International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

A Focus on the ICU's Mortality Prediction Using a CNN-LSTM Model

[1]Sakshi Hooda, [2]Suman Mann

ABSTRACT--*In healthcare, the ability to predict mortality accurately is an important aspect because it allows for empirical risk estimations, allowing further for prognostic hospital benchmarking, patient stratification, and decision-making. Currently, most of the techniques for prediction come in the form of scoring systems seeking to determine disease severity. Also, the models involve certain, rigid summarized physiological data and admission attributes. The implication is that there tends to be some degree of biasness in the systems, especially because they involve manual effort that also demands regular updating to address shortcomings that are reported frequently. Therefore, deep learning algorithms have evolved in a quest to allow for the automatic extraction of features, as well as their selection independent of human intervention. In this study, an approach combining notes for mortality prediction and a deep learning algorithm was proposed. The role of the CNN-LSTM entailed the mapping of various notes in relation to possibilities of mortality outcomes. In the custom architecture, there is a combination of recurrent and convolutional layers with previous capturing semantic associations in certain notes on an independent basis. Therefore, this experimental study strived to determine the performance of the CNN-LSTM approach towards mortality prediction via the use of notes for the initial 48, 24, and 12 hours of a given patient's hospital stay. In the findings, the study established that when CNN-LSTM is implemented, it exhibits superior performance when compared to the baseline. Thus, there was evidence of a proof-of-concept in relation to the efficacy of combining deep learning and notes towards improvements in outcome prediction.*

Keywords—*ICU's, Mortality Prediction, CNN-LSTM*

I. INTRODUCTION

The critical care unit or intensive care unit (ICU) forms one of the hospitals' special departments, especially because they seek to allow for patients with acute life-threatening health conditions to be served [1-3]. Indeed, in these settings, there tends to be a round-the-clock monitoring via the utilization of specially trained staff [4], as well as sophisticated equipment [5]. Also, ICUs exist in different types, including surgical and neonatal sections [6-8]. In most cases, even in the wake of differences in basic services, ICUs house multidisciplinary teams because of different medical specialties that these healthcare settings demand [9, 10].

In the current world, there is an increasing trend in data-driven medicine, using clinical data available towards healthcare decision-making augmentation, as well as information personalization and incorporation into medical practices. Similarly, different data-driven medicine approaches have been employed in areas such as machine

[1]*Research Scholar, IPU, New Delhi. Email id- sakshi.hooda@gmail.com*
[2]*Associate Professor, MSIT, New Delhi. Email-ID-sumanmann2007@gmail.com*

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

learning and artificial intelligence. Through this incorporation, there has been significant progress in precision medicine actualization [3-5].

Another important point that is noted is that the measurement of healthcare institutions' quality of care tends to be achieved in most cases via the determination of an outcome of mortality rate [9, 10]. In the ICU, mortality rate allows for unbiased estimations of empirical risks regarding clinical decisions. Also, it allows physicians to estimate potential mortality and prepare prognoses and plans of action while prioritizing patient care in an informed way. Previously, some of the scoring systems that have been proposed include MPM II, SAPS II, APACHE III, and APACHE II. Most of these systems discern disease severity by combining several features relative to psychological attributes and measurements of patients, eventually deriving mortality probabilities – after examining or assessing the severity score. Despite the promising nature of these previously proposed scoring systems, however, is that they exhibit certain deficiencies. These deficiencies include poor generalizability, effects of salience bias, problems associated with patient data, the manual selection of features, and the failure to account for certain temporal correlations in existence. In this study, an approach combining notes for mortality prediction and a deep learning algorithm was proposed. The role of the CNN-LSTM entailed the mapping of various notes in relation to possibilities of mortality outcomes. In the custom architecture, there is a combination of recurrent and convolutional layers with previous capturing semantic associations in certain notes on an independent basis. Therefore, this experimental study strived to determine the performance of the CNN-LSTM approach towards mortality prediction via the use of notes for the initial 48, 24, and 12 hours of a given patient's hospital stay.

II. Methodology

In this study, the main sequential components were three. They included modeling, experimental setup, and data preparation. During the data preparation stage, there was input data preprocessing, whereby it would be converted to an ideal form on which the proposed model would be applied. There was also the notes' fastText vector pre-training. For the case of the experimental setup stage, there was the experimental design process, as well as the specification of the criteria for cohort selection. Lastly, the modeling stage involved the development and implementation of the proposed CNN-LSTM architecture, upon which there would be the calculation of the mapping between the mortality result and the input notes. The figure below summarizes the methodological procedure that was adopted.

Figure 1: Summary of the methodological pipeline for implementing and evaluating the proposed model

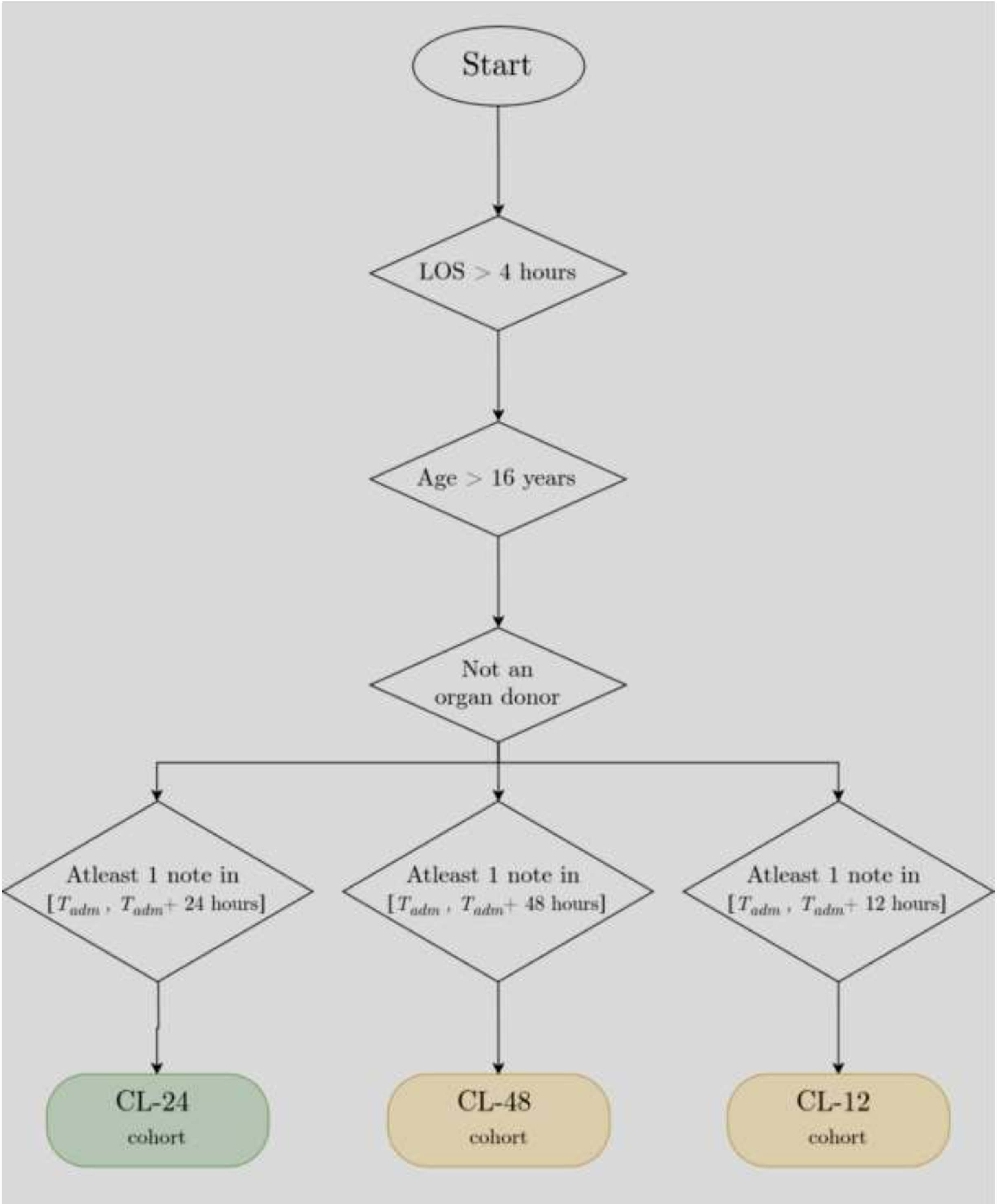

Figure 2: Model cohort selection criteria flowchart

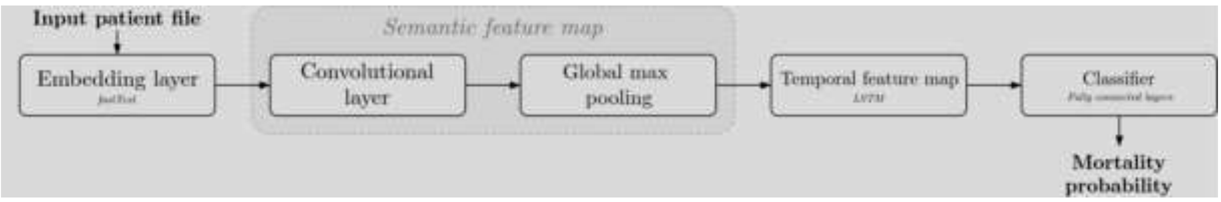

Figure 3: Components of the proposed model

III. Results and Discussion

To discern the nature of the proposed algorithm's performance, there was a comparison of its capacity for the task of mortality prediction with the case of the behavior of the baseline. Imperatively, the CNN-LSTM algorithm had its mortality prediction performance compared to various scoring systems applied to the parameter of disease severity, and they included OASIS, APS III, and SAPS II. Also, the performance metrics that were used towards the evaluation of the proposed algorithm and the baseline included the area under the precision-recall curve (AUPRC) and the area under the ROC curve (AUROC).

Regarding the AUROC parameter, the proposed model achieved a value of 0.85. At this point, this value was deemed significantly higher when compared to the remainder of the algorithms with which it was compared. Also, an even higher AUROC was realized during experiment CL-48, with experiment CL-12 demonstrating the framework as competitive, promising results that were obtained even at a time when the scoring systems used data from a longer period of time, compared to this model, which relied on data associated with a lesser period of time.

Experiment	CNN-LSTM	SAPS II	OASIS	APS III
CL-24	**0.85**	0.81	0.79	0.79
CL-12	**0.82**	0.81	0.78	0.79
CL-48	**0.87**	0.81	0.78	0.78

Table 1: Average AUROC comparative outcomes between the proposed model and the baseline techniques (for each fold)

Another parameter that was investigated involved the case of AUPRC. For the case of CNN-LSTM mode, a significantly greater AUPRC was observed, compared to the remainder of the algorithms with which it was compared. Of importance to note is that at this point, an imbalanced state of dataset was observed because for each mortal case, survival cases were 9. The eventuality is that the behavior of CNN-LSTM demonstrated a model that could be associated with generalization capability and higher discrimination – concerning the subject of mortality prediction.

Experiment	CNN-LSTM	SAPS II	OASIS	APS III
CL-24	**0.51**	0.41	0.39	0.36
CL-12	**0.45**	0.41	0.39	0.37
CL-48	**0.58**	0.40	0.39	0.36

Table 2: Average AUPRC comparative outcomes for CNN-LSTM performance relative to other mortality prediction models

IV. Conclusion

From the findings, it is worth noting that the CNN-LSTM approach is advantageous because of its ease of adaptability and setup. Therefore, the model was found to be relatively easier because the input involves notes, compared to the case of other techniques or scoring systems that are seen to demand different collections of patient attributes and physiological variables from the internal databases of different hospitals. It is also notable that a retrospective study could be used to develop the CNN-LSTM model, rather than demand expensive studies characterized by regulatory compliance protocols and complex selection criteria. It can be seen further that the capability of CNN-LSTM to adapt to various scenarios is high because of transfer learning in the machine learning algorithm. The implication is that hospitals could utilize CNN-LSTMs that have been trained on the data of other hospitals, courtesy of this adaptation capability – to different situations. Therefore, the merit makes the framework advantageous even further to healthcare institutions with less data, as they could utilize CNN-LSTM models developed in other hospitals to apply to their systems for mortality prediction. In the future, there is a need for more scholarly investigations to focus on the subject of some of the ways in which the proposed framework could be improved even further (towards mortality prediction) via architectural changes. Overall, it is concluded that when CNN-LSTM is implemented, it exhibits superior performance when compared to the baseline. Thus, there was evidence of a proof-of-concept in relation to the efficacy of combining deep learning and notes towards improvements in outcome prediction.

REFERENCES

1. Boag, W., Doss, D., Naumann, T., and Szolovits, P. (2018). What's in a note? Unpacking predictive value in clinical note representations. AMIA Jt Summits Transl Sci Proc, 2017:26-34.
2. Bojanowski, P., Grave, E., Joulin, A., and Mikolov, T. (2016). Enriching word vectors with subword information. arXiv preprint arXiv:1607.04606.
3. Brown, M. C. and Crede, W. B. (1995). Predictive ability of acute physiology and chronic health evaluation II scoring applied to human immunode_ciency virus-positive patients. Crit. Care Med., 23(5):848-853.
4. Chen, L. M., Martin, C. M., Morrison, T. L., and Sibbald, W. J. (1999). Interobserver variability in data collection of the APACHE II score in teaching and community hospitals. Crit. Care Med., 27(9):1999-2004.
5. Cheng, J. Y. (2017). Mortality prediction in status epilepticus with the APACHE II score. Pediatr. Crit. Care Med., 18(4):310-317.
6. Cooper, L. M. and Linde-Zwirble, W. T. (2004). Medicare intensive care unit use: analysis of incidence, cost, and payment. Critical care medicine, 32(11):2247-2253.
7. Curtis, J. R., Cook, D. J., Wall, R. J., Angus, D. C., Bion, J., Kacmarek, R., Kane-Gill, S. L., Kirchho_, K. T., Levy, M., Mitchell, P. H., and Others (2006). Intensive care unit quality improvement: A \how-to" guide for the interdisciplinary team. Crit. Care Med., 34(1):211-218.
8. Barrett, M. L., Smith, M. W., A, E., Honigman, L., and Pines, J. M. (2014). Utilization of intensive care services, 2011. HCUP statistical brief #185. Technical report, Agency for Healthcare Research and Quality

9. Haaland, O. A., Lindemark, F., Flaatten, H., Kv_ale, R., and Johansson, K. A. (2014). A calibration study of SAPS II with norwegian intensive care registry data. Acta Anaesthesiol. Scand., 58(6):701-708.
10. Halpern, N. A., Goldman, D. A., Tan, K. S., and Pastores, S. M. (2016). Trends in critical care beds and use among population groups and medicare and medicaid bene_ciaries in the united states: 2000-2010. Crit. Care Med., 44(8):1490-1499.