



MS-LSTMEA: Predicting clinical events for Hypertension using Multi-Sources LSTM Explainable Approach

Farida Zehraoui, Naziha Sendi, Nadia Abchiche-Mimouni

► To cite this version:

Farida Zehraoui, Naziha Sendi, Nadia Abchiche-Mimouni. MS-LSTMEA: Predicting clinical events for Hypertension using Multi-Sources LSTM Explainable Approach. 2022. hal-03812885

HAL Id: hal-03812885

<https://hal.science/hal-03812885>

Preprint submitted on 13 Oct 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Laboratoire IBISC, univ. Evry, université Paris-Saclay

Rapport de Recherche

MS-LSTMEA: Predicting clinical events for Hypertension using Multi-Sources LSTM Explainable Approach

Farida Zehraoui*, Naziha Sendi+, Nadia Abchiche-Mimouni*

*Laboratoire IBISC

+Visiomed Group

Octobre 2022

RR-IBISC_2022-035

MS-LSTMEA: Predicting clinical events for Hypertension using Multi-Sources LSTM Explainable Approach

Farida Zehraoui^a, Naziha Sendi^b, Nadia Abchiche-Mimouni^a

^a*Laboratoire IBISC Univ. Evry, Université Paris-Saclay, France, 23 Bd de France, Evry, 91034, France*

^b*VISIOMED group, 1 Avenue du Général de Gaulle, Paris, 92800, France*

Abstract

Hypertension is a major risk factor for cardiovascular disorders and diseases. Since it is expected to increase dramatically, effective Hypertension management becomes more and more critical. Early detection of patients with uncontrolled Hypertension would allow to employ personalized medicine for anticipating and providing the adequate drug class. In this paper, we propose MS-LSTMEA, a Multi-Source and Explainable Hypertension prediction Approach based on Long Short-Term Memory algorithm, for predicting both the drug class for patients and the date of their next medical appointment. MS-LSTMEA can successfully combine different sources of medical information about patients, represented by tabular data, while processing them separately and differently, and taking into account the temporal aspect of Electronic Health Records. In addition, it integrates an Attention mechanism which allows to improve the model results and to explain its outcomes.

Email addresses: farida.zehraoui@univ-evry.fr (Farida Zehraoui),
sendi.neziha@gmail.com (Naziha Sendi), nadia.abchichemimouni@univ-evry.fr
(Nadia Abchiche-Mimouni)

Experiments have been conducted on real Electronic Health Records data from 429,087 patients with the support of experts in healthcare. The results show significant improvements in accuracy over several existing models.

Keywords: Deep Learning, Explainable Artificial Intelligence, Attention mechanism, Hypertension, Electronic Health Records

Introduction

Hypertension is a major factor behind one of the most important causes of death in the world due to Cardiovascular diseases [1]. Improving upon the detection and the control of Hypertension, and working towards its early detection is of great importance in reducing the impact of cardiovascular diseases, their progression, and the associated risks. In order to confirm the diagnostic of Hypertension, repeated medical measurements are collected from patients.

With the availability of home kits for self-measurement allowing patients to be involved in the management of their disease, adherence to classic management rules could be advantageously replaced by personalized management. This enables to deeply interpret the characteristics of the patients and a better understanding of disease mechanisms, leading to better treatment choice and prevention. The collected measurements added to the available data sources are recorded in Electronic Health Records (EHRs).

An EHR can be viewed as a real-time patient health record (or a digital version of a patient's health information) with access to evidence-based decision support tools that can be used to aid physicians in their medical care task. An EHR contains the complete medical history of a patient. Such

history can contain diagnoses, drug class plans, medications, comorbidities, allergies, laboratory results, radiology images, etc.

In this study, we aim to improve appointment times and drug class quality to each patient. To achieve this goal, we use a real dataset provided by a medical prescription platform "CEGEDIM" (used by 23,000 doctors in France), that has been collected from 3,000 physicians. This dataset contains 429,087 patients. Each patient is associated with a sequence of visits. A visit contains several patient's characteristics (age, gender, height, diastolic pressure, systolic pressure, pulse rate, etc.) represented as tabular data. At each visit, a drug class is associated to each patient.

The purpose of the work presented in this paper is to propose a Machine Learning (ML) tool for predicting drug class of Hypertension and the next medical appointment for a patient, by using EHRs represented as tabular data that comes from different sources. Among ML approaches, Deep Learning (DL) showed impressive results in various areas like image recognition, translation, etc. DL is actually widely used to improve the quality of health-care.

Recently, recurrent neural networks (RNNs) [2] have encountered a great success in solving several ML problems due to their ability to learn hierarchical features in temporal domains [3]. It is why RNNs have been applied in different tasks such as representation learning, predictive learning, and classification. Nevertheless, simple RNNs are limited in their own since, for example, it is not easy to train them on long time sequences. This is due to the problems of vanishing and exploding gradients that are caused by back-propagation of errors over many time steps. The Long Short Term

Memory models (LSTM) [2] [4] deal with such problem by integrating memory units that allow to learn long temporal dynamics. LSTMs showed their effectiveness in various areas, especially in healthcare [5][3]. Despite their performance, and like most other ML algorithms, LSTMs are limited in their ability to explain results. They suffer from what is called the "black-box" problem.

In almost all approaches, the tabular feature sources that characterize the patients are fused before using the ML models (with or without preprocessing step). The problem is that these features generally come from different sources and their difference must be captured by the ML algorithm. This can be achieved by learning the tabular source representation when building the ML model.

To deal with the problems cited above, we propose a novel deep architecture through a Multi-Sources LSTM Explainable Approach (MS-LSTMEA) applied to Hypertension management. For that purpose, the approach successfully treats separately and differently several sources of tabular medical information, and takes into account long temporal trajectories of events embedded in EHRs. In addition, MS-LSTMEA uses an Attention mechanism [6] which consists in highlighting the information that has been most crucial in the predictive results. Such a mechanism can improve the model's explanation; a very important property in critical areas such as healthcare.

MS-LSTMEA is able to predict the drug class and the next medical appointment for a hypertensive patient. Indeed, the choice of the relevant and personalized drug class for each patient is very important since a bad choice can lead to disastrous effects that can conduct to the death of the

patient. The next medical appointment time is also crucial for physicians when making decisions, because it provides information on the gravity of the patient’s condition. The scientific contributions and challenges of this work are summarized below:

1. Novel multi-source LSTM architecture for dealing with the temporal trajectory of multi-sources events embedded in EHRs (temporal and heterogeneous aspects).
2. Specific processing of data sources of patient characteristics (personal, clinical and comorbidity), that are usually merged in the literature, leading to a high level patient representation learned during the model construction at each time step.
3. Attention mechanism allowing to:
 - (a) Highlight the information that has contributed most to the provided results (decision explainability).
 - (b) Adapt the drug class by identifying important features that are specific for each patient (drug class personalization).
4. Advising support system with the potential to improve patient health-care by predicting the date of his/her next visit (self-management of Hypertension).

In the next section, we will discuss related works in terms of temporal aspect, multi-sources of EHRs, and use of Attention mechanism that improves explainability of DL approaches. Then, Section 3 describes the proposed approach (MS-LSTMEA). Section 4 presents the experiment design, the results and a discussion allowing to highlight the main contributions of this

work. Section 5 concludes and draws some improvements of our work and its perspectives.

1. Related works

Researchers have recently applied DL methods to EHRs in order to predict medical events. We can distinguish the proposed approaches using several criteria [7][8]: the kind of features they consider, the feature representation they used, the type of DL architectures they used, the tasks they performed, their methods attempting to explain the obtained results, etc. We focus on related works that consider the temporal aspects and the multi-sources nature of EHRs and those that use Attention mechanism for model explanation.

1.1. Temporal aspect of EHRs

There are different neural network architectures in the literature that deal with the temporal aspects of EHRs. These architectures are mainly composed of RNNs or Convolutional Neural Networks (CNNs).

RNNs are specially designed to deal with sequences and have been very successful in the NLP field. Since EHRs data are similar to text documents from many perspectives, most of the RNNs that represent EHRs were inspired from Natural Language Processing (NLP) applications.

Pham et al. [9] proposed Deepcare, that is based on LSTMs. It used temporal parameterization that takes into account the dynamics of the diseases. DeepCare deals with the variable size of the input sequences, their irregular timing and the confounding interactions between the intervention and disease progression. It was used to predict unscheduled readmission of patients

with mental disorders and diabetes in an Australian hospital. Lipton et al. [10] used LSTM with target replication to represent the sequential aspect of data to perform diagnoses prediction. They evaluated their approach for predicting different diseases on EHRs data from patients in intensive care units.

Doctor AI [11] was proposed to predict the diseases diagnoses, the treatments, and the date of the next visit. It projected the high dimensional inputs to a lower dimensional space and then used a Gated Recurrent Unit (GRU) network that is based on the same principle as LSTM with less training parameters. Doctor AI used a joint loss function combining the cross entropy for diagnoses and drug class prediction with squared loss for the next visit prediction. In [12], authors addressed the class imbalance problem using LSTMs for the prediction of unplanned readmission risk within 30-day. The proposed approach is based on contextual embedding and medical features of clinical concepts to generate visits representations. The authors succeeded to obtain good results on real data from a Swedish Hospital. Dasilva et al. [13] developed a model based on LSTM to assess the severity of patient's condition in order to anticipate its appropriate treatment. The authors showed good results when predicting vital signs of patients and therefore predict the prognostic indices to treat patients before the deterioration of their health condition.

In [14] and [15], authors proposed the use of bidirectional RNNs (BiGRU or BiLSTM) that can be considered as variants of RNNs. These models consist first in applying the RNNs from the first to the last visit then, from the last to the first visit. The resulting representations are concatenated before

performing predictions.

CNNs have also been adapted to deal with EHRs sequential data. EHRs are represented as matrices, with the horizontal dimensions representing the time and the vertical dimensions representing medical events [16][17][18].

Other works have chosen to combine different DL algorithms for the purpose of improving the prediction results. For example, a CNN was combined with LSTM in [19] and with BiLSTM in [20]. More recently, several works were published about the prediction of COVID-19 comparing the different architectures that deal with temporal data [21][22]. In [22], authors compared different architectures (CNN, BiLSTM, LSTM, and combined CNN-LSTM) for the prediction of COVID-19 and find that LSTM and combined CNN-LSTM give better results.

1.2. Multi-Sources DL and EHR

Medicine practices rely on information provided by different sources: clinical data, laboratory data, imaging data, narrative data, etc. In recent DL applications in the medical field, a fusion of complementary data sources was proposed aiming to improve the performance results [23]. Most recent works combined the tabular EHR data with medical imaging or textual data [23]. In [24][25], authors used CNNs to represent the images, combined the extracted high level features with clinical characteristics and then used a Feed forward neural network. These models did not deal with the temporal aspect of EHRs. In [26] [27], authors aggregated the prediction results using different approaches. Reda et al. [26] built a new classifier that takes the prediction probabilities from the different sources as input for early diagnosis of prostate cancer. Qiu et al. [27] used different ways to aggregate the predic-

tions of the different sources: majority voting, max and min. The results of the three methods were merged and then fused with clinical models. Qiao et al. [28] proposed a multimodal approach based on Attention that combines CNN and Bi-GRU to represent the textual and clinical data for diagnosis prediction. The authors captured the correlations between the two sources before using Joint fusion, where features interactions were performed before using bidirectional RNNs with Attention.

Bagheri et al. [29] combined clinical temporal data with textual data for cardiovascular risk prediction using Joint fusion approach. They used a BiLSTM model that provided word embedding. These embedding were then concatenated with clinical data, resulting in a larger embedding which was then fed to dense layers before performing the prediction.

The multi-sources models proposed in the literature deal only with heterogeneous data (tabular EHR with images or/and text). Even if the data in EHR comes from different sources, these sources are merged and processed in the same way.

Our approach learns the tabular source representation during the model construction by simultaneously taking into account their temporal aspect. To our knowledge, it is the first approach that deals with this problem.

As a critical application area, healthcare requires explanation of the results provided by DL tools to be thoroughly explained. The interpretation of the models and/or the predictions is crucial since it can highlight the features that are involved in the predictions and therefore allows models to be validated or trusted and/or the discovery of unknown relationships between the features and the predictions. Therefore, different approaches that aim to ex-

plain DL models used in healthcare were proposed in the literature [30][31]. Recently, Attention mechanism has received a lot of interest in several application areas like Natural Language Processing (NLP), where it has been used to improve model results and to explain prediction. Most of the models applied to EHRs are inspired by those used in the field of NLP. We focus on the related works that use Attention mechanism for model or predictions interpretation.

1.3. Attention based interpretation

Attention mechanism in DL allows to highlight elements/features that are involved in the prediction. Self-attention provides a representation of the same sequence by linking its different positions.

Attention mechanisms are mainly combined with different RNNs models at different levels in the architectures. Shickel et al. [32] proposed DeepSOFA for the disease severity prediction. It used a GRU to handle the temporal aspect of EHRs and Self-attention to point up the most relevant time points when predicting hospital mortality. After embedding EHRs, the model assigns weights to time points that precede the current time point. The Attention scores allow to identify and highlight the time periods that influences the predictions. In [33], Ma et al. proposed a model that is based on bidirectional RNN, called Dipole. It is used to predict the future health information of a patient. To improve and interpret the prediction results, authors proposed three variants of their algorithm by introducing three Attention mechanisms that allow to compute different context vectors. These mechanisms differ by the way they capture the relationships between the hidden states of the RNN.

Edward Choi and al. [34] developed a DL model, called RETAIN (REverse Time AttentIoN) that predicts heart failure in the outpatient setting. It is able to compute the importance of a variable or a visit through two RNNs and a two-level Attention mechanism which can identify important visits and the relevant clinical features in the highlighted visits. The tool GRNN-HA proposed in [14], based on a Bi-GRU and hierarchical Attention, performs mortality prediction. It is able to compute two levels of Attention: one level for medical codes and another one for the visits of a patient. Another model that introduced hierarchical Attention mechanisms, called Patient2Vec, was proposed by Zhang et al. [15]. Patient2Vec used a BiGRU with a two-level hierarchical Self-attention in order to predict the future risk of hospitalization. Authors proposed to group visits into times periods called sub-sequences and proposed hierarchical self-attention mechanisms at two levels: medical codes (variables) and sub-sequences. Kaji et al. [35] applied an Attention mechanism at the level of input variables, then used LSTM in order to predict several medical outcomes (daily sepsis, vancomycin antibiotic, myocardial infarction, and administration) over two-week-long patient Intensive Care Unit (ICU) courses using the MIMIC-III dataset. They provided, from a simple Attention mechanism, an interpretation at the level of variables and at the level of visits. More recently, Men et al. [36] proposed to predict multiple diseases based on EHR data, from a hospital in Southeast China, using LSTM extended with time-aware (for handling temporal irregularities in EHRs) and Attention mechanism (for prediction interpretation). The latter allows to identify the relevance of each visit for the disease prediction. The authors showed that their approach outperforms classical ML and DL

methods for the prediction of the diagnoses of future disease.

In our work, we propose a novel approach, which is based on a joint fusion of data sources, that takes into account the temporal aspect of EHRs using LSTMs and the different sources of EHRs tabular data even though they can be merged directly at the level of inputs (early fusion) as done in related works. We showed that this representation improves the results. Moreover, unlike existing approaches, the construction of the data sources embedding is not separated from the overall architecture, it is learned during the training step of the whole NN. Our approach deals simultaneously with multiple sources over time in the RNN since at each visit, the learned high-level representation of the sources is transmitted to the LSTM units. We also use a Self-attention mechanism that helps to improve the results and to highlight the most important variables, sources, and patient visits. Unlike existing approaches that embed the Attention mechanism inside RNN units, or that use complex multi-level Attention, our approach implements a simple attention mechanism at the input level. This mechanism allows a direct interpretation of the model predictions.

2. Method

As already said, we present a novel explainable Multi-Source Approach that is based on LSTM architecture, called MS-LSTMEA. This approach is dedicated to predicting clinical events for each patient, such as drug class and the date of the next medical appointment.

We chose LSTM architecture due to its ability to remember previous entries, which makes it well suited for time series prediction.

MS-LSTMEA successfully integrates multiple tabular data sources containing medical patients information and takes into account the temporal aspect of EHRs. The architecture of MS-LSTMEA is showed in Figure 1. Every data source is passed to fully connected layers that build individual embedding for the source. These embeddings constitute abstract high level representations of the data sources that are concatenated before being processed by an LSTM unit. The goal is to predict the drug class at each visit and the date of the next medical appointment for each patient.

In addition, we introduced a Self-attention mechanism at the level of the

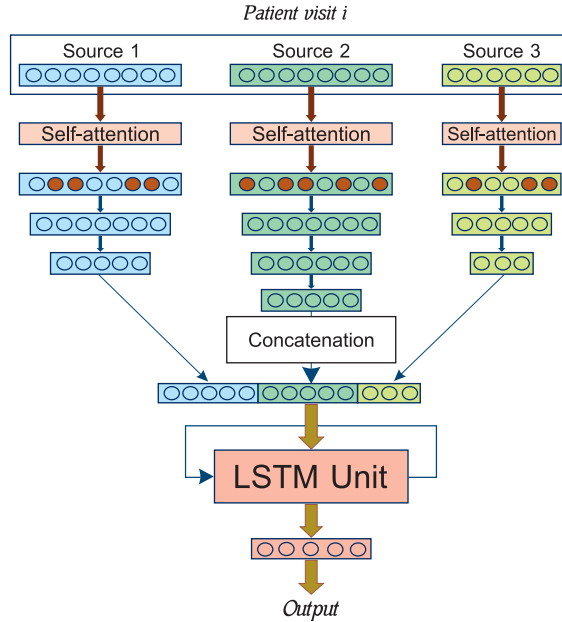


Figure 1: MS-LSTMEA architecture

input. This improves both the results and the explanation of the predictions and the model. Indeed, we are able to determine the most relevant features that are used for each prediction thanks to the attention mechanism that is

located at the input level. The importance of the sources and the visits can then be deduced from the scores of the important features.

We represent each patient $p \in P$ by a succession of n_p medical events (visits) v_t ($1 \leq t \leq n_p$): $p = [v_1, \dots, v_{n_p}]$.

The goal is to predict the date of the next medical appointment (see Section 3.3) or the drug class among a set of drug classes: $Y = \{y_1, \dots, y_m\}$, where m is the number of drug classes. The different knowledge sources are usually concatenated (early fusion) and not separated in the literature since they share a tabular representation. In our approach, we used joint fusion where the high-level representations (embeddings) of sources are combined.

Guided by the EHR data that we used, we defined the three following sources¹:

- $s(1)$: refers to the personal features (gender, age, etc.)
- $s(2)$: corresponds to clinical features (clinical codes related to the diagnoses, the procedures of medications, etc.);
- $s(3)$: contains comorbidity features that correspond to the presence of other diseases in addition of hypertension (disease of interest).

Each source of information is transmitted to fully connected layers after learning attention scores.

¹Our architecture is enough general since we can add easily other data sources that contain additional information about patients.

2.1. Attention mechanism

A visit v_t at time t of a patient p is represented by a vector that contains the values of the sources of information (see Figure 1):

$v_t = [s_t(1), s_t(2), s_t(3)]$, where $s_t(1)$, $s_t(2)$ and $s_t(3)$ are the three sources that describe the patients p at the visit v_t . The sources $s(k)_t$ ($k \in \{1, 2, 3\}$) are represented by numerical vectors of dimension d_k . We introduce a set of weights \mathcal{W}^k ($k \in \{1, 2, 3\}$) that allows to calculate the feature-level self-attention scores $\alpha_t(k)$ associated to the source $s_t(k)$ ($k \in \{1, 2, 3\}$). These weights are learned during the training step and the attention scores $\alpha(k)_t$ ($k \in \{1, 2, 3\}$) are computed as follows:

$$\alpha_t(k) = \text{softmax}(\mathcal{W}^k \cdot s_t(k)) \quad (1)$$

The Attention scores $\alpha_t(k)$ associated to the input patient visit v_t and then transmitted to the fully connected layers in order to build the high level of representation as follows:

$$\hat{s}_t(k) = \alpha_t(k) \odot s_t(k), \quad k \in \{1, 2, 3\} \quad (2)$$

where \odot denotes element-wise product.

Note that, unlike most of existing approaches that introduce Attention scores inside the LSTM unit (mostly with the aim to improve the prediction results), we add an Attention layer at the level of input sources to preserve interpretability. The computed score is based on learned weights \mathcal{W}^k ($k \in \{1, 2, 3\}$), that provide variables importance like in [35].

After this step, we obtain, for each patient p , high level representations $\tilde{s}(k)$,

of the sources $s(k)$ ($k \in \{1, 2, 3\}$) using, a specific number l_k of connected hidden layers.

at time t , these embeddings are concatenated in order to form one multidimensional vector \tilde{v}_t defined as follows²:

$$\tilde{v}_t = [\tilde{s}_t(1); \tilde{s}_t(2); \tilde{s}_t(3)] \quad (3)$$

For the rest of the paper, the components of the high level representation of the visit at time t , \tilde{v}_t , are denoted as \tilde{v}_{tj} , i.e.,

$$\tilde{v}_t = [\tilde{v}_{t1}, \dots, \tilde{v}_{td}], \quad \text{where } d \text{ is the dimension of } \tilde{v}_t \quad (4)$$

2.2. Visits representation

The embeddings built for the visit of each patient are fed sequentially to the LSTM units that take into account the temporal aspect of these visits. We use a one-to-one LSTM in order to provide a prediction at each time step. Figure 2 illustrates the temporal aspect of our model.

A visit $v_t \in V$, at time t , of patient $p \in P$ is embedded into a representation space where it is transformed to a vector \tilde{v}_t . The resulting vector is fed to an LSTM unit that produces the prediction of the drug class or the next medical appointment. An LSTM unit (see Figure 3) is composed of a memory cell c , that keeps the observed information from the input to the current time step, and three gates that control the state of the cell: forget gate f that controls the retention or deletion of old information; input gate

²Note that the temporal aspect of the visits is not taken into account in the fully connected layers. It is handled by the LSTM.

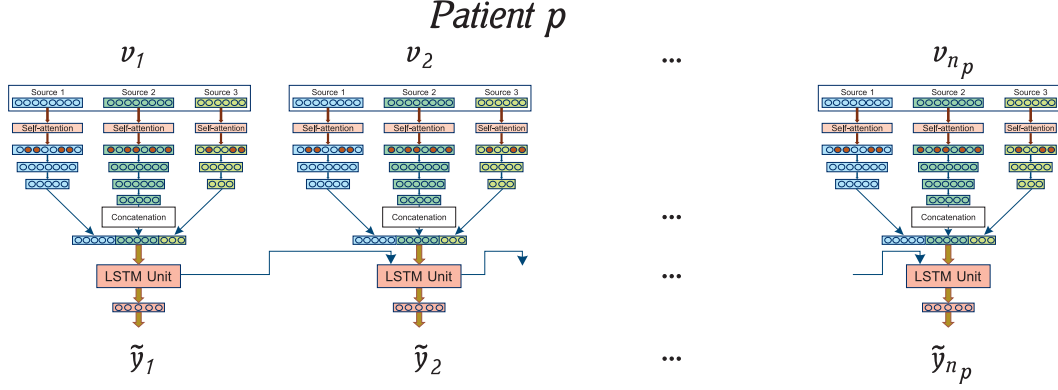


Figure 2: Unfolded MS-LSTMEA Architecture

i , which selects the information that will be added to the cell state c ; output gate o that selects the information that will be returned by the cell state.

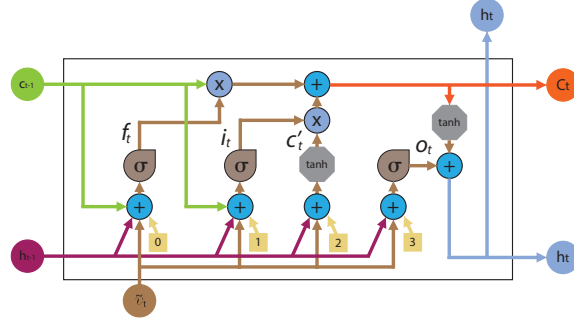


Figure 3: LSTM Unit

The LSTM unit receives three inputs: c_{t-1} , h_{t-1} and \tilde{v}_t . c_{t-1} represents the memory cell and h_{t-1} corresponds to the state of the LSTM unit at time step $(t - 1)$. The value of the initial LSTM state h_0 is equal to 0.

At time step t , three gates f_t , i_t , and o_t are defined as follows:

$$\begin{aligned}
f_t &= \text{sigmoid}(\mathcal{W}_f[\tilde{v}_t; h_{t-1}] + b_f) \\
i_t &= \text{sigmoid}(\mathcal{W}_i[\tilde{v}_t; h_{t-1}] + b_i) \\
o_t &= \text{sigmoid}(\mathcal{W}_o[\tilde{v}_t; h_{t-1}] + b_o)
\end{aligned} \tag{5}$$

Where (\mathcal{W}_f, b_f) , (\mathcal{W}_i, b_i) , and (\mathcal{W}_o, b_o) are the (weights, biases) couples associated to the three gates.

The unit state h_t at the current time step t depends on the input \tilde{v}_t , the states of the three gates and the previous LSTM state h_{t-1} . It is calculated by:

$$\begin{aligned}
\tilde{c}_t &= \tanh(\mathcal{W}_c[\tilde{v}_t; h_{t-1}] + b_c) \\
c_t &= f_t \cdot c_{t-1} + i_t \cdot \tilde{c}_t \\
h_t &= o_t \cdot \tanh(c_t)
\end{aligned} \tag{6}$$

At each time step t , the state h_t of the LSTM unit is updated before being propagated to the following output layer in order to predict the drug class or the date of the next medical appointment. For the classification task, we used a softmax activation function in the output layer in order to compute the probabilities π_t associated to the predictions as following:

$$\pi_t = \text{softmax}(\mathcal{W}.h_t + b) \tag{7}$$

Where \mathcal{W} represents the learned weights and b is the bias.

The prediction \tilde{y}_t at time t of our model is defined by: $\tilde{y}_t = \underset{i=1}{\operatorname{argmax}}^{\varphi} \pi_{ti}$.

A categorical cross entropy function E is used to train the model³, it is defined as follows:

³In the case of regression tasks, we can use a linear activation function in the output

$$E = -\frac{1}{\#P} \sum_{p \in P} \frac{1}{n_p} \sum_{t=1}^{n_p} \sum_{i=1}^{\varphi} [y_{ti} \log(\pi_{ti})] \quad (8)$$

Where $\#P$ represents the patients number in the training set, π_t is the model output probabilities and y_t is the true target (class)⁴.

3. Experimentation

In order to validate our approach, we performed different experiments whose results are presented in this section. For these experiments, we have considered Hypertension, which is known as one of the major factors of heart disease and stroke. Gathering regular medical measurements helps to consolidate the diagnosis of Hypertension. These measurements are exploited and combined into EHRs with different sources in order to improve the prediction of the right drug class for a patient and the date of patient’s next medical appointment. This allows to help patients to self-manage their disease.

3.1. The Dataset

The dataset considered in the experiments comes from a real data collected by 3,000 doctors, provided by CEGEDIM (medical prescription platform used by 23,000 doctors in France). This dataset contains the characteristics of 429,087 patients which are under follow-up for Hypertension. Each patient is represented by a sequence of visits and received an average of twenty visits. We have retained, for each visit, 15 characteristics (Table 1 describes the most important ones). A class of drug is associated to each

layer and a mean squared error loss function.

⁴Note that we represent the classes using One-Hot Encoding.

patient at each visit. Six classes of drugs were distinguished: Alpha blockers (AB), Beta-blockers (BB), Diuretics (D), Angiotensin II receptor blockers (ARAI), ACE inhibitors (IEC), and Calcium channel blockers (IC) (see Table 2).

The dataset was anonymized by removing sensitive data such as names,

Features		Signification	Source type
Gender		Indicates the person's gender	
Age		Patient's age at time of prescription	s(1)
Weight		Patient's last recorded weight	
Height		Patient's last recorded height	
Pulse		Heart rate associated with medical consultation	
Diastolic	Blood	Diastolic blood pressure associated with the medical consultation	
Pressure			
Systolic	Blood	Systolic blood pressure associated with the medical consultation	s(2)
Pressure			
Measure	blood	Patient's last measurement of blood sugar	
		glucose	
Insulin treatment		Insulin drug class (if any)	
Other treatment		If there is drug class (other than insulin)	s(3)

Table 1: Most important features of a visit.

addresses and phone numbers.

Number of patients	429087
Number of input features	15
Average number of visits	20
Type of data	Real/Boolean
Gender repartition	Female: 41%, Male: 59%
Average age at the time of prescription	68.8
Average duration between visits (days)	63.8
Treatment classes	BB, AB, IEC, ARAII, IC, D

Table 2: Datasets description.

3.2. Experiment setup

An LSTM model has been trained for predicting the drug class and the date of their next medical appointment for a patient. The proposed model receives a multidimensional vector v_t representing the features related to the visit at time t for a patient p . As explained previously, a visit v_t is composed of Personal features ($s(1)$), Clinical features ($s(2)$), and Comorbidities features ($s(3)$), as shown in Table 1.

First, the data is split up into training, validation, and test sets. 60% of patients data were used for training the models, 20% as the validation set and 20% as the test set. An Attention mechanism is applied to each data source. The output of the Attention mechanism is used, for embeddings' construction, as input to a two fully connected layers with sigmoid activation function. Then, each visit of each patient is represented by concatenating the high level representations sources. It is this representation which is used to proceed to the predictions by using LSTM units. The model has been

trained for 1000 epochs ⁵. Dropout has been used between LSTM layers to avoid overfitting. All model hyperparameters are fixed using grid search. Note that we tested different variations of LSTM: LSTM with stacked layers, Bi-LSTM and we did not see any improvement in results compared to vanilla LSTM.

3.3. *Experimental results*

Our algorithm has been validated by comparing the results given by MS-LSTMEA with those of other proposed methods that are based on different DL architectures. We considered: Doctor AI [11] (based on (GRU), Med2vect [37] (based on Auto-encoders), Deepr [17] (based on CNN), DeepCare [9] (based on LSTM), DeepPatient [38] (based on Multilayer Perceptrons). The following aspects have been evaluated:

- Prediction results.
- Influence of source type on the prediction results.
- Influence of the high level representation of sources on the prediction results.
- Determining relevant clinical factors.
- Determining individual relevant clinical factors.

Prediction results

The results of the comparison of our MS-LSTMEA based Attention algorithm with different algorithms, regarding the prediction of the drug class

⁵One epoch corresponds to one iteration over the entire training data.

Approach	Accuracy	Precision	Recall	F1-Measure
MS-LSTMEA	80.4 ± 0.03	78.4 ± 0.02	83.9 ± 0.2	81.0 ± 0.04
Doctor AI	68.3 ± 0.07	69.5 ± 0.01	65.2 ± 0.02	67.3 ± 0.03
M2vect	65.4 ± 0.02	59.5 ± 0.07	69.2 ± 0.11	64.0 ± 0.05
DeepCare	55.5 ± 0.11	54.6 ± 0.03	64.9 ± 0.01	59.3 ± 0.09
DeepPr	51.9 ± 0.04	52.2 ± 0.02	46.6 ± 0.03	49.2 ± 0.02
DeepPatients	58.8 ± 0.01	57.2 ± 0.03	69.3 ± 0.01	62.7 ± 0.13

Table 3: Comparative results while predicting the best drug class.

and the date of the next medical appointment, are reported in Tables 3 and 4. Table 3 shows that our approach greatly outperformed the existing ones while predicting drug class. Indeed, MS-LSTMEA obtained an accuracy by up to 80.4% while Doctor AI, which resulted in the second best score, obtained barely 68.3%. This can be attributed to the use of LSTM instead of GRU in DoctorAI since LSTMs remember longer sequences than GRUs. While this is the main reason why LSTMs outperform other algorithms in tasks requiring modeling long sequences, we will see hereafter that the introduction of Attention mechanism further improves the results.

The prediction results are in the favor of our approach, compared to other approaches. This confirms the efficiency of the proposed visit representation, the use of an Attention mechanism and fully connected layers.

The prediction of the next medical appointment should be modeled as a regression problem. The time between visits (time period) could be biased because it depends not only on the availability of doctors and patients, but also on patient’s health condition. In our approach, we decided to discretize the time period. Thus, the regression problem was transformed into a clas-

sification problem.

After having explored several discretization values, and having had several discussions with physicians, we opted for a time period covering one month. In the used EHRs, the smallest time between two visits is equal to 10 days and the largest one is up to 200. So, seven classes C_k have been fixed, depending on the time period as follows:

Let k be the number of months since the previous visit, so seven classes are fixed, depending on the time period as follows:

- For $k = 1$ to 6 : Class $C_k = k$, if $(k - 1)$ months $\leq \text{time period} < k$ months
- Class $C_7 = 7$ if $\text{time period} \geq 6$ months

In practice, physicians⁶ plan the next visit of their patient according to the patient's situation. When a patient's Systolic Blood Pressure is very high, physicians usually set their next visit very close (less than 25 days) for therapeutic adjustments. The visit is planned within three months if the Hypertension is severe, i.e., with associated risk factors (e.g., a patient who had a previous stroke) or with comorbidities (e.g., diabetes). Renewal of the drug class requires close monitoring. In the case of a renewal of a drug class for a person with few health problems and for young patients whose blood pressure has been stabilized for a long while, the delay goes beyond six months. The use of classes incremented by month applies well to patients in self-measurement at home as it creates an opportunity to detect any prob-

⁶Source: interview with Dr François Teboul, Emergency Doctor, Paris, May 2021

Approach	Accuracy	Precision	Recall	F1-Measure
MS-LSTMEA	79.5 ± 0.04	78.6 ± 0.01	80.9 ± 0.01	79.7 ± 0.03
Doctor AI	68.2 ± 0.02	66.7 ± 0.11	72.2 ± 0.04	69.4 ± 0.03
M2vect	57.6 ± 0.01	56.2 ± 0.01	66.8 ± 0.1	61.1 ± 0.03
DeepCare	74.7 ± 0.02	73.4 ± 0.01	77.1 ± 0.05	75.2 ± 0.01
DeepPr	59.4 ± 0.03	58.2 ± 0.05	65.6 ± 0.07	61.7 ± 0.01
DeepPatients	62.2 ± 0.02	60.3 ± 0.03	70.9 ± 0.02	65.2 ± 0.01

Table 4: Comparative results while predicting the time before the next medical appointment.

lematic situations beyond the above framework.

Table 4 also illustrates the superiority of our MS-LSTMEA based approach, which gives an accuracy of up to 79.5%.

The influence of source type on the prediction results

For studying the influence of combining multiple sources of information on the performance, we have tested the different combinations⁷. Indeed, determining the significant sources of features plays an important role in predicting Hypertension. Choosing the relevant combination of features plays also a crucial role in improving the performance of the prediction model. Table 5 shows that Clinical features outperform the Personal and the Comorbidity features when predicting the drug class or the date of the next medical appointment. The prediction accuracy is improved by up to 44.7%

⁷The fact of having three distinct input sources, necessitated to set up seven experiments.

(s1)	(s2)	(s3)	Accuracy (T)	Accuracy (V)
0	0	1	22.2 \pm 0.03	13.7 \pm 0.04
0	1	0	49.3 \pm 0.01	51.8 \pm 0.03
0	1	1	64.3 \pm 0.02	59.1 \pm 0.05
1	0	0	24.5 \pm 0.04	19.4 \pm 0.03
1	0	1	36.3 \pm 0.06	46.2 \pm 0.02
1	1	0	67.7 \pm 0.07	57.1 \pm 0.01
1	1	1	80.4 \pm0.03	72.3 \pm0.14

Table 5: Influence of source type (1: included, 0: not included) on the prediction Accuracy.

when they are included to predict drug class. However, using only Personal features gives a low accuracy (score of 24.5%). This can be interpreted by the fact that the age, gender, weight etc. without Blood pressure measurements do not really make sense for drug class prediction. What is important to note is that the results show that we can produce more accurate predictions (accuracy=80.4% for the prediction of the drug class) when merging the different information sources, which provide measurements that can be different and complementary in their nature.

Including solely Clinical features for predicting the time before the next medical appointment gives 51.8% of accuracy that is higher than the accuracy obtained with personal data only (19.4%) or with only Comorbidity features (13.7%).

Influence of high level representation of sources on the prediction results

We evaluated the influence of high level representation of sources (construction of embeddings). For that purpose, we tested our model with and

Embeddings	Accuracy (T)	Accuracy (V)
Yes	69.3 \pm 0.02	65.7 \pm 0.03
No	80.4 \pm 0.03	72.3 \pm 0.14

Table 6: Influence of embeddings construction on the prediction results.

without construction of embeddings. In Table 6, it clearly appears that construction of embeddings from EHRs provides new data representations and improves prediction results whether when predicting the drug class or the date of a patient’s next medical appointment. This is due to the ability of our approach to build relevant high level representation from the initial data sources using the fully connected layers.

Determining relevant clinical factors

The Attention scores over visits were used to construct heatmaps for the models trained to predict the drug class or the date of a patient’s next medical appointment. The softmax scores for each feature were averaged over test patients who were predicted to belong to the same class in order to obtain patients-averaged Attention maps. This is helpful for highlighting the individual characteristics that had the most influence on the prediction of a class of interest.

Figure 4(A) shows a heatmap while averaging the Attention scores of patients when predicting the drug class (BB) at each visit. This shows, for each visit, the most relevant features that lead to the prediction. Figure 4(B) shows the averaged Attention scores obtained in Figure 4(A) on the features for each patients visit. This shows the visits which contributed the most to the prediction. We can see that Visits 4 and 10 are the most efficient visits

for the prediction. Figure 4(C) shows the averaged Attention scores on the visits to highlight the most significant features that are used for the prediction. As we can see, the most significant features are 'Pulse' and 'Diastolic pressure' with an Attention score equals respectively to 0.63, and 0.62. The feature 'Systolic pressure' has an Attention score up to 0.58. In summary, we can confirm, once more, that the most significant features are the clinical ones.

Finally, the Attention mechanism has been also applied to each data source to observe the importance of sources. By averaging the scores of the features that constitute each data source, we measured their importance scores. Figure 4(D) shows that the most important source type for prediction of the drug class is the Clinical features with an Attention score of up to 0.53.

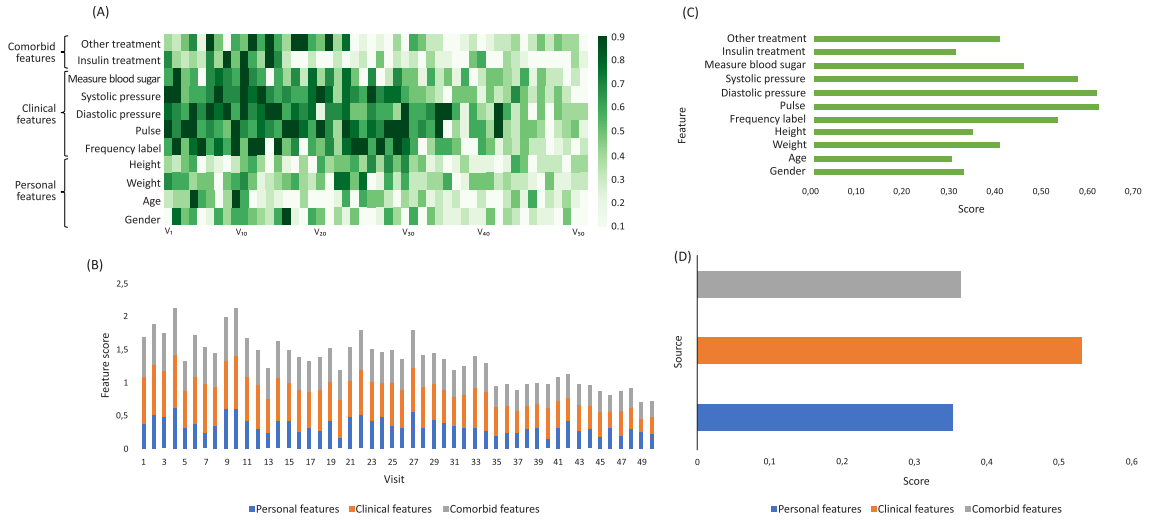


Figure 4: Attention heatmaps averaged over all test patients with visit Attention, features, and sources with the highest mean activation per visit while predicting the drug class BB.

The second heatmap of Figure 5 shows the patients-averaged Attention scores when predicting the date of the next medical appointment (class C_1). The results shows that we can extract the features with the highest Attention scores for each visit from Figure 5(A): for visit 1, 'Pulse' and 'Systolic Blood Pressure' have the highest score. The most relevant visits can be extracted from Figure 5(B): as we can see, the visits 9, 10, and 11 contribute the most to the prediction. The most important features are given by the Figure 5(C): as shown, the features 'Diastolic pressure' and 'Pulse' are the most relevant in predicting the next medical appointment with respectively 0,616, 0,61 and 0,60 as Attention scores. The importance of data sources is given by the Figure 5(D): we can observe that, similarly, the most important source for this prediction is the Clinical Features source with an Attention score of up to 0.54.

Determining relevant individual clinical factors

We have generated individualized predictions and Attention visualization heatmaps for each patient for the two predictions task (the drug class and the date of the next medical appointment).

Figures 6 and 7 show the same information as in the previous section for each patient. Heatmaps are built for each patient individually in (A). Then, we plotted the Attention scores related to visits (B), features (C), and data sources (D). Figure 6(B) shows that visit 5 is the most relevant one for predicting the drug class for this patient. Figure 6(C) shows that 'Diastolic pressure', 'Pulse', and 'Prescription blood sugar' are the most important features with respectively 0.78, 0.74, and 0.74 as Attention scores. Figure 6(D) confirms, once more, that the most important features belong to the clinical

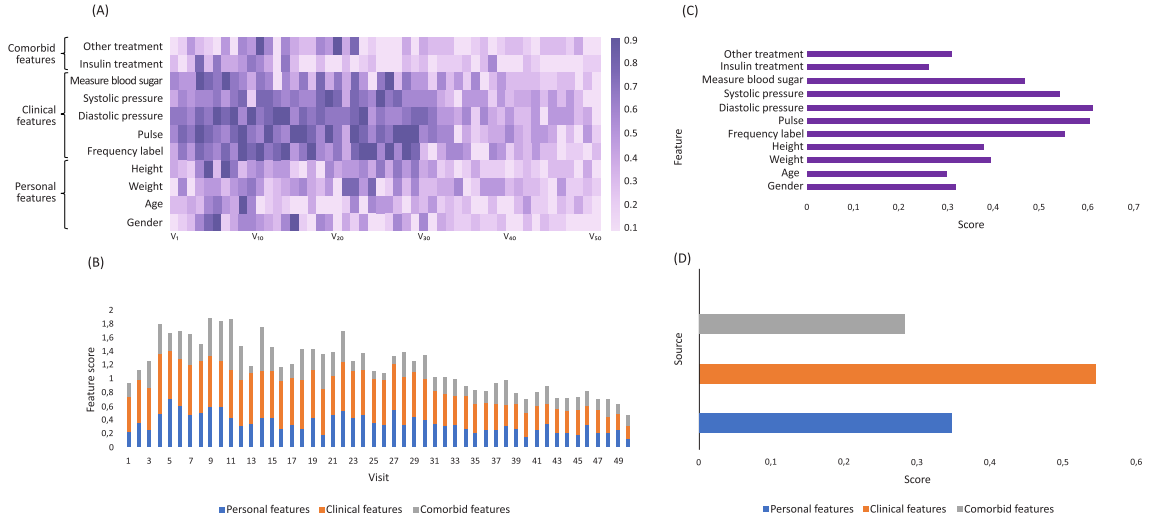


Figure 5: Attention heatmaps averaged over all test patients with visit Attention, features, and sources with the highest mean activation per visit while predicting the date of the next medical appointment (class C_1).

features with an Attention score up to 0.7.

Similarly, the heatmap of the Figure 7(B) shows that for the same patient, the visit 4 is the most determinant one, the features 'Duration' and 'Diastolic pressure' are the most determinant features regarding the prediction of the next medical appointment with respectively 0.71 and 0.60 as Attention scores (Figure 7(C)). This figure illustrates the sources scores and definitely confirms that clinical features represent the most determinant ones with an Attention score of up to 0.47.

3.4. Discussion

This study presents a deep learning model based on LSTM and an Attention mechanism for performing two prediction tasks that are important in

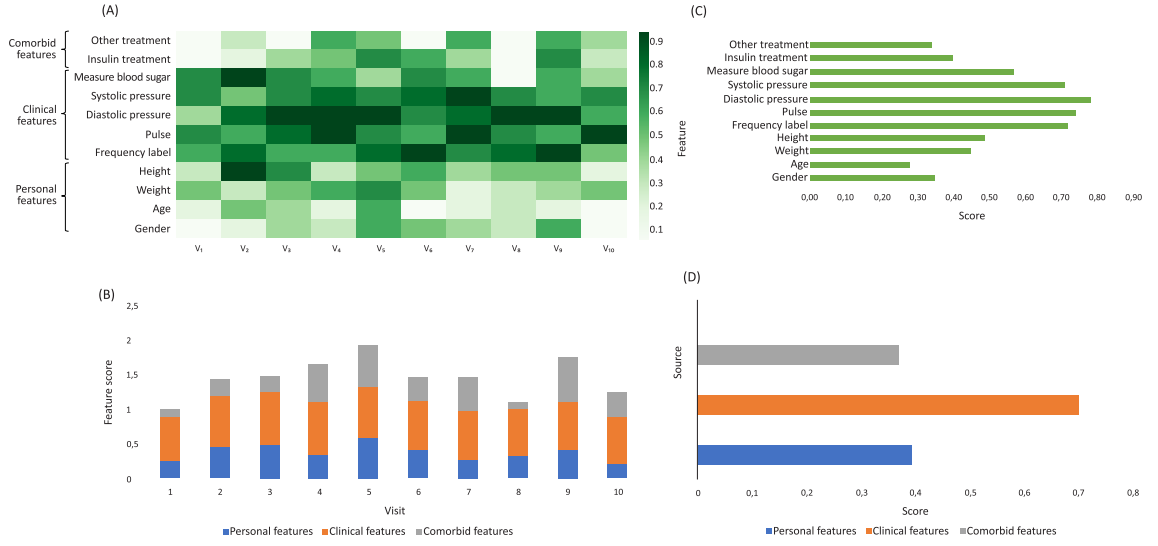


Figure 6: Personalized Attention heatmap while predicting the drug class BB for one patient.

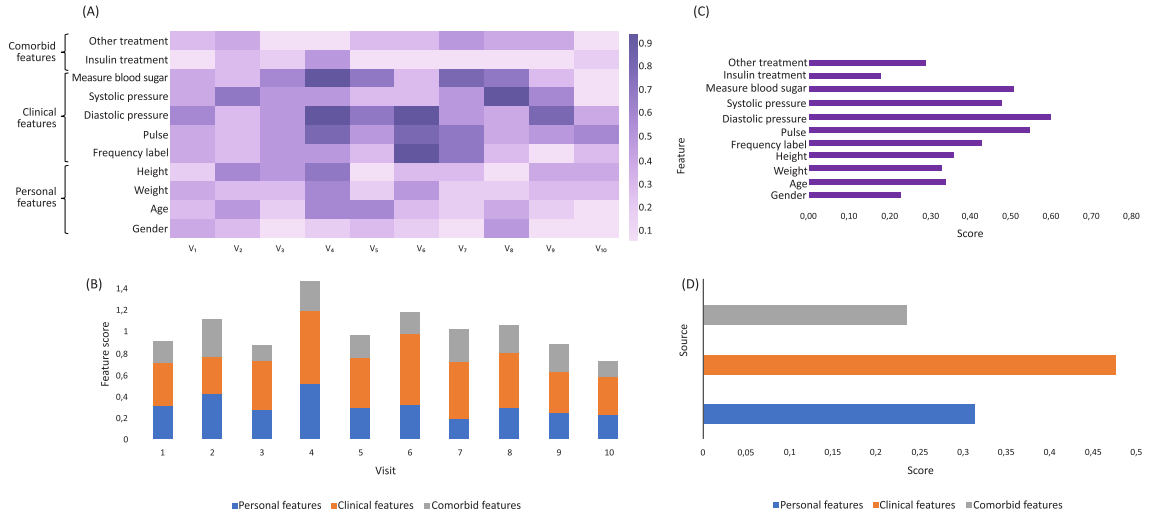


Figure 7: Personalized Attention map for the prediction of the next medical appointment for one patient (class C_1).

Hypertension management: the drug class for patients and the date of their next medical appointment. The model gives high accuracy of up to 80.4% while predicting the drug class and an accuracy of up to 72.3% while predicting the date of the next medical appointment. The results showed that high level representation of Personal, Clinical, and Comorbidity features realized separately and differently, from EHRs using a fully connected layers, with a simple Attention mechanism, is a strategy which outperforms models that ignore any of these aspects. Moreover, we showed that combining different features results gives better performance than using either of them alone. In addition, our model proved itself to be able to capture the significant features, sources or visits that contribute the most to the predictions thanks to the use of the Attention mechanism. Thus, our model can easily provide an explainable prediction.

There are a many works regarding the prediction of high Blood pressure and cardiovascular risk, but few address the prediction of the drug class for Hypertension. Close to this is the work of Xiangyang and his colleagues [39] on the prediction of the optimal therapeutic pathway for Hypertension. They used RNNs including LSTM to predict the probability of reaching a target BP (Blood Pressure) control with different drug class pathways. As in our case, LSTMs reveal a high degree of accuracy in prediction. We think that the strategies of predicting the drug class and the optimal pathway can be considered as two complementary ways that can help patients and doctors collaborate and improve Hypertension management. Another related work concerns the personalization of hypertensive drugs [40]. The authors considered five antihypertensive drug prescriptions for a 14,581 inpatient population

where each patient is described by 15 features and belongs to the successful group if its DBP (Diastolic BP) < 90 and SDB (Systolic BP) < 140 , to the unsuccessful group otherwise. To reveal the difference in characteristics of the features between the two groups, the authors applied statistical and machine learning methods to assign a weight to each feature. Each weight represents the importance of a feature for a patient regarding the effectiveness of a drug. The results of this study can help doctors in choosing the most adequate drugs for patients. Unfortunately, unlike in our study, the considered dataset does not take into account the temporal aspects related to Hypertension drug class. Moreover, the study could have gained in reliability if the considered dataset had been larger.

A limitation in the construction of the predictive model of current research is related to the use of only demographic features [41] and clinical features [42]. The former discards a huge proportion of information in each patient records, whereas the latter ignores the knowledge and guidelines coming from human intelligence. Our experiments showed that using information regarding comorbidities improves the performance of deep learning algorithms. Such expert knowledge is considered as a domain-specific knowledge which is also a big challenge since it impacts the progression of the disease and the state of the patient.

When dealing with deep learning models for predicting clinical events with EHRs, it is important to ensure that the model is flexible. Such flexibility ensures the possibility of adding new data sources. In fact, the combination of multiple data inputs (biological, pathological, medical images, etc.) raises expectations and hopes in terms of understanding the causes and mechanisms

of diseases as well as for the personalization of medical monitoring. Some of existing works did not take into account the integration of new data sources such as the work in [17]. Our MS-LSTMEA approach is integrative since it can easily handle new data sources, either being discrete or categorical, regardless of heterogeneity. This is because we build for each data source differently and separately an embedding to represent a visit.

4. Conclusion

In this work, we have proposed MS-LSTMEA, which is an LSTM-based model that can learn efficient patient representation from a large amount of electronic patient records and predict future events for patients, such as the drug class or the date of their next medical appointment. We have tested MS-LSTMEA on a large dataset of real EHRs, which achieved 82.90% accuracy and significantly outperformed many existing approaches. We have also shown that high level representation of Personal, Clinical, and Comorbidities features from EHRs, with an LSTM network highly improves the performance.

In the conducted experimentation, we have used a real set of data collected from 3,000 doctors, provided by CEGEDIM (medical prescription platform used by 23,000 doctors in France). The provided dataset describes the characteristics of 429,087 patients, each of whom is represented by a series of visits. Our method is generic enough to integrate several kinds of knowledge sources (Ex. genomics, images...). The approach and the results can be used to design a Decision Support System (DSS) that could be used by hospitals and physicians for Telemonitoring (follow-up) of Hypertensive pa-

tients in terms of predicting the drug class and the date of their next medical appointment. At the last, but not the least, the proposed approach can help patients self-manage their Hypertension while providing explanations for the decisions of the model, contributing to the challenges of explainable Artificial Intelligence (AI), a subject of high interest in the current AI landscape [43][44].

This work raises different research perspectives. Indeed, we used the LSTM to model the temporal trajectory of patient visits. However, other models like the transformer [45] have recently proved their efficiency in the processing of sequential information and especially the EHRs [46]. Future work will focus on these models. In order to improve the explainability and transparency of our approach, we plan to combine the neural networks with symbolic AI approaches. We will focus on applying a multi-agent argumentation protocol that will be able to deal with prior knowledge injection from the official guidelines behind Hypertension. This has already been successfully done with an ensemble method based on multilayer perceptrons in [47].

CRediT authorship contribution statement

Farida Zehraoui (FZ) is the main contributor and the supervisor of the work. She designed and formalized the method. She supervised the experiments. She also wrote the state of the art and highly contributed to the writing of the paper. **Naziha Sendi** realized the implementation and performed the experiments. She also wrote the first draft of results analysis. **Nadia Abchiche-Mimouni (NAM)** participated to the conception of the method and to the writing and proof-reading of the paper. The Funding

acquisition and the Project administration have been ensured by **NAM** and **FZ**. All authors approved the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

- [1] G. A. Roth, al., Global burden of cardiovascular diseases and risk factors, 1990–2019: Update from the gbd 2019 study, *Journal of the American College of Cardiology* 76 (2020) 2982–3021.
- [2] A. Sherstinsky, Fundamentals of recurrent neural network RNN and long short-term memory LSTM network, *Physica D: Nonlinear Phenomena* 404 (2020) 132306.
- [3] V. K. R. Chimmula, L. Zhang, Time series forecasting of covid-19 transmission in canada using lstm networks, *Chaos, Solitons Fractals* 135 (2020) 109–864.
- [4] M. C. . N. G. Van Houdt, G., A review on the long short-term memory model., *Artificial Intelligence Review* 53 (2020) 5929–5955.
- [5] M. Iqbal, F. Al-Obeidat, F. Maqbool, S. Razzaq, S. Anwar, A. Tubaishat, M. S. Khan, B. Shah, Covid-19 patient count prediction using lstm, *IEEE Transactions on Computational Social Systems* 8 (2021) 974–981.

- [6] Z. Niu, G. Zhong, H. Yu, A review on the attention mechanism of deep learning, *Neurocomputing* 452 (2021) 48–62.
- [7] B. Shickel, P. J. Tighe, A. Bihorac, P. Rashidi, Deep ehr: a survey of recent advances in deep learning techniques for electronic health record (ehr) analysis, *IEEE journal of biomedical and health informatics* 22 (2017) 1589–1604.
- [8] J. R. Ayala Solares, F. E. Diletta Raimondi, Y. Zhu, F. Rahimian, D. Canoy, J. Tran, A. C. Pinho Gomes, A. H. Payberah, M. Zottoli, M. Nazarzadeh, N. Conrad, K. Rahimi, G. Salimi-Khorshidi, Deep learning for electronic health records: A comparative review of multiple deep neural architectures, *Journal of Biomedical Informatics* 101 (2020) 103337.
- [9] T. Pham, T. Tran, D. Phung, e. J. Venkatesh, Svetha", L. Khan, T. Washio, G. Dobbie, J. Z. Huang, R. Wang, DeepCare: A deep dynamic memory model for predictive medicine, in: *Advances in Knowledge Discovery and Data Mining*, Springer International Publishing, Cham, 2016, pp. 30–41.
- [10] Z. C. Lipton, D. C. Kale, C. Elkan, R. C. Wetzal, Learning to diagnose with LSTM recurrent neural networks, in: Y. Bengio, Y. LeCun (Eds.), *4th International Conference on Learning Representations, ICLR 2016, San Juan, Puerto Rico, May 2-4, 2016, Conference Track Proceedings*, 2016.
- [11] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, J. Sun, Doctor AI:

- Predicting clinical events via recurrent neural networks., in: F. Doshi-Velez, J. Fackler, D. Kale, B. Wallace, J. Wiens (Eds.), Proceedings of the 1st Machine Learning for Healthcare Conference, volume 56 of *Proceedings of Machine Learning Research*, PMLR, 2016, pp. 301–318.
- [12] A. Ashfaq, A. Sant’Anna, M. Lingman, S. Nowaczyk, Readmission prediction using deep learning on electronic health records, *Journal of biomedical informatics* 97 (2019) 103256.
 - [13] D. B. da Silva, D. Schmidt, C. A. da Costa, R. da Rosa Righi, B. Eskofier, Deepsigns: A predictive model based on deep learning for the early detection of patient health deterioration, *Expert Systems with Applications* 165 (2021) 113905.
 - [14] Y. Sha, M. D. Wang, Interpretable predictions of clinical outcomes with an attention-based recurrent neural network, in: Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics, 2017, pp. 233–240.
 - [15] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, L. E. Barnes, Patient2vec: A personalized interpretable deep representation of the longitudinal electronic health record, *IEEE Access* 6 (2018) 65333–65346.
 - [16] Y. Cheng, F. Wang, P. Zhang, J. Hu, Risk prediction with electronic health records: A deep learning approach, in: Proceedings of the 2016 SIAM International Conference on Data Mining, SIAM, 2016, pp. 432–440.

- [17] P. Nguyen, T. Tran, N. Wickramasinghe, S. Venkatesh, **Deepr**: A convolutional net for medical records, *IEEE J. Biomed. Health Informatics* 21 (2017) 22–30.
- [18] X. Li, H. Wang, H. He, J. Du, J. Chen, J. Wu, Intelligent diagnosis with chinese electronic medical records based on convolutional neural networks, *BMC bioinformatics* 20 (2019) 62.
- [19] S. M. Lauritsen, M. E. Kalør, E. L. Kongsgaard, K. M. Lauritsen, M. J. Jørgensen, J. Lange, B. Thiesson, Early detection of sepsis utilizing deep learning on electronic health record event sequences, *Artificial Intelligence in Medicine* 104 (2020) 101820.
- [20] An attention-based cnn-bilstm hybrid neural network enhanced with features of discrete wavelet transformation for fetal acidosis classification, *Expert Systems with Applications* 186 (2021) 115714.
- [21] A. Zeroual, F. Harrou, A. Dairi, Y. Sun, Deep learning methods for forecasting covid-19 time-series data: A comparative study, *Chaos, Solitons Fractals* 140 (2020) 110121.
- [22] H. Verma, S. Mandal, A. Gupta, Temporal deep learning architecture for prediction of covid-19 cases in india, *Expert Systems with Applications* 195 (2022) 116611.
- [23] S.-C. Huang, A. Pareek, S. Seyyedi, I. Banerjee, M. Lungren, Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines, *NPJ Digital Medicine* 3 (2020).

- [24] S. E. Spasov, L. Passamonti, A. Duggento, P. Liò, N. Toschi, A multi-modal convolutional neural network framework for the prediction of alzheimer’s disease, in: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2018, pp. 1271–1274.
- [25] A. Yala, C. Lehman, T. Schuster, T. Portnoi, R. Barzilay, A deep learning mammography-based model for improved breast cancer risk prediction, *Radiology* 292 (2019) 182716.
- [26] I. Reda, A. Khalil, M. Elmogy, A. Aboelfetouh, A. Shalaby, M. Abou-El-Ghar, A. Elmaghraby, M. Ghazal, A. El-Baz, Deep learning role in early diagnosis of prostate cancer, *Technology in Cancer Research & Treatment* 17 (2018) 153303461877553.
- [27] J. X. Qiu, H.-J. Yoon, P. A. Fearn, G. D. Tourassi, Deep learning for automated extraction of primary sites from cancer pathology reports, *IEEE Journal of Biomedical and Health Informatics* 22 (2018) 244–251.
- [28] Z. Qiao, X. Wu, S. Ge, W. Fan, MNN: Multimodal attentional neural networks for diagnosis prediction, in: *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, 2019, pp. 5937–5943.
- [29] A. Bagheri, T. K. J. Groenhof, W. B. Veldhuis, P. A. de Jong, F. W. Asselbergs, D. L. Oberski, Multimodal learning for cardiovascular risk prediction using ehr data, in: *Proceedings of the 11th ACM International Conference on Bioinformatics, Computational Biology and Health*

Informatics, BCB '20, Association for Computing Machinery, New York, NY, USA, 2020.

- [30] S. Shen, S. X. Han, D. R. Aberle, A. A. Bui, W. Hsu, An interpretable deep hierarchical semantic convolutional neural network for lung nodule malignancy classification, *Expert Systems with Applications* 128 (2019) 84–95.
- [31] F. A. Zeiser, C. A. da Costa, G. de Oliveira Ramos, H. C. Bohn, I. Santos, A. V. Roehe, Deepbatch: A hybrid deep learning model for interpretable diagnosis of breast cancer in whole-slide images, *Expert Systems with Applications* 185 (2021) 115586.
- [32] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, P. Rashidi, Deepsofa: a continuous acuity score for critically ill patients using clinically interpretable deep learning, *Scientific reports* 9 (2019) 1–12.
- [33] F. Ma, R. Chitta, J. Zhou, Q. You, T. Sun, J. Gao, 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*, 2017, pp. 1903–1911.
- [34] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, W. Stewart, Retain: An interpretable predictive model for healthcare using reverse time attention mechanism, in: *Advances in Neural Information Processing Systems*, 2016, pp. 3504–3512.

- [35] D. A. Kaji, J. R. Zech, J. S. Kim, S. K. Cho, N. S. Dangayach, A. B. Costa, E. K. Oermann, An attention based deep learning model of clinical events in the intensive care unit, *PloS one* 14 (2019) e0211057.
- [36] L. Men, N. Ilk, X. Tang, Y. Liu, Multi-disease prediction using lstm recurrent neural networks, *Expert Systems with Applications* 177 (2021) 114905.
- [37] E. Choi, M. T. Bahadori, E. Searles, C. Coffey, M. Thompson, J. Bost, J. Tejedor-Sojo, J. Sun, Multi-layer representation learning for medical concepts, *KDD '16*, Association for Computing Machinery, New York, NY, USA, 2016, p. 1495–1504.
- [38] R. Miotto, L. Li, B. A. Kidd, J. T. Dudley, Deep patient: An unsupervised representation to predict the future of patients from the electronic health records, in: *Scientific reports*, 2016.
- [39] W. Farhan, Z. Wang, Y. Huang, S. Wang, F. Wang, X. Jiang, A predictive model for medical events based on contextual embedding of temporal sequences, *JMIR Med Inform* 4 (2016) e39.
- [40] L. Chunyu, L. Ran, Z. Junteng, W. Miye, X. Jing, S. Lan, Z. Yixuan, Z. Rui, F. Yizhou, W. Chen, Y. Hongmei, Z. Qing, Characterizing the critical features when personalizing antihypertensive drugs using spectrum analysis and machine learning methods, *Artificial Intelligence in Medicine* 104 (2020) 101841.
- [41] S. Basu Roy, A. Teredesai, K. Zolfaghar, R. Liu, D. Hazel, S. Newman, A. Marinez, Dynamic hierarchical classification for patient risk-

- of-readmission, in: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, Association for Computing Machinery, New York, NY, USA, 2015, p. 1691–1700.
- [42] W. Farhan, Z. Wang, Y. Huang, S. Wang, F. Wang, X. Jiang, A predictive model for medical events based on contextual embedding of temporal sequences, *JMIR Medical Informatics* 4 (2016) e39.
- [43] X.-H. Li, C. C. Cao, Y. Shi, W. Bai, H. Gao, L. Qiu, C. Wang, Y. Gao, S. Zhang, X. Xue, L. Chen, A survey of data-driven and knowledge-aware explainable ai, *IEEE Transactions on Knowledge and Data Engineering* (2020) 1–1.
- [44] F. K. Došilović, M. Brčić, N. Hlupić, Explainable artificial intelligence: A survey, in: 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2018, pp. 0210–0215.
- [45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser, I. Polosukhin, Attention is all you need, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), *Advances in Neural Information Processing Systems*, volume 30, Curran Associates, Inc., 2017.
- [46] L. Rasmy, Y. Xiang, Z. Xie, C. Tao, D. Zhi, Med-BERT: pretrained contextualized embeddings on large-scale structured electronic health records for disease prediction, *NPJ digital medicine* 4 (2021) 1–13.

- [47] N. Sendi, N. Abchiche-Mimouni, F. Zehraoui, Towards a transparent deep ensemble method based on multiagent argumentation, in: D. Calvaresi, A. Najjar, M. Schumacher, K. Främling (Eds.), *Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, Springer International Publishing, Cham, 2019, pp. 3–21.