A BLSTM with Attention Network for Predicting Acute Myeloid Leukemia Patient's Prognosis using Comprehensive Clinical Parameters

Chih-Chuan Lu[1], Jeng-Lin Li[1], Yu-Fen Wang[2], Bor-Sheng Ko[3], Jih-Luh Tang[2,3], and Chi-Chun Lee[1]

Abstract— The prognosis management is crucial for high-risk disease like Acute Myeloid Leukemia (AML) in order to support decisions of clinical treatment. However, the challenges of accurate and consistent forecasting lie in the high variability of the disease outcomes and the complexity of the multiple clinical measurements available over the course of the treatment. In order to capture the multi-dimensional and longitudinal aspect of these comprehensive clinical parameters, we utilize an attention-based bi-directional long short-term memory (Att-BLSTM) network to predict AML patient's survival and relapse. Specifically, we gather a 10-year worth of real patient's clinical data including blood test, medication, HSCT status, and gene mutation information. Our proposed Att-BLSTM framework achieves 77.1% and 67.3% AUC in tasks of predicting the next 2-year mortality and disease relapse with these comprehensive clinical parameters, and our further analysis demonstrates that a next 0 to 3 months prediction performs equally well, i.e., 74.8% and 67% AUC for mortality and relapse respectively.

I. Introduction

Acute Myeloid Leukemia (AML) is the most common type of leukemia disease notoriously known for its poor prognosis outcome, i.e., low survival rate (below 25% in 5 years after diagnosis) and high relapse rate (about 50%). Major clinical treatments rely on intensive chemotherapy and allogeneic hematopoietic stem cell transplantation (HSCT). However, outcomes of the treatment vary greatly from patient to patient; e.g., it is known that younger and healthier patients may extend their remission duration as compared to older patients, and different selections of chemo drugs may result in complete remission (CR) but also risk of mortality. Hence, the prognosis management and treatment plan are often jointly considered in the current clinical setting to handle such potential life-threatening risk for patients while undergoing standard intervention strategy.

Forecasting a patient's occurrence of death and relapse over the treatment course is critical and often assessed by the physician based on the patient's prognostic factors. Challenges in obtaining accurate and consistent prognosis management come from issues of complex integrative assessment of heterogeneous and longitudinal clinical variables, including images, laboratory test results, medical records,

and even the interviews between patients and doctors. It is further troubled by the occurrences in the variety of treatment outcomes even when receiving similar therapies [1]. Differences in the individual doctor's clinical experiences and the overwhelming clinical parameters available lead to a current situation that there exists no consensus and standard guideline to approach a clinical prognostic decision.

Most of the prior research in prognostic stratification is based on statistical analysis of conventional risk factors, such as demographic, peripheral blood, and cytogenetic [2], [3], [4]. In order to improve prediction accuracy, researchers have started to explore machine learning techniques. For example, Gupta et al. have used machine learning methods to predict survival rates of various diseases using electronic administrative records [5]. Pan et al. have investigated sociodemographic, clinical, immunological and cytogenetic variables using random forest classifier to predict relapse of acute lymphoblastic leukemia [6]. Recently, Lin et al. have proposed to predict AML patient's diagnosis to death using cytogenetics, age and mutations with a deep learning model [7]. While various works demonstrate promising applications of machine learning for outcome prediction, most if not all of these works consider the patient's clinical variables as static attributes without modeling their temporal aspects.

In this work, we propose an attention-based bi-directional long short-term memory (Att-BLSTM) that models comprehensive aspects (5 major dimensions) of clinical variables of an AML patient over his/her treatment course to predict the prognostic outcomes, specifically mortality and relapse. We use a dataset collected retrospectively from National Taiwan University Hospital over a 10-year window consists of blood test, medication usage, HSCT status, and gene mutation information. Our method obtains 77.1% and 67.3% AUC for 2-year mortality and relapse prediction and 74.8% and 67% AUC for 3-month mortality and relapse prediction respectively. To the best of our knowledge, this is one of the first works that presents a longitudinal deep learning approach by using multimodal (over 5 clinical events' records) to predict AML patient's outcomes. The rest of this paper is organized as follows: Methodology and experiments would be detailed in Section II. Section III shows the results analysis and Section IV concludes this work and future work.

II. Methodology

Figure 1 illustrates our methodological framework. We further explain the following components in details: the database description and preprocessing method, the Att-BLSTM, and final outcome prediction.

[1] CC Lee, CC Lu, JL Li are with Department of Electrical Engineering, National Tsing Hua University, Hsinchu, Taiwan, and Joint Research Center for AI Technology and All Vista Healthcare, Ministry of Science and Technology, Taiwan (phone: +88635162439. e-mail: cclee@ee.nthu.edu.tw, {lu721, cllee}@gapp.nthu.edu.tw).

[2] YF Wang, JL Tang are with Tai-Cheng Stem Cell Therapy Center, National Taiwan University, Taipei, Taiwan.

[3] BS Ko, JL Tang are with Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan.
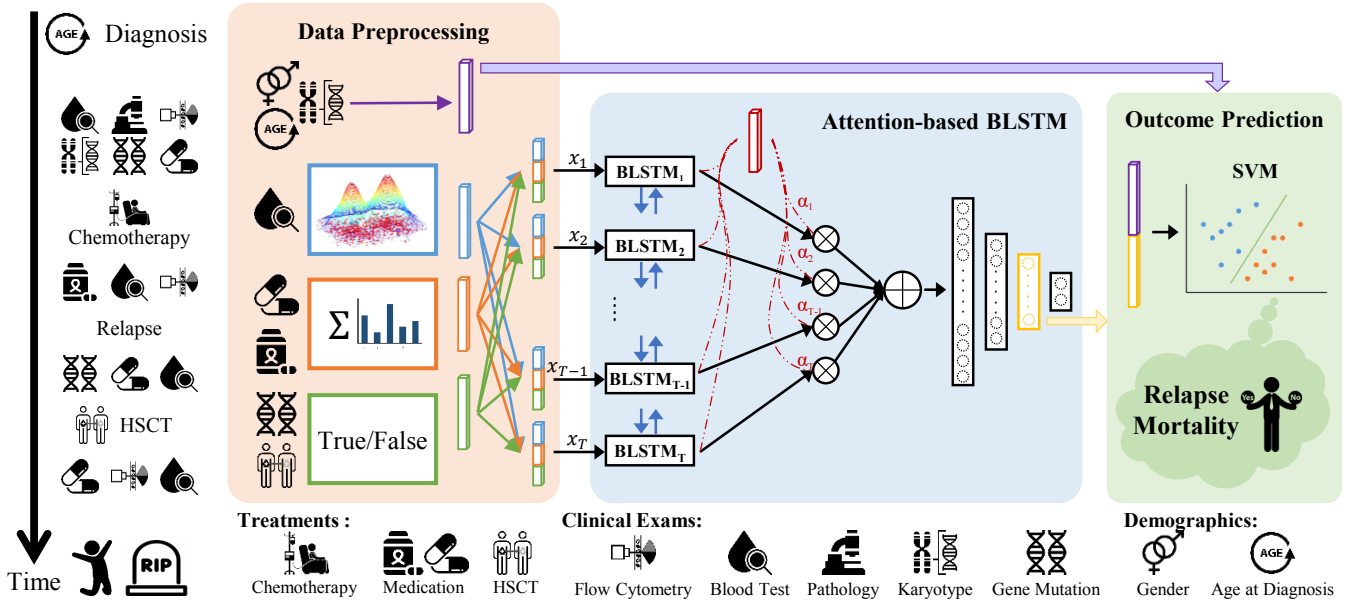
Fig. 1: An illustration of our framework for mortality and relapse prediction using Att-BLSTM.

A. Database and Preprocessing

We collect retrospective clinical data of AML patients during the periods of Jan. 2006 to Feb. 2017 from the Integrated Medical Database, National Taiwan University Hospital (NTUH-iMD). There are 637 out of 913 total patients meet the minimal preprocessing requirement, i.e., at least 3 months follow-up duration since diagnosis. Each patient has both personal *static* and *time-dependent* variables. The personal static variables include basic demographics (age at diagnosis and gender), as well as first cytogenetics test at diagnosis. The cytogenetics test categorizes each patient into 3 risk groups: favorable, intermediate, adverse. Table I includes the distribution of all static personal features and key treatment and status of the data used in this work. On the other hand, time-dependent variables consist of laboratory results of complete blood count (CBC), white blood composition (WBC), gene mutation, and treatment history such as allogeneic HSCT and medication history. Our study is approved by the Research Ethic Committee of the National Taiwan University Hospital.

For the time-dependent variables, each variable is processed differently to generate features as input to the Att-BLSTM network. For every 10-day window (termed as a time step), we gather each patient's sequence of the following measured items and encode each of them to form a single vector as input. The brief description of encoding approach for each item is shown below:

- **CBC&WBC:** The exam result includes 9 and 12 dimensions for CBC and WBC respectively. We first gather all available records (i.e., may be different amount for each patient within each time step). Then, for every time step, we encode the sequences of blood test result measurements into a fixed length vector using a technique based on Gaussian Mixture Model based Fisher-vector encoding (GMM-FV) [8].
- **Medication:** There are a total of 30 types of anti-

neoplastic medications identified according to L01 and L03 ATC code. The summation of dose usage is calculated according these ATC codes within each time step to be used as features. Additionally, ANOVA F-test feature selection is performed to identify the most informative subset of medication-related features.

- **HSCT:** We have the date of HSCT and whether relapse occurs after HSCT. Each time step will be given a binary feature value indicating whether the patient has received HSCT and without relapse, i.e., in that particular time step, a value of 1 is given if a patient has gone through HSCT and no relapse has occurred and a value of 0 otherwise.
- **Gene mutation:** The test examines 10 types of genes indicating whether there is mutant. We generate a 10-dimensional feature vector for each time step indicating whether the particular gene type has mutated or not.

In this work, we take each patient's 3-month worth of data since diagnosis and before first CR as input (resulting in a 9-time-step sequence of feature inputs derived from concatenating the above 5-item time-dependent clinical features vectors) for our mortality and relapse outcome prediction tasks. Table II summarizes the key statistics for all static and time-dependent input features.

B. Attention-based BLSTM Network

In this work, we utilize an Att-BLSTM network proposed by Zhou et al. [9] to model the sequence of time-dependent feature vectors. The BLSTM is an improved version of LSTM by considering both forward and backward time-dependent relationship to ensure the temporal gradient can be equally transmitted. The use of attention mechanism can be thought of as having a learnable weight to emphasize the important part of the sequence output from BLSTM.

Given the input feature sequence of T time step $x = \{x_1, x_2, ..., x_T\}$. The corresponding output sequence h_t of

TABLE I: The distribution of key population statistics used in this study.

	Total	Gender		Age at Diagnosis			Cytogenetics Risk Group				CR		Relapse		HSCT	
		M	F	-30	30-60	60-	Favourable	Intermidiate	Adverse	N/A	Y	N	Y	N	Y	N
N	637	215	322	66	367	204	46	420	107	64	482	155	253	384	462	175
%	100	33.8	66.2	10.7	57.6	32.4	7.2	65.9	16.8	10.0	75.7	23.3	39.7	60.3	72.5	27.5

tth time step from BLSTM layers is the concatenation of forward and backward LSTM output sequence $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$

$$h_i = [\overrightarrow{h_t}; \overleftarrow{h_t}] \tag{1}$$

Then, the attention weight $\alpha \in \mathbb{R}^T$ is calculated by:

$$\alpha_t = \frac{exp(u^T h_t)}{\sum_t^T exp(u^T h_t)} \tag{2}$$

where u is a trainable vector having the same dimension as h_t and u^T is its transpose. This attention weight for each time step t is integrated back to the BLSTM output h_t to complete the following equation:

$$s = tanh(\sum_t^T \alpha_t h_t) \tag{3}$$

The output s is then input to a stacked fully connected layers with rectified linear units (ReLU) as activation function, and the latent layer before network output as final Att-BLSTM representation s^*.

$$s^* = relu(FC(s)) \tag{4}$$
$$\tilde{y} = softmax(FC(s^*)) \tag{5}$$

The network parameter is learned by minimizing cross-entropy loss between true label y and network output $\tilde{y}$.

C. Outcome Prediction

Once the Att-BLSTM network is trained, we can input the processed time-dependent encoded feature vector (section II-A) to generate the representation s^* (section II-B) for each patient. The final prediction model is based on training a support vector machine (SVM) with linear kernel by input the concatenation of time-series representation s^* and static features (section II-A).

D. Experimental Setup

There are two prognosis outcomes as prediction targets: mortality and relapse. We collect 3-month worth of patient's data from diagnosis date and before first CR date to derive our training set features. The prediction target is whether

TABLE II: A summary on the key statistics of our feature set used in this work (see section II-A).

Charactiristics	Dimension	Total	Subset	
			Mortality	Relapse
Patient Number		913	637	482
Demographic	2	913	637	482
Cytogenetic	1	2223	573	488
CBC & WBC	9+12	50633	11193	6314
Medication	30	98519	19689	10317
HSCT	1	462	29	30
Gene Mutation	10	1769	911	657

the patient would survive and relapse within the coming N months. The GMM-FV encoding is computed with Gaussian mixture number set to 4. We utilize ADAM optimizer [10] with an initial learning rate of 0.00005 in learning the Att-BLSTM network. The size of mini-batch is 16.

We conduct two different experiments. Firstly, we compare our proposed framework in 2-year outcome (i.e., whether the patient would survive or relapse within the next 2 years) prediction tasks with three other methods listed below as well as different input modalities mentioned in II-A. Secondly, we investigate the accuracy obtained by varying different targeted future prediction periods, i.e., 0-3, 3-12, and 12-24 months.

- **SVM:** The time-dependent representation is directly concatenated from the encoded features without using Att-BLSTM. Then, this time-dependent representation is concatenated with static features to train a SVM classifier.
- **LR:** It is similar to SVM but the classifier is changed to Logistic Regression (LR).
- **BLSTM:** The time-dependent representation is learned using BLSTM without attention mechanism. Then, the representation is concatenated with static feature to train a SVM classifier.
- **Att-BLSTM:** Our proposed framework.

The metric used in this work is unweighted accuracy (UAR) and area under receiver operating characteristic curve (AUC). We use a 5-fold patient-independent cross validation scheme for all of our experiments. In each fold, 80% samples are using as training data, and the rest 20% samples are using to evaluate the performance.

III. RESULTS

In this work, there are 913 patients in total. Due to the different follow-up duration and treatment condition, there are different numbers of patients included in each experiment. To meet the minimal data requirement, i.e., 3-month follow-up, 637 patients are included, in which 482 patients achieved CR. Figure 2 summarizes the accuracy of our first experiment mentioned in Section II-D. Our proposed method obtains the best accuracy using comprehensive clinical parameters, i.e., all of the input items: 77.1% (AUC) and 71.4% (UAR) for mortality prediction and 67.3% (AUC) and 62.8% (UAR) for relapse prediction. By comparing with conventional machine learning methods, i.e., **SVM** and **LR** methods, it is evident that BLSTM-based technique provides a more discriminative representation that learns to predict better on time-series data. Moreover, by further integrating BLSTM with attention mechanism, the prediction accuracy further improves.

In our second experiment, we observe that our proposed framework can not only obtain a better accuracy than other
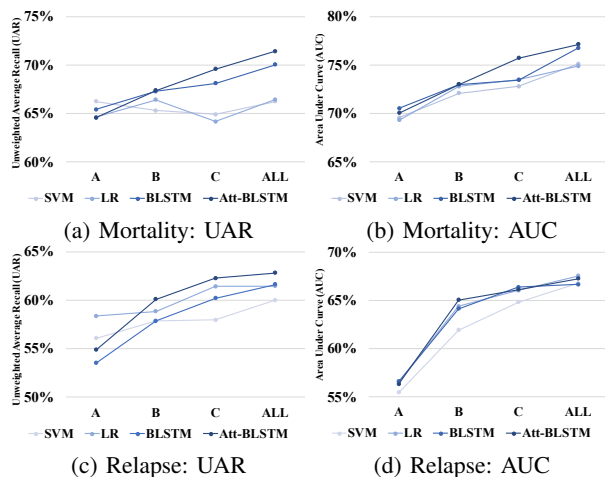
(a) Mortality: UAR (b) Mortality: AUC

(c) Relapse: UAR (d) Relapse: AUC

Fig. 2: The illustration of result comparison of **SVM**, **LR**, **BLSTM**, and **Att-BLSTM** with different input feature settings: **A:** CBC&WBC, **B:** B+medications, **C:** B+cytogenetics+gene mutation, **ALL:** C+demographics+HSCT

baseline machine learning methods, but also maintain its modeling power when training it with different target prediction periods. Table III shows accuracy results obtained for second experiment. The best accuracy occurs when predicting the outcome in the target period of the coming 0 to 3 months: 74.8% (AUC) for mortality and 67.0% (AUC) for relapse. We notice that when predicting the patient's outcome in the next 3 to 12 months and 1 to 2 years, the lower accuracy may partly due to the inadequate number of available data that meets the longer follow-up duration requirement. More importantly, various treatments that would occur prior to that target periods but not included in our training features potentially have a larger effect on predicting the patient's final outcome.

IV. DISCUSSIONS AND CONCLUSIONS

In this work, we present one of first study in utilizing time-dependent deep learning model learned from a retrospective collection of 10-year worth of real patient's comprehensive clinical variables to predict AML patients' outcome in order to address the current clinical challenges in advancing the AML treatment with better prognosis prediction. Our proposed method is capable of integrating both static attributes and time-dependent progression of clinical variables via attention-based BLSTM model. It consistently outperforms other methods without considering temporal aspect. It further provides a methodological approach in integrating heterogeneous and longitudinal clinical variables that are challenging to be modeled using conventional statistical methods.

There are multiple future directions: first, there are over a hundred exam items used during an AML patient's treatment course. We will further explore additional items from the electronic database to perform the prognosis prediction. Second, we would like to include multi-site clinical data to expand our population cohort and at the same time to evaluate the robustness of our prognosis prediction framework. Our aim is to assist physician to better stratify the mortality

TABLE III: Results of mortality and relapse prediction in the 0-3, 3-12, and 12-24 months using CBC&WBC, medications, HSCT, and gene mutations, demographics, and cytogenetics.

Mortality					
N	Total	Label		UAR	AUC
(Months)	(Samples)	Alive	Death	(%)	(%)
0-3	637	598	39	70.1	74.8
3-12	597	457	140	65.2	71.0
12-24	444	367	77	61.4	69.9

Relapse					
N	Total	Label		UAR	AUC
(Months)	(Samples)	Remission	Relapse	(%)	(%)
0-3	482	443	39	62.4	67.0
3-12	482	357	125	55.6	61.3
12-24	482	440	42	56.7	62.3

and relapse risk through comprehensive machine learning approach based on the multi-variate clinical variables measured over the time course of each patient individually, and eventually help provide predictive analytics in positively affecting the needed treatment strategy.

REFERENCES

[1] C. S. Hourigan, R. Peter Gale, and R. B. Walter, "Refining aml outcome prediction," *Leukemia*, p. 1, 12 2018.

[2] T. Herold, V. Jurinovic, A. M. Batcha, S. A. Bamopoulos, M. Rothenberg-Thurley, B. Ksienzyk, L. Hartmann, P. A. Greif, J. Phillippou-Massier, S. Krebs, *et al.*, "A 29-gene and cytogenetic score for the prediction of resistance to induction treatment in acute myeloid leukemia," *haematologica*, vol. 103, no. 3, pp. 456–465, 2018.

[3] M. L. Sorror, B. E. Storer, A. T. Fathi, A. T. Gerds, B. C. Medeiros, P. Shami, A. M. Brunner, M. A. Sekeres, S. Mukherjee, E. Peña, *et al.*, "Development and validation of a novel acute myeloid leukemia–composite model to estimate risks of mortality," *JAMA oncology*, vol. 3, no. 12, pp. 1675–1682, 2017.

[4] C. D. Godwin, M. Othus, M. A. Powell, S. A. Buckley, E. H. Estey, and R. B. Walter, "Prediction of early death in adults with relapsed or refractory acute myeloid leukemia," *Leukemia & lymphoma*, vol. 57, no. 10, pp. 2421–2424, 2016.

[5] S. Gupta, T. Tran, W. Luo, D. Phung, R. L. Kennedy, A. Broad, D. Campbell, D. Kipp, M. Singh, M. Khasraw, *et al.*, "Machine-learning prediction of cancer survival: a retrospective study using electronic administrative records and a cancer registry," *BMJ open*, vol. 4, no. 3, p. e004007, 2014.

[6] L. Pan, G. Liu, F. Lin, S. Zhong, H. Xia, X. Sun, and H. Liang, "Machine learning applications for prediction of relapse in childhood acute lymphoblastic leukemia," *Scientific reports*, vol. 7, no. 1, p. 7402, 2017.

[7] M. Lin, V. Jaitly, I. Wang, Z. Hu, L. Chen, M. Wahed, Z. Kanaan, A. Rios, A. N. Nguyen, *et al.*, "Application of deep learning on predicting prognosis of acute myeloid leukemia with cytogenetics, age, and mutations," *arXiv preprint arXiv:1810.13247*, 2018.

[8] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: Theory and practice," *International journal of computer vision*, vol. 105, no. 3, pp. 222–245, 2013.

[9] P. Zhou, W. Shi, J. Tian, Z. Qi, B. Li, H. Hao, and B. Xu, "Attention-based bidirectional long short-term memory networks for relation classification," in *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, vol. 2, 2016, pp. 207–212.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proceedings of the International Conference on Learning Representations (ICLR)*, 2015.