerpts of the BN graph structure. Fig. 4(a) focuses on Angiotensin II receptor blockers and their relationship to other drugs and diagnoses. Angiotensin II receptor blockers are used to treat hypertension, kidney diseases, and heart failure [34]. Furthermore, our graph shows a connection to essential hypertension and several ATC subgroups, namely ACE inhibitors, Dihydropyridine derivates, HMG CoA reductase inhibitors, and Thiazides.

TABLE IV
FEATURES WITH A SIGNIFICANT STATISTICAL EFFECT ON ARM

Feature Names	Corrected p-value
Type 2 diabetes	<0.0001
Heart failure with reduced EF	<0.0001
Shortness of breath	<0.0001
Constipation	<0.0001
Morbid obesity	<0.0001
Screening for malignant neoplasms	<0.0001
Dementias	<0.0001
Chronic airway obstruction	0,0003
Cough	0,0018
Screening for infectious and parasitic diseases	0,0083
Congestive heart failure (CHF) NOS	0,0119

Notes: We performed a logistic regression and corrected for the confounding variables age, sex, and state of residency. The p-values were corrected for multiple testing using the Holm-Sidak method.

Fig. 4(b) depicts morbid obesity and other diagnoses and drugs in its immediate neighborhood. There is a link to the class of Biguanides, which includes the drug Metformin, commonly used to treat diabetes [35]. Furthermore, morbid obesity is linked to hypertension, type 2 diabetes, obstructive sleep apnea, and obesity.

We aimed for an understanding of the statistical and potential causal effects of those features on the endpoint, which were either among the 20 most important features or sink nodes in the BN. The latter are nodes without outgoing connections and, therefore, do not influence any other features, according to our BN analysis. For each of those features, we performed a univariate logistic regression analysis while using IPTW case weights to correct for potential confounding effects of age and gender. Our analysis shows significant effects of several prior diagnoses on the ARM onset, namely, type 2 diabetes, obesity, dementia, cardiovascular diseases, and respiratory diseases (see Table IV). These morbidities have previously been reported as risk factors for severe COVID-19 disease progression [36], [37], [38], [39], [40], [41], and also the underlying molecular mechanisms have been discussed [42], [43]. Besides, we found significant effects between constipation, screening for malignant neoplasms, and infectious/parasitic diseases and the ARM onset. This might be explained by the fact that such procedures are more frequently executed in older patients with bad health conditions, resulting in a higher risk of severe COVID-19 progression. In the same type of patients, constipation is also a frequent problem, e.g., due to lifestyle.

D. Transfer Learning on Austrian Hospital Data

Fine-tuning of Ex-MedBERT on a small set of pseudonymized in-patient data from an Austrian hospital group resulted in a prediction performance almost identical to the one observed for an RF trained de novo on the same data (Supplementary Table B.2). At the same time, prediction performances were significantly lower than the ones observed on IBM Explorys

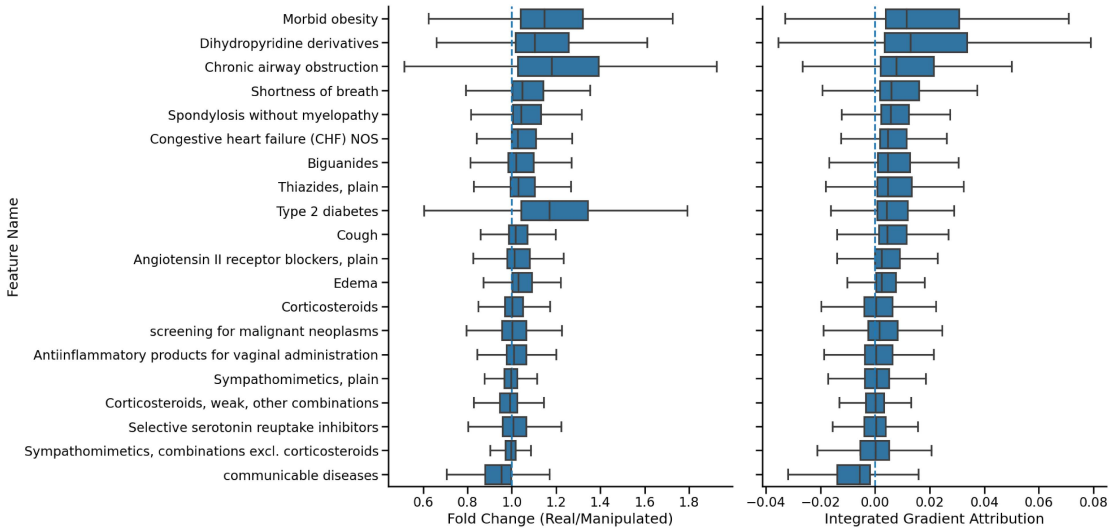
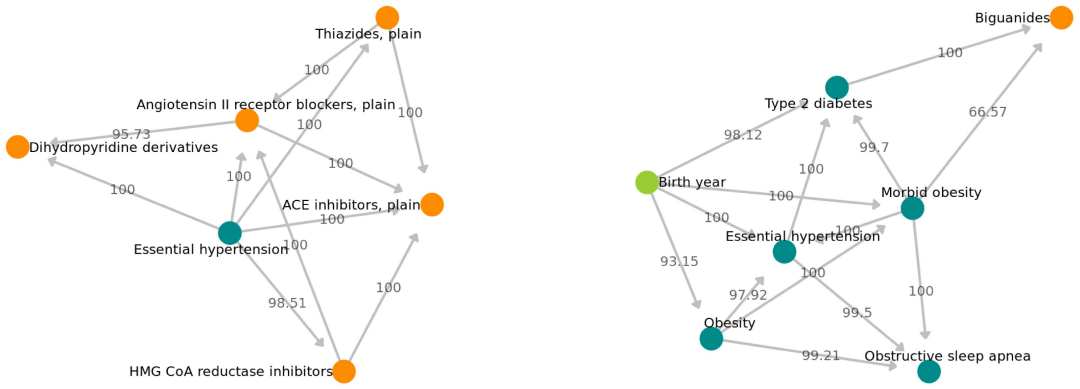


Fig. 3. Integrated Gradients Attributions for ExMed-BERT GRU. Depicted are all calculated fold changes (FC) and IG attributions for the 20 most important features for the prediction of ARM onset. The dashed blue lines indicate neutral attributions. Everything greater than the neutral value positively affects the prediction and vice versa.



(a) Angiotensin II receptor blockers (b) Morbid obesity

Fig. 4. Selected features and the direct neighborhood in the inferred Bayesian network. The numbers indicate the bootstrap strength of the respective edges in percentage. That means a bootstrap strength of 100 indicates that the corresponding edge has been found in each of the 1000 BN reconstructions learned from different bootstrap samples.

(AUC $\approx 60\,\%$). We will elaborate on potential reasons in the subsequent discussion.

IV. Discussion

Pandemics such as COVID-19 pose immense challenges to global healthcare systems. Utilizing patient-level risk models to support doctors and clinics is one way to maximize the use of available resources. Following previous research, we trained a transformer-based model on structured EHR data in this study. In contrast to the prior approaches, such as BEHRT or Med-BERT, we incorporated additional data modalities and developed risk models for COVID-19 disease progression. Prediction performances achieved by our ExMed-BERT model are altogether superior to those reported by Lazzarini et al. [44] for the closely related endpoint of acute respiratory distress syndrome (ARDS) [44]. The authors trained an XGB based on US administrative claims data from 290,000 patients and achieved an AUROC of 69% and an AUPR of 7%. For comparison, using data from intensive care units (ICUs), Bendavid et al. reported an AUROC of 83% for an XGB trained to predict the initiation of invasive mechanical ventilation [45], and Singhal et al. achieved an AUROC of 89% for predicting the onset of ARDS [46]. Importantly, ICU data are structurally and content-wise very different from the data used in our study, which comprises in-patient as well as out-patient information over a more extended period (here: one year), but only contains limited quantitative information. Altogether, our findings align with previous studies [5], [7], [8], showing that transformer-based models are well-suited for structured EHR data similar to ours. Even without additional quantitative information, our ExMed-BERT outperformed the RF, XGB, and RETAIN

models. With the inclusion of quantitative clinical measures, our ExMed-BERT models further increased in prediction performance. For that purpose, we proposed a novel approach to combine quantitative clinical measures with the embeddings of EHR codes learned by ExMed-BERT, which resulted in the overall best-performing model. Our results thus demonstrate the importance of combining diagnosis and prescription codes with quantitative clinical measures for developing risk models. Even though these quantitative clinical measures were only taken at a single time point in the two weeks preceding the COVID-19 infection and not for every patient, using these data could provide better performance.

Using a combined strategy consisting of feature importance analysis, BN structure learning and statistical hypothesis testing, we were able to identify diagnoses and prescriptions that have a significant impact on model prediction and may causally influence the endpoint. Our analysis supports that socioeconomic and psycho-social health risks play an important role in addition to well-known risk factors such as obesity, diabetes, cardiovascular diseases, and dementia, which have already been reported as known risk factors for severe COVID-19 disease progression in several studies [36], [39], [47], [48], [49]. This confirms the validity of our approach, which can be applied to other datasets as well.

Our work demonstrates the potential of a transformer-based pre-training/fine-tuning strategy to develop risk models for precision medicine. This strategy provides the chance to perform transfer learning of our model on data from other organizations and thus use the pre-trained ExMed-BERT as a basis for future model development. Our experiment with data from an Austrian hospital group demonstrated the potential as well as the limitations of such an approach: The data from the Austrian hospital group only comprises in-patient information, and the number of patients is far smaller than during the fine-tuning phase on the IBM Explorys data (6,335 patients instead of 80,211). Furthermore, the ratio of ARM-positive patients is significantly lower (6.1% instead of 13.4%). Notably, there could also be different medical coding practices in the two countries. Finally, constraints on the technical equipment within the Austrian hospital group only allowed us to fine-tune our model for a small number of epochs and without hyperparameter tuning. Due to all these factors, our ExMed-BERT model fine-tuned on the Austrian data achieved a performance that was comparable to an RF model trained de novo on the same data but significantly lower than prediction performances achieved on US data. We thus conclude that having a sufficiently large dataset with a number of patients in a range comparable to the IBM Explorys data would be a prerequisite to obtaining better models in a transfer learning setting. Furthermore, appropriate technical equipment is important. Finally, the integration of in-patient and out-patient data is required, at least for our model.

Another limitation is the lack of previously published transformer-based models, such as BEHRT or Med-BERT, and the associated data, which hindered direct comparison with our model. As the pre-training of transformer-based models is computationally extremely expensive, it is often not feasible to run comprehensive ablation studies. Despite these limitations, our model was rigorously assessed by comparing it against established approaches (RF, XGBoost, and RETAIN), and its potential was adequately demonstrated. By making our model publicly available, future studies can use it as a foundation for further development.

V. Conclusion

Our work demonstrates the potential of customized transformer-based models for analyzing structured EHR data. We showed that it is possible to integrate quantitative clinical data into such models, which can significantly improve prediction performance. Furthermore, we introduced a general approach for explaining ExMed-BERT model predictions. Transfer learning strategies open the possibility of leveraging our pre-trained ExMed-BERT model for the prediction of clinical endpoints different from the one addressed within this article. For that purpose, we allow users to apply for access to our pre-trained ExMed-BERT model on https://doi.org/10.5281/zenodo.7324178 or by sending an email to the corresponding author. Our code is available at https://github.com/SCAI-BIO/ExMed-BERT.

Acknowledgment

The COPERIMOplus Consortium: Fraunhofer Data Protection Office (Anne Funck Hansen), Fraunhofer IAIS (Sabine Kugler, Stefan Rüping), Fraunhofer IGD (Jan Burmeister, Jörn Kohlhammer), Fraunhofer IKTS (George Sarau, Silke Christiansen), Fraunhofer IME (Oliver Keminer), Fraunhofer ITMP (Aimo Kannt, Andrea Zaliani, Ann Christina Foldenauer, Carsten Claussen, Eduard Resch, Kevin Frank), Fraunhofer MEVIS (Hendrik Laue, Horst Hahn, Jochen Hirsch, Marco Wischnewski, Matthias Günther, Saulius Archipovas), Fraunhofer SCAI (Alpha Tom Kodamullil, Andre Gemünd, Bruce Schultz, Carina Steinborn, Christian Ebeling, Daniel Domingo Fernández, Helena Hermanowski, Holger Fröhlich, Jürgen Klein, Manuel Lentzen, Marc Jacobs, Martin Hofmann-Apitius, Meike Knieps, Michael Krapp, Philipp Johannes Wendland, Philipp Wegner, Sepehr Golriz Khatami, Stephan Springstubbe, Thomas Linden), ZB MED Information Centre for Life Sciences (Juliane Fluck).

Conflict of Interest Statement: SV, DK, and WL received salaries from Steiermärkische Krankenanstaltengesellschaft m.b.H. (KAGes) (Graz, Austria). The company had no influence on the scientific results presented in this article.

Contributions: Initiated and supervised project: HF; programmed code and conducted experiments: ML, TL, SV, DK; supervised transfer learning on Austrian hospital data: WL[†]. Drafted the manuscript: ML, HF. All authors have read and approved the current version of the article.

References

[1] R. Miotto, F. Wang, S. Wang, X. Jiang, and J. T. Dudley, "Deep learning for healthcare: Review, opportunities and challenges," *Brief. Bioinf.*, vol. 19, no. 6, pp. 1236–1246, Nov. 2018.

[2] B. A. Goldstein, A. M. Navar, M. J. Pencina, and J. P. A. Ioannidis, "Opportunities and challenges in developing risk prediction models with electronic health records data: A systematic review," *J. Amer. Med. Inform. Assoc.: JAMIA*, vol. 24, no. 1, pp. 198–208, Jan. 2017.

[3] T. Linden, J. D. Jong, C. Lu, V. Kiri, K. Haeffs, and H. Fröhlich, "An explainable multimodal neural network architecture for predicting epilepsy comorbidities based on administrative claims data," *Front. Artif. Intell.*, vol. 4, 2021, Art. no. 610197. [Online]. Available: https://www.frontiersin.org/article/10.3389/frai.2021.610197

[4] L. Rasmy et al., "Recurrent neural network models (CovRNN) for predicting outcomes of patients with COVID-19 on admission to hospital: model development and validation using electronic health record data," *Lancet Digit. Health.*, vol. 4, no. 6, pp. e415–e425, Jun. 2022, doi: 10.1016/S2589-7500(22)00049-8.

[5] Y. Li et al., "BEHRT: Transformer for electronic health records," *Sci. Rep.*, vol. 10, no. 1, Apr. 2020, Art. no. 7155. [Online]. Available: https://www.nature.com/articles/s41598-020-62922-y

[6] Y. Li et al., "Hi-BEHRT: Hierarchical transformer-based model for accurate prediction of clinical events using multimodal longitudinal electronic health records," *IEEE J. Biomed. Health Inf.*, vol. 27, no. 2, pp. 1106–1117, 2022.

[7] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, and D. Zhi, "Med-BERT: Pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction," *npj Digit. Med.*, vol. 4, no. 1, pp. 1–13, May 2021. [Online]. Available: https://www.nature.com/articles/s41746-021-00455-y

[8] Y. Meng, W. Speier, M. K. Ong, and C. W. Arnold, "Bidirectional representation learning from transformers using multimodal electronic health record data to predict depression," *IEEE J. Biomed. & Health Inf.*, vol. 25, no. 8, pp. 3121–3129, Aug. 2021.

[9] J. Shang, T. Ma, C. Xiao, and J. Sun, "Pre-training of graph augmented transformers for medication recommendation," in *Proc. Int. Joint Conf. Artif. Intell.*, 2019. [Online]. Available: https://www.ijcai.org/proceedings/2019/

[10] J. Lee et al., "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, no. 4, pp. 1234–1240, Feb. 15, 2020, doi: 10.1093/bioinformatics/btz682.

[11] A. Elnaggar et al., "ProtTrans: Towards cracking the language of life's code through self-supervised deep learning and high performance computing," May 2021, *arXiv:2007.06225*.

[12] Y. Ji, Z. Zhou, H. Liu, and R. V. Davuluri, "DNABERT: Pre-trained bidirectional encoder representations from transformers model for DNA-language in genome," *Bioinformatics*, vol. 37, no. 15, pp. 2112–2120, Aug. 2021. [Online]. Available: https://academic.oup.com/bioinformatics/article/37/15/2112/6128680

[13] S. Madan, V. Demina, M. Stapf, O. Ernst, and H. Fröhlich, "Accurate prediction of virus-host protein-protein interactions via a siamese neural network using deep protein sequence embeddings," *Patterns*, vol. 3, no. 9, Sep. 2022, Art. no. 100551. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389922001568

[14] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol.*, Minneapolis, Minnesota: Assoc. Comput. Linguistics, vol. 1, Jun. 2019, pp. 4171–4186. [Online]. Available: https://aclanthology.org/N19-1423

[15] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun, "RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism," in *NIPS'16: Proc.30th Int. Conf. Neural Inf. Process. Syst.*, Red Hook, NY, USA: Curran Associates Inc., 2016.

[16] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[17] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. & Data Mining*, 2016, pp. 785–794, doi: 10.1145/2939672.2939785.

[18] World Health Organization, "ICD-10 : International statistical classification of diseases and related health problems : Tenth revision," World Health Organization, Tech. Rep., 2004. [Online]. Available: https://apps.who.int/iris/handle/10665/42980

[19] C. J. McDonald et al., "LOINC, a universal standard for identifying laboratory observations: A 5-Year update," *Clin. Chem.*, vol. 49, no. 4, pp. 624–633, Apr. 2003. [Online]. Available: https://academic.oup.com/clinchem/article/49/4/624/5641953

[20] P. R. Rosenbaum and D. B. Rubin, "The central role of the propensity score in observational studies for causal effects," *Biometrika*, vol. 70, no. 1, pp. 41–55, 1983. [Online]. Available: https://dash.harvard.edu/handle/1/3382855

[21] P. R. Rosenbaum, "Model-based direct adjustment," *J. Amer. Stat. Assoc.*, vol. 82, no. 398, pp. 387–394, Jun. 1987. [Online]. Available: https://www.tandfonline.com/doi/abs/10.1080/01621459.1987.10478441

[22] P. C. Austin and M. M. Mamdani, "A comparison of propensity score methods: A case-study estimating the effectiveness of post-AMI statin use," *Statist. Med.*, vol. 25, no. 12, pp. 2084–2106, 2006, [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.2328

[23] A. Kline and Y. Luo, "PsmPy: A package for retrospective cohort matching in python," in *Proc. 44th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2022, pp. 1354–1357, doi: 10.1109/EMBC48229.2022.9871333.

[24] S. J. Nelson, K. Zeng, J. Kilbourne, T. Powell, and R. Moore, "Normalized names for clinical drugs: RxNorm at 6 years," *J. Amer. Med. Inform. Assoc.*, vol. 18, no. 4, pp. 441–448, Jul. 2011. [Online]. Available: https://academic.oup.com/jamia/article-lookup/doi/10.1136/amiajnl-2011-000116

[25] P. Wu et al., "Mapping ICD-10 and ICD-10-CM codes to phecodes: Workflow development initial evaluation," *JMIR Med. Inform.*, vol. 7, no. 4, Nov. 29, 2019, Art. no. e14325. doi: 10.2196/14325.

[26] "WHO collaborating centre for drug statistics methodology, ATC classification index with DDDs," 2022. [Online]. Available: https://www.whocc.no/atc_ddd_index_and_guidelines/atc_ddd_index/

[27] A. Vaswani et al., "Attention Is All You Need," *Adv. Neural Inf. Process. Syst.*, I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, Dec. 2017.

[28] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25rd ACM SIGKDD Int. Conf. Knowl. Discov. Data Mining*, 2019, pp. 2623–2631.

[29] D. J. Stekhoven and P. Bühlmann, "MissForestnon-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2011.

[30] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," Jun. 2017, *arXiv:1703.01365*.

[31] M. Scutari, "Learning Bayesian networks with the bnlearn R package," *J. Stat. Softw.*, vol. 35, no. 3, pp. 1–22, 2010.

[32] F. Glover and C. McMillan, "The general employee scheduling problem. an integration of MS and AI," *Comput. Operations Res.*, vol. 13, no. 5, pp. 563–573, Jan. 1986. [Online]. Available: https://www.sciencedirect.com/science/article/pii/030505488690050X

[33] F. Glover, "Tabu search—Part I," *ORSA J. Comput.*, vol. 1, no. 3, pp. 190–206, 1989.

[34] Z. H. Israili, "Clinical pharmacokinetics of angiotensin II (AT1) receptor blockers in hypertension," *J. Hum. Hypertension*, vol. 14, no. 1, pp. S73–S86, Apr. 2000. [Online]. Available: https://www.nature.com/articles/1000991

[35] R. Vigneri and I. D. Goldfine, "Role of metformin in treatment of diabetes mellitus," *Diabetes Care*, vol. 10, no. 1, pp. 118–122, Jan. 1987. [Online]. Available: https://diabetesjournals.org/care/article/10/1/118/790/Role-of-metformin-in-treatment-of-diabetes

[36] W.-J. Guan et al., "Comorbidity and its impact on 1590 patients with COVID-19 in China: A nationwide analysis," *Eur. Respir. J.*, vol. 55, no. 5, May 2020, Art. no. 2000547, doi: 10.1183/13993003.00547-202.

[37] A. C. Tahira, S. Verjovski-Almeida, and S. T. Ferreira, "Dementia is an age-independent risk factor for severity and death in COVID-19 inpatients," *Alzheimer's Dement.*, vol. 17, no. 11, pp. 1818–1831, 2021. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/alz.12352

[38] S. Toniolo et al., "Dementia and COVID-19, a bidirectional liaison: Risk factors, biomarkers, and optimal health care," *J. Alzheimer's Dis.*, vol. 82, no. 3, pp. 883–898, Jan. 2021. [Online]. Available: https://content.iospress.com/articles/journal-of-alzheimers-disease/jad210335

[39] S. Peric and T. M. Stulnig, "Diabetes and COVID-19," *Wiener klinische Wochenschrift*, vol. 132, no. 13, pp. 356–361, Jul. 2020, doi: 10.1007/s00508-020-01672-3.

[40] F. Demeulemeester, K. d. Punder, M. v. Heijningen, and F. v. Doesburg, "Obesity as a risk factor for severe COVID-19 and complications: A review," *Cells*, vol. 10, no. 4, Apr. 2021, Art. no. 933. [Online]. Available: https://www.mdpi.com/2073-4409/10/4/933

[41] S. Kwok et al., "Obesity: A critical risk factor in the COVID-19 pandemic," *Clin. Obesity*, vol. 10, no. 6, 2020, Art. no. e12403, https://onlinelibrary.wiley.com/doi/pdf/10.1111/cob.12403

[42] R. Concha, E. Ohayon, and A. Lam, "Neuroinflammation in COVID-19 and ADRD: Similarities, differences, and interactions," *Alzheimer's Dement.*, vol. 17, no. S3, 2021, Art. no. e056282. [Online]. Available: https://onlinelibrary.wiley.com/doi/pdf/10.1002/alz.056282

[43] C. Steenblock et al., "COVID-19 and metabolic disease: Mechanisms and clinical management," *Lancet. Diabetes Endocrinol.*, vol. 9, no. 11, pp. 786–798, Nov. 2021.

[44] N. Lazzarini, A. Filippoupolitis, P. Manzione, and H. Eleftherohorinou, "A machine learning model on real world data for predicting progression to acute respiratory distress syndrome (ARDS) among COVID-19 patients," *PLoS One*, vol. 17, no. 7, 2022, Art. no. e0271227.

[45] I. Bendavid et al., "A novel machine learning model to predict respiratory failure and invasive mechanical ventilation in critically ill patients suffering from COVID-19," *Sci. Rep.*, vol. 12, no. 1, Jun. 2022, Art. no. 10573. [Online]. Available: https://www.nature.com/articles/s41598-022-14758-x

[46] L. Singhal et al., "eARDS: A multi-center validation of an interpretable machine learning algorithm of early onset acute respiratory distress syndrome (ARDS) among critically ill adults with COVID-19," *PLos One*, vol. 16, no. 9, Sep. 2021, Art. no. e0257056. [Online]. Available: https://journals.plos.org/plosone/article?id=, doi: 10.1371/journal.pone.0257056.

[47] Y. Du et al., "Clinical features of 85 fatal cases of COVID-19 from wuhan. a retrospective observational study," *Amer. J. Respir. Crit. Care Med.*, vol. 201, no. 11, pp. 1372–1379, Jun. 2020. [Online]. Available: https://www.atsjournals.org/doi/10.1164/rccm.202003-0543OC

[48] B. Li et al., "Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China," *Clin. Res. Cardiol.*, vol. 109, no. 5, pp. 531–538, May 2020, doi: 10.1007/s00392-020-01626-9.

[49] T. Linden et al., "Machine learning based prediction of COVID-19 mortality suggests repositioning of anticancer drug for treating severe cases," *Artif. Intell. Life Sci.*, vol. 1, Dec. 2021, Art. no. 100020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2667318521000209