How to Cite:
Geethamani, D., & Rangaraj, R. (2022). Feature augmented stacked bagging for length of stay prediction on mimic III. *International Journal of Health Sciences*, *6*(S2), 8742–8751.
https://doi.org/10.53730/ijhs.v6nS2.7261

Feature augmented stacked bagging for length of stay prediction on mimic III

D. Geethamani
Assistant Professor, Department of Computer Science, Dr. N.G.P. Arts and Science College (Autonomous), Affiliated to Bharathiyar University, Coimbatore, Tamil Nadu, India

R. Rangaraj
Professor and Head, PG & Research Department of Computer Science, Hindusthan College of Arts and Science Coimbatore, Tamil Nadu, India.

***Abstract*---**Managing hospital resources is one of the major requirements in providing qualitative and timely treatments do patients. From the patient's standpoint, identifying the number of days of stay can aid in effective planning from the monetary perspective.This work presents the Feature Augmented Stacked Bagging (FASB) model for effective prediction of length of stay of patient based on their hospital records. The ensemble model created is composed of two levels; the first level operates on the hospital data and provides the intermediate predictions, while the second level uses predictions from the first level model to perform the final predictions. Experiments where performed using the MIMIC III data, and the results obtained were compared with existing state of the art models for analysis. Comparisons indicate that the proposed model exhibits 95% accuracy levels, which is 13% greater than the compared model. This indicates that the proposed model is highly effective in identifying the length of stay of patients.

***Keywords*---**length stay prediction, ensemble modelling, bagging, stacking, feature augmentation.

Introduction

Increasing of the patient demand for hospitalization and budget restrictions, inpatient bed management has become very challenging for the hospital management team. The major challenge occurs in the allocation of beds in ICU wards. Allocation of appropriate resources, especially in the ICU wards at appropriate time can be lifesaving. Further, managing cost in the ICU department is also critically important, as the ICU department is the one with the highest

Manuscript submitted: 9 March 2022, Manuscript revised: 18 April 2022, Accepted for publication: 1 May 2022

operational cost. Correct identification of Length of Stay (LoS) of patients place a critical role in effectively allocating resources. Length of stay is the time elapsed between admission of the patient and their discharge. High length of stay levels is generally associated with increased risk for mortality and also hospital acquired infections. Further, accurate identification of length of stay can effectively reduce the waiting time for patients.

Estimations for identifying length of stay are currently performed manually in most cases. Increase in the electronic health records and automations can be highly beneficial in automating this process. Automation of health care services, specifically determining the length of stay of patients is of high significance, as it is directly related to determining the mortality levels of the patient. Effective automation of these processes can aid in improving the forecasting efficiency. This can in turn reduce the administrative burden on the hospital management, and can also enable better in low cost scheduling. Initial techniques used automated statistical processes like real time demand capture management (RTDC) [6] for the prediction process. However, these tools exhibit several issues, the major being the fact that they can only work in surgical departments. Several other improved algorithms were later developed for prediction of LOS [7, 8, 9]. Prediction time still remains to be the biggest challenge while working on clinical records. The domain demands highly accurate and faster algorithms. However, data hugeness and the high imbalance levels associated with the domain data proves to be the challenges in designing such algorithms. It has been determine from literature that faster models generally tend to face accuracy issues [10]. This work presents an ensemble modeling technique that can be used to effectively determine the length of stay of patients. The model has been designed as a classification algorithm, and considers 10 days as the threshold for binarizing the problem.

Related works

Length of stay prediction is beginning to receive importance due to the advantages it brings to both the patient and the hospitals. This section discusses some of the recent and significant works in this domain. A convolutional neural network based model for length of stay prediction has been proposed by Rocheteau et al. [11]. This work uses a combination of temporal convolution network and pointwise convolution network. The model concentrates mainly on handling challenges like missing data, data imbalance and irregular sampling of data.Attention based models that concentrate on improving performance in the length of stay prediction process has been proposed by Vaswani et al. [12] and Song et al. [13]. These models operate on clinical data and has successfully proved that attention based models slightly outperform LSTM based models. Single classification model based length of stay prediction has been proposed by Ma et al. [14]. This work uses a combination of just in time learning and one class extreme learning machine for the prediction process.

A neural network based model used for predicting inpatient flow and length of stay prediction has been proposed by He et al. [15]. This work uses inpatient flow as a major factor in determining the length of stay of patient. This work uses the artificial neural networks and model set based on multitask learning technique to perform the predictions. Machine learning based model for predicting length of stay for newborns has been proposed by Thompson et al. [16].This work uses Weka, and analysis nine models available in the Weka environment to perform the predictions. An inpatient length of stay prediction model for brain tumor surgery patients has been proposed by Muhlesteinet al. [17]. This work creates an ensemble model and uses regression technique as the prediction process. A machine learning based model for predicting length of stay for ICU patients has been proposed by Jauk et al. [18]. This work uses a combination of neural network, logistic regression, and gradient boosting technique to perform the analysis. The analysis is performed as a classification task.

A model that operates on codes contained in the electronic health records for length of stay prediction has been proposed by Harerimana et al. [19]. This work categorizes the length of stay levels as short, medium and long. The model also predicts the mortality level of patient based on the current admission details. A hierarchical attention network model has been created for the prediction process. A deep learning model for length of stay prediction on clinical records was proposed by Zebin et al. [20]. The work uses auto encoder dense neural network, and performs the classification of length of stay as being short or long, considering seven as the threshold days. A two level classification model for forecasting the length of stay has been proposed by Livieris et al. [21]. It uses statistical analysis to perform the forecasting process. A three layer neural network model operating on clinical records to predict the length of stay has been proposed by Gentimis et al. [22]. The model uses 5 days as the threshold for the prediction process. An early identification model to determine prolonged ICU stay has been proposed by Kramer et al. [23]. The model considers diagnosis details of the patient to perform the predictions. Other similar techniques used for forecasting the length of stay based on threshold values include works bySotoodeh et al. [24], andChuang et al. [25].

Feature Augmented Stacked Bagging (FASB)

Length of stay prediction on hospital data is a complex process, mainly due to the hugeness of the data and the high complexity associated with the hospital data. This work presents a feature augmentation based stacked bagging model, FASB for effective length of stay prediction. The proposed model is composed of three major phases, the feature argumentation phase, multiple model creation and final prediction using the stacked model.

Feature Augmentation for Training Data Building

This work uses MIMIC data for training the model. Training data is a collection of medical records obtained from various departments in the hospital. The records contain medical details related to diagnostic information, treatment details, symptoms experienced by patient, vitals measured from the patient and stay details [26]. Most of the features were observed to be categorical in nature. In specific, several features are string values. Hence, preprocessing becomes a mandatory step while processing medical information.

Initial process is to identify the file that contains LOS details. Further processes deal with determining features and performing augmentation to enhance the data quality. Age is a date time feature. It is a very significant feature. Hence, the feature has to be converted to integer format to be made usable. The procedures performed on patients and diagnostic information play a vital role in determining the length of stay off the patient. Both the information are represented as strings and also in the form of code numbers. Procedures and diagnosis codes for patients during the particular hospital stay are represented as integers. These values, when appropriately formatted can improve the quality of data to a large extent. Effective augmentation of these features can aid in improved prediction quality. Further, augmentation is also performed based on demographic features and personal details of the patients.

Procedure and diagnostic details of patients are merged with patient personal information and LoS data to obtain the tentative training data. As several data files are merged together, the resultant data bill contains several null values. Records containing null values are eliminated. The data also contains several categorical data representing marital status, codes for diagnosis and procedures, gender, ethnicity, etc. Categorical data cannot be directly used by machine learning models. Hence, they are converted to numerical data using encoding techniques. General encoding techniques increases the number of columns based on the number of distinct entities in the feature. Medical data is already huge. Further increasing the number of features delete to curse of dimensionality. Hence, label encoding is performed for the conversion process. The length of stay feature is represented as a continuous value. Analysis reveals that patients staying for more than 10 days in hospitals exhibits high mortality rates. Hence, 10 days is considered as the threshold limit. The machine learning problem is converted to a binary classification problem by considering this threshold [27, 28]. The augmented and encoded features combine together to form the final training data.

Multiple Model Creation

The processed data is segregated into 3 sections; training data, validation data and test data. The processing models are created in two levels. The first level is architected using two distinct machine learning models; a tree based prediction model, and an ensemble model. Decision tree is used as the tree based learning model, and random forest is used as the ensemble modeling technique.

Decision tree is a tree based learning algorithm. The model operates by determining the entropy levels and performing the tree split based on the feature with highest in entropy. The model is dynamic in nature and can effectively operate on huge data sets. Random forest is a bagging ensemble model that is based on combining a group of decision tree models. Multiple trees are used by the model. However, each tree is provided with a different subset of the training data. Hence, every tree generates different decision rules. A combination of these two models has been identified to effectively work together to provide a better performing model.

The entire training data is passed to both the models and the models are trained. Each model creates a different set of decision rules. The validation data is passed to both the models and the predictions are obtained. The two set of predictions are combined with the actual length of stay values to form the training data for the second level stacked model.

Final Prediction using Stacked Model

Predictions obtained from the first level model is used as a training data for the second level model. Logistic regression is used as the machine learning model of choice. The intentional use of linear model is to ensure that the model is free from unnecessary complexities. As it is used in the second stacking level, this model is called the meta-model. This model operates based on predictions from the previous level. The test data is initially predicted using the level one models. The obtained predictions are passed to the level 2 stacked meta-model for the final prediction.

Results and Discussion

The proposed Feature Augmented Stacked Bagging (FASB) model has been implemented using Python. The multiple record files obtained from MIMIC data are read using the pandas library. The files are aggregated and feature and feature augmentation is performed using the concatenation functions. The multilevel models are created using scikit library available in Python. Performance of the FASB model is measured using standard classifier performance metrics.

The ROC curve representing performance of the model in terms of true positive rate (TPR) and false positive rate (FPR) is shown in figure 1. High TPR levels and low FPR levels depict a good classifier model. The performance indicates TPR levels greater than 0.9 and FPR levels less than 0.2. High TPR levels indicate that the proposed model exhibits high accuracy in determining patients with high probability levels of staying for more than 10 days. Low FPR levels indicate that the model exhibits low false alarm levels, which means that the FASB model has low probability of misclassifying a patient with less than 10 days of stay.

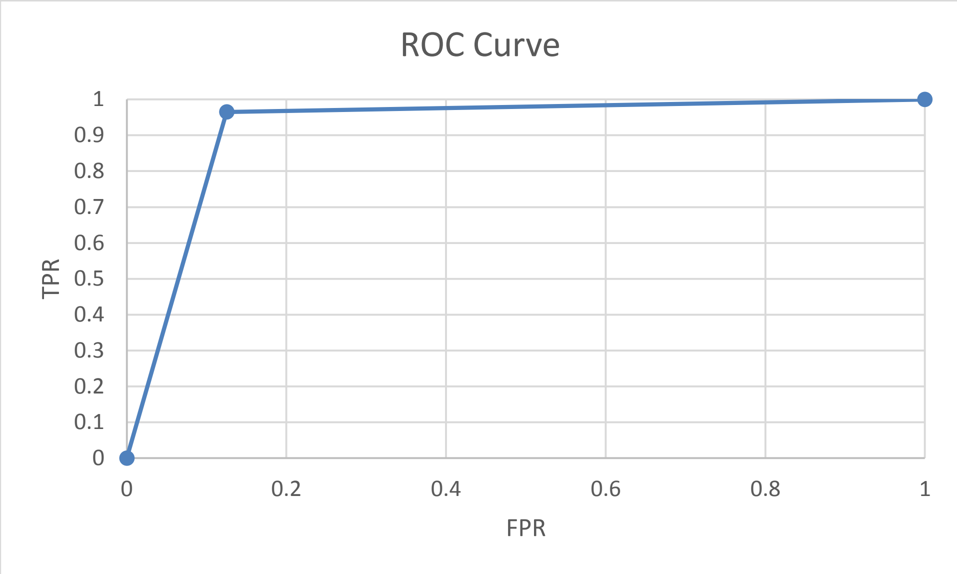

Figure 1:ROC Curve of FASB

The PR curve representing the precision and recall levels is shown in figure 2. Precision and recall represents the efficiency of a classifier in correctly identifying patients staying for more than 10 days. High values for precision and recall represent a good classifier. The precision and recall values for the FASB model has been identified to be greater than 0.9. Such high values indicate that the model is highly effective in identifying patients with high LoS values.

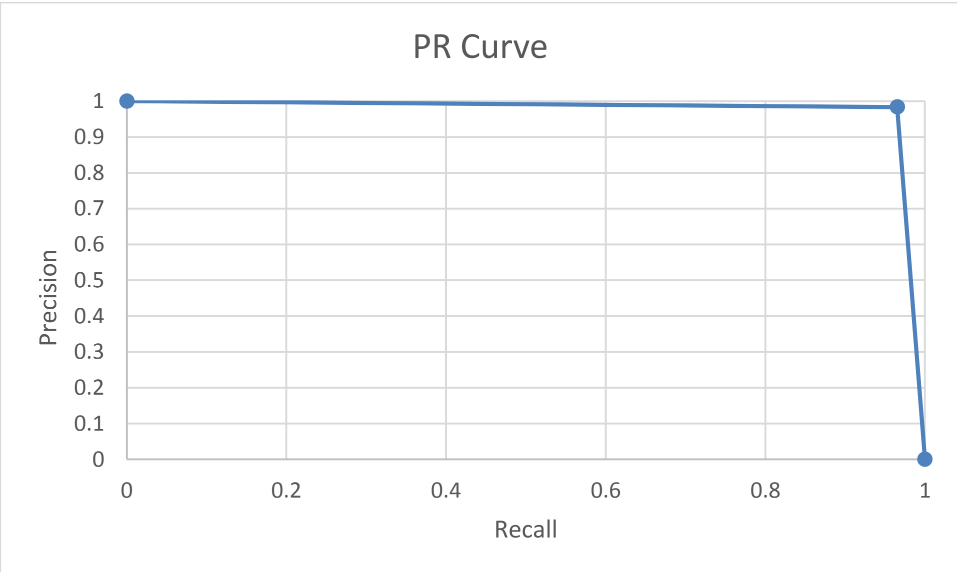

Figure 2: PR Curve of FASB

A tabulated view of the classification report is shown in table 1. The table distinctly represent precision, recall and F1 score in terms of both the classes.

The table indicates that the FASB model is highly accurate in identifying Class 1 which represents patients with high probability of staying for more than 10 days. However, the precision, recall and F1 score for class 0 is observed to be average, ranging between 77% and 87%. This shows that the model still has scope for improvement in determining patients who has high probability of leaving before 10 days. The overall accuracy level off the model is observed to be 95%, indicating that the overall predictions is highly effective.

Table 1: Classification Report of FASB

	Precision	Recall	F1-Score	Support
0	0.77	0.87	0.82	8979
1	0.98	0.97	0.97	68784
Accuracy			0.95	77763
Macro	0.87	0.92	0.9	77763
Avg	0.96	0.95	0.96	77763

The proposed FASB model has been compared with the model proposed by Ma et al [14]. A comparison of the aggregate metrics, accuracy, sensitivity, specificity, and precision is shown in figure 3. It could be determined from the figure that except for specificity, all the other metrics exhibit performance greater than 90%. The performance is also observed to be comparatively higher than the Ma et al. model.

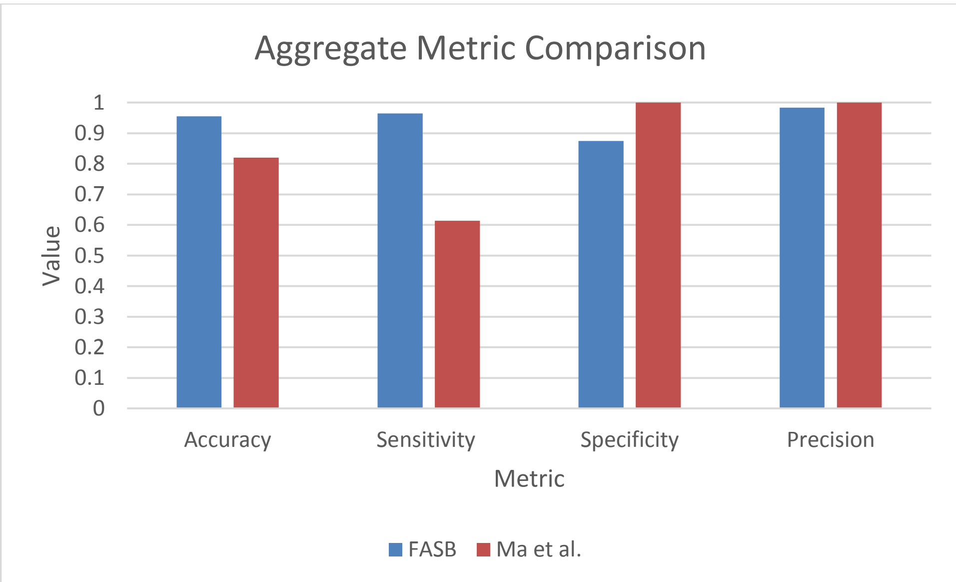

Figure 3: Comparison of Aggregate Metrics of FASB

A tabulated view of the performance is shown in table 2. The best performances are highlighted in bold. It could be observed that the FASB model exhibits 13% increased accuracy levels, 35% improved sensitivity levels, and 13% increased G-

Mean. Further, a 1.02 decrease in the Lift levels are also observed, exhibiting the improved performance of FASB.

Table 2: Performance Comparison of FASB

	FASB	Ma et al.
Accuracy	0.9547	0.82
Sensitivity	0.965	0.614
Specificity	0.8748	1
Precision	0.983	1
G-Mean	0.918	0.784
Lift	1.11	2.13

Conclusion

Length of stay prediction is a complicated process, mainly due to the large number of factors involved in the prediction process. Accurate prediction leads to several benefits both for the patient and the hospitals. This work presents a stacking ensemble based LOS prediction model, FASB. The work is composed of two levels to handle the complexities of the data. The initial level is composed of two machine learning models, while the second stacked level uses a linear model. Predictions from the initial levels is used to train the second level model. The second level meta model learns from the errors and the correct predictions of the first level model. Performance comparisons indicate that the proposed FASB model exhibits improved accuracy levels off up to 13% depicting that it provides better predictions. Limitations of the proposed model is that the FASB model exhibits reduced specificity and precision levels. Future enhancements will be based on improving the specificity and precision levels of the current model.

References

1. Alvin Rajkomar, Eyal Oren, Kai Chen, et al. 2018. Scalable and Accurate Deep Learning with Electronic Health Records. In *npj Digital Medicine*
2. B. Mozes, E. Schiff, B. Modan, Factors affecting inappropriate hospital stay, Int. J. Qual. Health Care 3 (3) (1991) 211–217.
3. H. Song, D. Rajan, J.J. Thiagarajan, A. Spanias, Attend and diagnose: Clinical time series analysis using attention models, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
4. Predicting in-hospital length of stay for very-low-birth-weight preterm infants using machine learning techniques
5. Improving clinical outcome predictions using convolution over medical entities with multimodal learning
6. R. Resar, K. Nolan, D. Kaczynski, K. Jensen, Using real-time demand capacity management to improve hospitalwide patient flow, Joint Commiss. J. Qual. Patient Saf. 37 (5) (2011), 217–AP3.
7. H. Harutyunyan, H. Khachatrian, D.C. Kale, G.V. Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, arXiv preprint arXiv: 1703.07771 (2017).

8. H. Song, D. Rajan, J.J. Thiagarajan, A. Spanias, Attend and diagnose: Clinical time series analysis using attention models, in: Thirty-Second AAAI Conference on Artificial Intelligence, 2018.
9. P. Gupta, P. Malhotra, L. Vig, G. Shroff, Using features from pre-trained timenet for clinical predictions, KHD@ IJCAI (2018) 38–44.
10. S. Purushotham, C. Meng, Z. Che, Y. Liu, Benchmarking deep learning models on large healthcare datasets, J. Biomed. Informat. 83 (2018) 112–134.
11. Temporal Pointwise Convolutional Networks for Length of Stay Prediction in the Intensive Care Unit
12. Ashish Vaswani, Noam Shazeer, NikiParmar, JakobUszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and IlliaPoLoSukhin. 2017. Attention is All You Need. In Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., 6000–6010.
13. Huan Song, DeeptaRajan, Jayaraman J. Thiagarajan, and Andreas Spanias. 2018. Attend and Diagnose: Clinical Time Series Analysis using Attention Models. In 32nd AAAI Conference on Artificial Intelligence, AAAI 2018. 4091–4098.
14. Length of stay prediction for ICU patients using individualized single classification algorithm
15. Neural network-based multi-task learning for inpatient flow classification and length of stay prediction
16. Brandon Thompson, Karim Elish, Robert Steele, Machine learning-based prediction of prolonged length of stay in newborns, in: 2018 17th IEEE International Conference on Machine Learning and Applications, ICMLA, IEEE, 2018, pp. 1454–1459.
17. Whitney E Muhlestein, Dallin S Akagi, Jason M Davies, Lola B Chambless, Predicting inpatient length of stay after brain tumor surgery: Developing machine learning ensembles to improve predictive performance, Neurosurgery 85 (3) (2019) 384–393.
18. Stefanie Jauk, Diether Kramer, Günther Stark, Karl Hasiba, Werner Leodolter, Stefan Schulz, Johann Kainz, Development of a machine learning model predicting an ICU admission for patients with elective surgery and its prospective validation in clinical practice., Stud. Health Technol. Inf. 264 (2019) 173–177.
19. A deep attention model to forecast the Length Of Stay and the in-hospital mortality right on admission from ICD codes and demographic data
20. T. Zebin, S. Rezvy, T.J. Chaussalet, A deep learning approach for length of stay prediction in clinical settings from medical records, in: 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), IEEE, 2019, pp. 1–5.
21. I.E. Livieris, T. Kotsilieris, I. DimopouLoS, P. Pintelas, Decision support software for forecasting patient's length of stay, Algorithms 11 (12) (2018) 199.
22. T. Gentimis, A. Ala'J, A. Durante, K. Cook, R. Steele, Predicting hospital length of stay using neural networks on mimic iii data, in: 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data

Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/ CyberSciTech), IEEE, 2017, pp. 1194–1201.
23. A.A. Kramer, J.E. Zimmerman, A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay, BMC Med. Informat. Decis. Making 10 (1) (2010) 27.
24. M. Sotoodeh, J.C. Ho, Improving length of stay prediction using a hidden markov model, AMIA Summits Translat. Sci. Proc. 2019 (2019) 425.
25. M.-T. Chuang, Y.-H. Hu, C.-F. Tsai, C.-L. Lo, W.-C. Lin, The identification of prolonged length of stay for surgery patients, in: 2015 IEEE International Conference on Systems, Man, and Cybernetics, IEEE, 2015, pp. 3000–3003.
26. All Data Inclusive, Deep Learning Models to Predict Critical Events in the Medical Information Mart for Intensive Care III Database (MIMIC III)
27. Jen J. Gong, Tristan Naumann, Peter Szolovits, and John V. Guttag. 2017. Predicting Clinical Outcomes Across Changing Electronic Health Record Systems. In *KDD*.
28. Bret Nestor, Matthew B. A. McDermott, Geeticka Chauhan, Tristan Naumann, Michael C. Hughes, Anna Goldenberg, and MarzyehGhassemi. 2018. Rethinking Clinical Prediction: Why Machine Learning must Consider Year of Care and Feature Aggregation. *CoRR*abs/1811.12583 (2018). arXiv:1811.12583