Predicting User-specific Future Activities using LSTM-based Multi-label Classification

Mohammad Sabik Irbaz[1], Fardin Ahsan Sakib[2], Lutfun Nahar Lota[1]
Department of Computer Science and Engineering, Islamic University of Technology, Bangladesh[1]
Department of Computer Science, George Mason University, USA[2]
sabikirbaz@iut-dhaka.edu, fsakib@gmu.edu, lota.nahar.407@gmail.com

Paper ID: 49

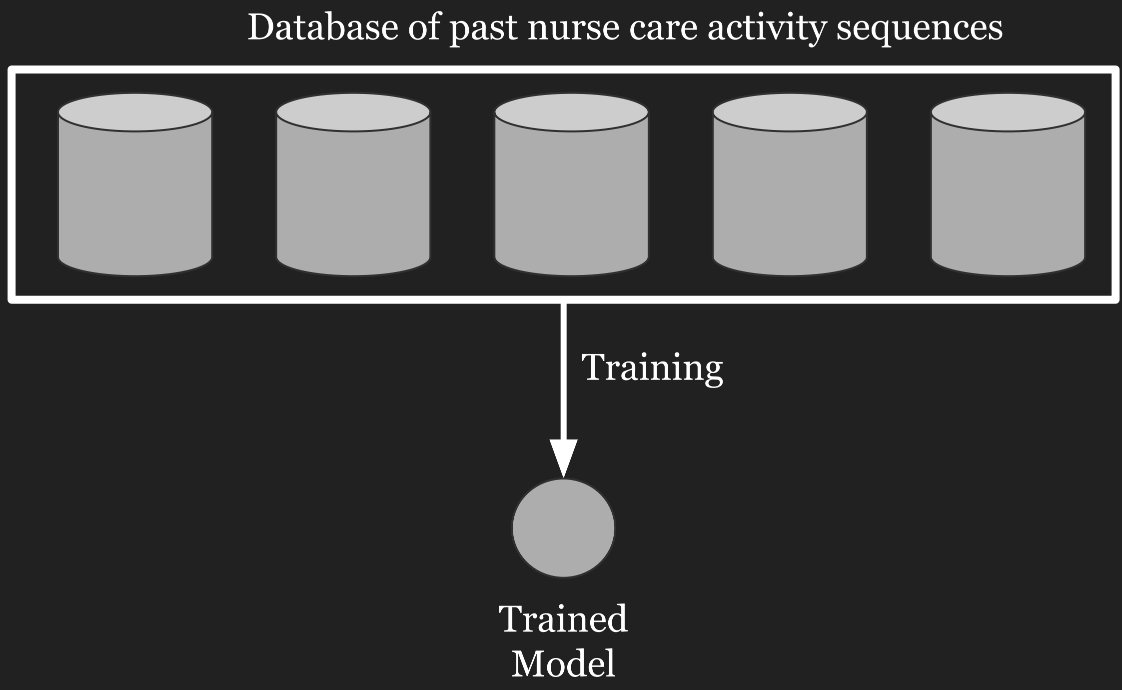
The 4th International Conference on Activity and Behavior Computing, 2022

London, United Kingdom (Hybrid)

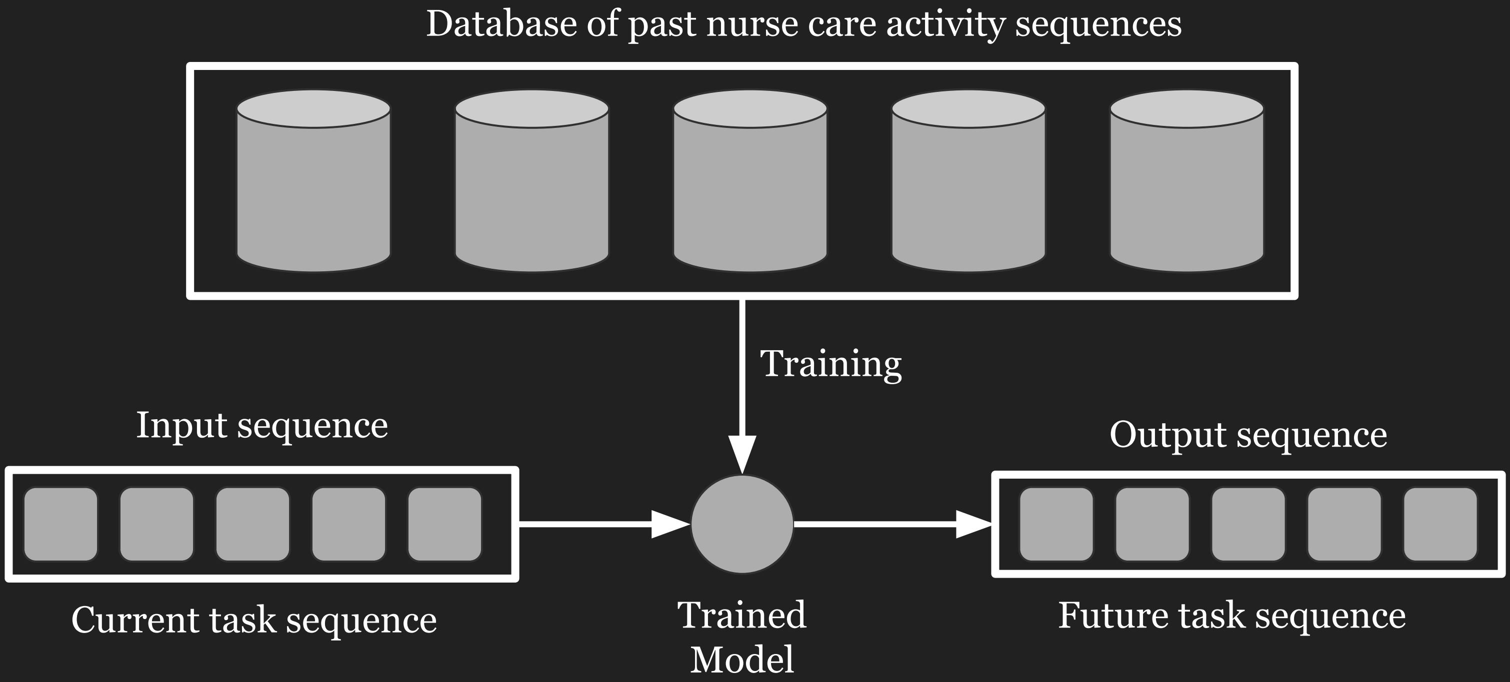
Introduction

- User-specific future activity prediction in the healthcare domain can drastically improve the efficiency and robustness of the services provided by the nurses.
- The fourth nurse care activity recognition challenge [1] aims to predict future activities of a specific nurse based on the nurse’s past activities and the sequence of activities
- In this paper, our team, **“Not A Fan of Local Minima”** , employ a novel 2-stage training approach (user-agnostic pre-training and user-specific fine-tuning) and achieve a test accuracy of **92%**.
- We concluded that proper data pre-processing and a **2-stage training process** resulted in better performance

Outline


Database of past nurse care activity sequences
Training
Trained
Model

Outline



Database of past nurse care activity sequences
Training
Input sequence
Output sequence
Current task sequence
Trained Model
Future task sequence

Problem Statement

Given a dataset of previous activity sequences of nurses, we need to predict the future task sequences of a particular nurse based on his/her current task sequences.

Research Challenges

Raw and skewed dataset

Research Challenges

Raw and skewed dataset

Inconsistency between train and test data

Research Challenges

Raw and skewed dataset

Inconsistency between train and test data

Processing the data for multi-label classification

Related Works

- Activity Recognition on health care data has not been studied extensively due to the complexity and limited availability of the dataset.

- First nurse care activity challenge collected a multi-modal dataset in a controlled environment [3]. Kadir et al. [2] used kNN classifier and achieved 87% accuracy on 10-fold cross-validation and 66% accuracy on leave-one-subject-out cross-validation.

- Second nurse care activity recognition challenge [4] used both lab and real-world data. Irbaz et. al [5] achieved a validation accuracy of 75% and test accuracy of 22.35%. They used both a high pass and a low pass filter to shape the data in the spatial domain during preprocessing and used the kNN classifier.

- Third care nurse care activity recognition challenge [6] focused on recognizing the activities of a nurse based on the accelerometer data as it is the cheapest and the most feasible way of collecting activity data. Sayem et al [7] achieved a validation accuracy of 72% using random forest.

Datasets

We used both of them because:

- their structures are almost similar
- their data collection process is same
- both of them contain total 28 labels or activity types performed by caregivers or nurses

3rd Nurse Care Data

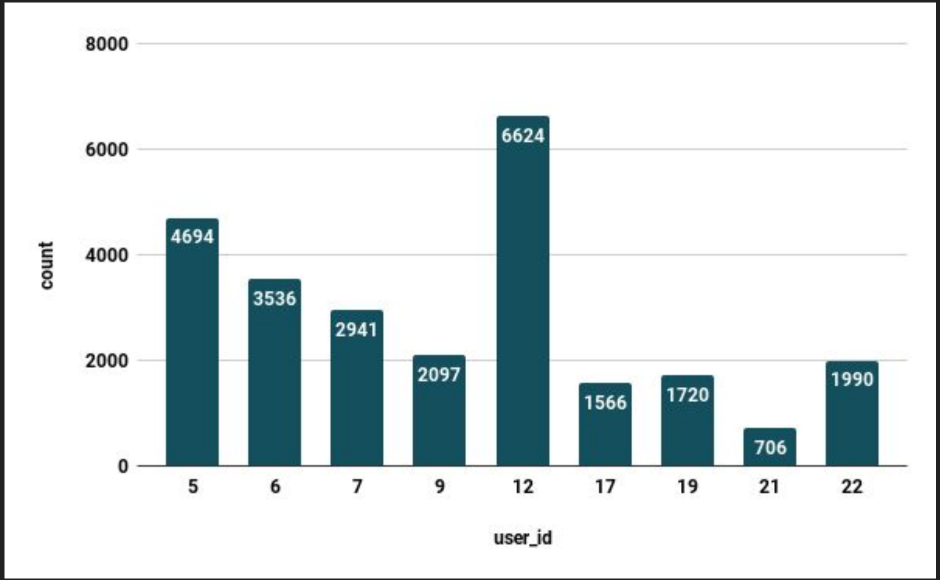
Size: **25,874**

Users: **9**

User IDs: ***5, 6, 7, 9, 12, 17, 19, 21, 22***

Max activity records: **6,224 (user 12)**

Min activity records: **706 (user 21)**



4th Nurse Care Data

Train Data Size: **10,985**

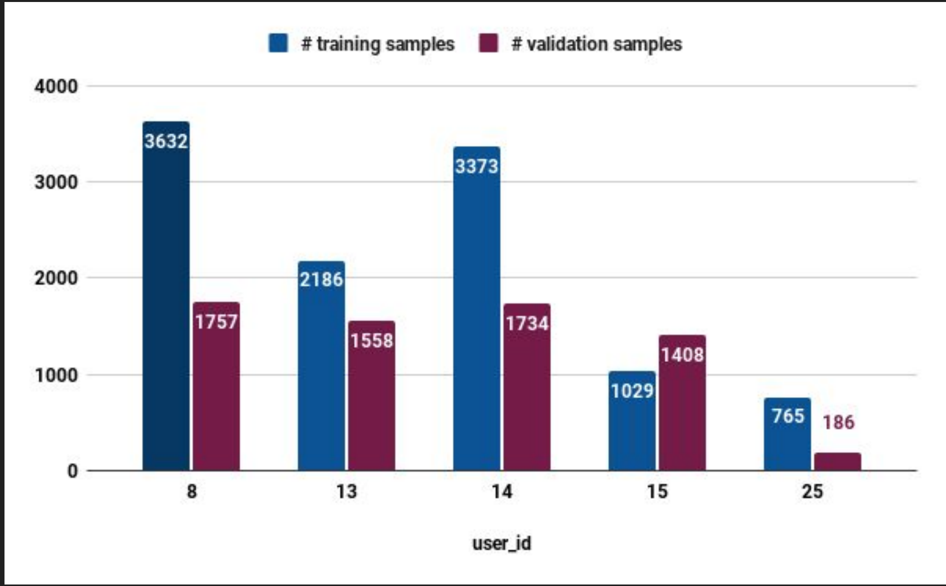
Validation Data Size: **6,643**

Users: **5**

User IDs: ***8, 13, 14, 15, 25***

Max activity records: **3,632 (user 8)**

Min activity records: **765 (user 25)**



Proposed Methology

Three Stages

Proposed Methology

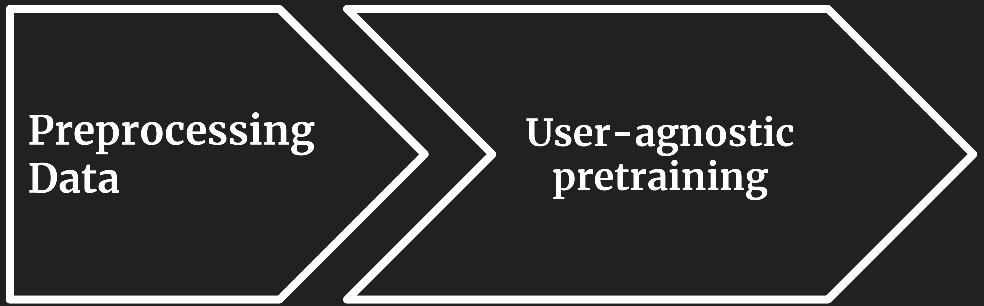
Three Stages

Preprocessing Data

Organizing the Raw Data for training and testing

Proposed Methology

Three Stages



Training without specifying
any particular user

Proposed Methology

Three Stages

Preprocessing Data

User-agnostic pretraining

User-specific finetuning

Finetuned on a specific user data

Stage 1: Preprocessing Data

- ➢ For each user and a particular date and hour, all the activities are listed sequentially based on the starting time of those activities. For a specific date and hour, the unique activities are also listed.
- ➢ If there is only one instance of a particular date, we remove that record.
- ➢ To build the final dataset, we gather the hourly tasks of a particular date and iterate through them.
- ➢ Finally after preprocessing, we get the following format

user_id	year-month-date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month-date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month -date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

[7,

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month -date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

[7, 85, frequency([10, 23, 6, 6, 6, 6]),

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month-date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

[7, 85, frequency([10, 23, 6, 6, 6, 6]), 87, presence([6, 10, 23]),

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month-date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

[7, 85, frequency([10, 23, 6, 6, 6, 6]), 87, presence([6, 10, 23]), 89

8, 85, frequency([6]), 87, presence([6]),

Stage 2: User-agnostic Pretraining

Reorganizing for the input sequence

user_id	year-month -date	previous hours	previous activities	previous activities unique	next_hour	next activities
25	2018-05-30	[7, 8]	[[10, 23, 6, 6, 6, 6], [6]]	[[6, 10, 23], [6]]	9	[10]
25	2018-05-26	[6, 7, 8]	[[23], [6, 10, 10, 6, 6, 6], [18, 6]]	[[23], [6, 10], [6, 18]]	9	[7, 18, 22]
25	2018-06-03	[7, 8, 10, 11]	[[6, 6, 6, 23, 10, 6], [6], [10, 24], [6, 6, 6]]	[[6, 10, 23], [6], [10, 24], [6]]	12	[6]

[7, 85, frequency([10, 23, 6, 6, 6, 6]), 87, presence([6, 10, 23]), 89

8, 85, frequency([6]), 87, presence([6]), 89, 9]

Stage 2: User-agnostic Pretraining (cont.)

Configurations:

Max Length: **1200**

LSTM: **1200 x 128**

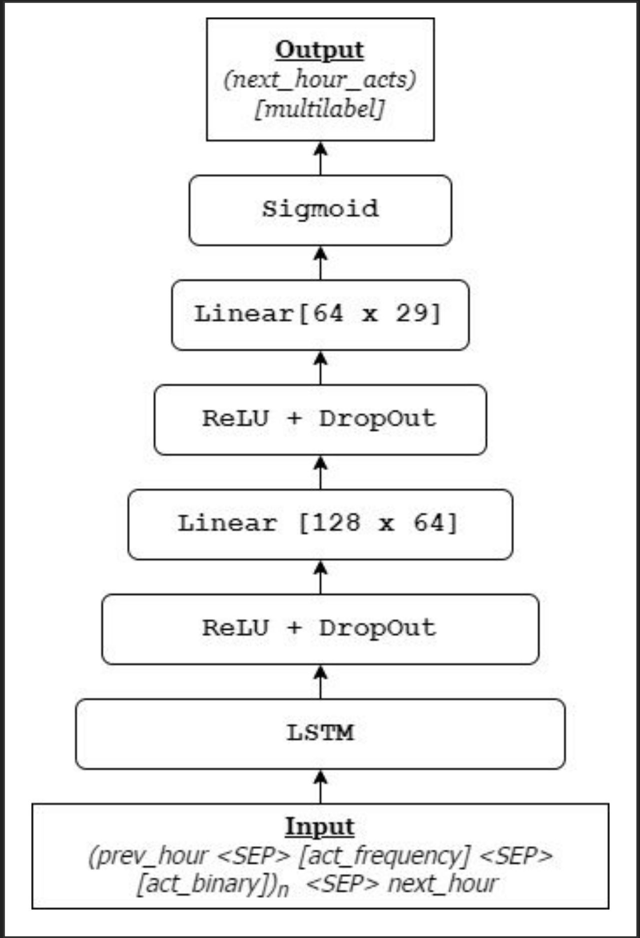
Dropout: **0.1**

Output Size: **29**

Learning Rate: 4e-4

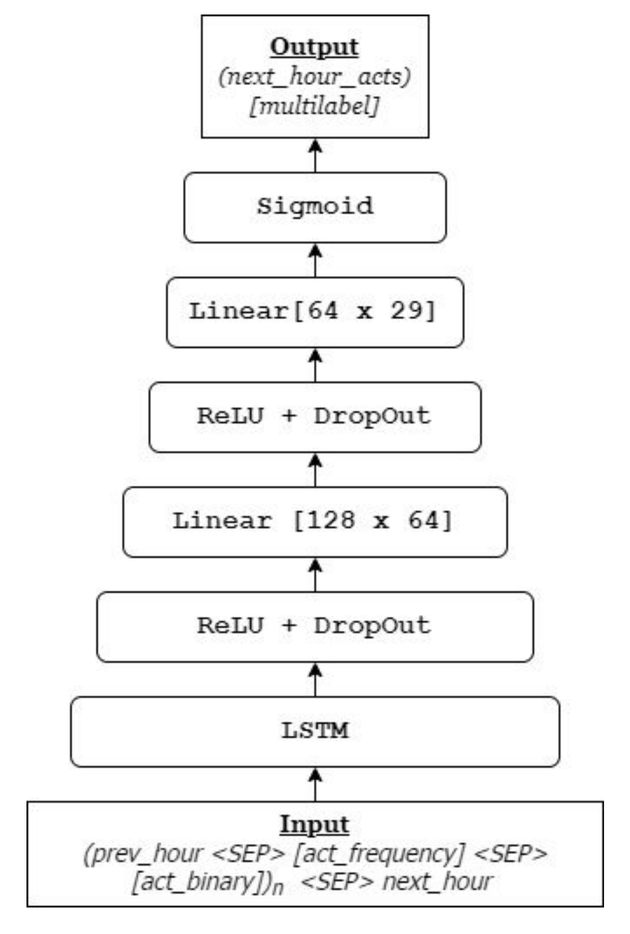
Epoch: 50

Threshold: 0.5



Stage 3: User-specific Finetuning

- ➔ Pretrained model weights are utilized and finetuned using the data of specific users using the same configurations and model architecture.
- ➔ These finetuned models are used for final inference on the test data.



Evaluation Metrics

- ➔ Four evaluation metrics:
  - **Accuracy (Exact Match):** Ratio of correct predictions to the total observations
  - **Precision:** Ratio of correct predictions to the total predicted positive observations
  - **Recall:** Ratio of correct predictions to the all observations in actual class
  - **F1-Score:** Weighted average of Precision and Recall
- ➔ **Accuracy** is the strictest metric for multilabel classification
- ➔ **F1-Score** provides a balanced score

User-agnostic pre-training performance

- We trained 2 different models with 2 different combination of datasets without specifying any user.
- We trained BiLSTM and LSTM models using only 4th Nurse Care dataset and, in other two cases, we trained BiLSTM and LSTM models combining 3rd and 4th Nurse Care dataset.

Model	Data
BiLSTM	4th Nurse Care
LSTM	4th Nurse Care
BiLSTM	3rd and 4th Nurse Care
LSTM	3rd and 4th Nurse Care

User-agnostic pre-training performance (cont.)

Model	Data	Accuracy	Precision	Recall	F1-score
BiLSTM	4th Nurse Care	0.2414	0.4848	**0.6781**	0.5327
LSTM	4th Nurse Care	0.2452	**0.5111**	0.6776	**0.5485**
BiLSTM	3rd and 4th Nurse Care	0.2414	0.4848	0.6781	0.5327
LSTM	3rd and 4th Nurse Care	**0.2548**	0.4908	0.6862	0.5453

- ➔ The evaluation scores show that we get the best F1-score with LSTM model pre-trained on 4th Nurse Care data and the best accuracy with LSTM model pre-trained on both 3rd and 4th Nurse Care data.
- ➔ We chose the BiLSTM and LSTM model pre-trained on both 3rd and 4th Nurse Care data for further finetuning since it reaches the best accuracy (exact match) result.

User-specific fine-tuning performance

- After finetuning the pretrained LSTM model on each specific user data, we averaged the scores and got the best accuracy **(31.58%)** and F1-score **(60.38%)**.

- We chose these finetuned LSTM models for final inference.

Model	user_id	#valid_samples	Accuracy	Precision	Recall	F1-score
LSTM	**8**	128	0.0703	0.5067	0.6353	0.5313
	13	74	0.3108	0.5532	0.7005	0.5958
	14	119	0.3333	0.6523	0.7582	0.6729
	15	139	0.5286	0.6472	0.7674	0.675
	25	66	0.3182	0.4758	0.5227	0.4786
	avg		**0.3158**	**0.5794**	**0.6931**	**0.6038**
BiLSTM	**8**	128	0.0469	0.4459	0.6505	0.4985
	13	74	0.3108	0.5491	0.7288	0.601
	14	119	0.2583	0.6768	0.7019	0.6683
	15	139	0.5143	0.6865	0.7618	0.6875
	25	66	0.303	0.4192	0.596	0.4658
	avg		**0.2875**	**0.5729**	**0.6957**	**0.5972**

Performance on Test Dataset

LSTM models pretrained on both 3rd and 4th Nurse Care Datasets and later finetuned on specific user data for inference.

Accuracy >> **92%**
F1-Score >> **19%**

Interesting Findings

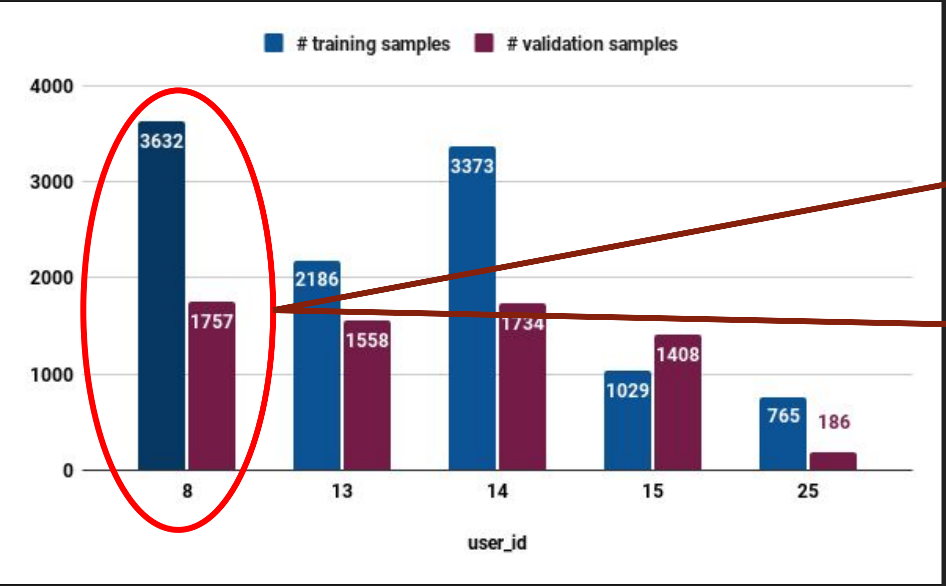
Adding 3rd Nurse Care Data with the 4th Nurse Care Data resulted in better performance in pretraining and finetuning. If we did not conduct the 2-stage training approach, our highest validation accuracy might have been around 24-25%. Employing this approach, resulted in 5-6% more accuracy and F1-score.

Model	Data	Accuracy	Precision	Recall	F1-score
BiLSTM	4th Nurse Care	0.2414	0.4848	**0.6781**	0.5327
LSTM	4th Nurse Care	0.2452	**0.5111**	0.6776	**0.5485**
BiLSTM	3rd and 4th Nurse Care	0.2414	0.4848	0.6781	0.5327
LSTM	3rd and 4th Nurse Care	**0.2548**	0.4908	0.6862	0.5453

Model	user_id	#valid_samples	Accuracy	Precision	Recall	F1-score
LSTM	**8**	128	0.0703	0.5067	0.6353	0.5313
	13	74	0.3108	0.5532	0.7005	0.5958
	14	119	0.3333	0.6523	0.7582	0.6729
	15	139	0.5286	0.6472	0.7674	0.675
	25	66	0.3182	0.4758	0.5227	0.4786
	avg		**0.3158**	**0.5794**	**0.6931**	**0.6038**
BiLSTM	**8**	128	0.0469	0.4459	0.6505	0.4985
	13	74	0.3108	0.5491	0.7288	0.601
	14	119	0.2583	0.6768	0.7019	0.6683
	15	139	0.5143	0.6865	0.7618	0.6875
	25	66	0.303	0.4192	0.596	0.4658
	avg		**0.2875**	**0.5729**	**0.6957**	**0.5972**

Interesting Findings (cont.)

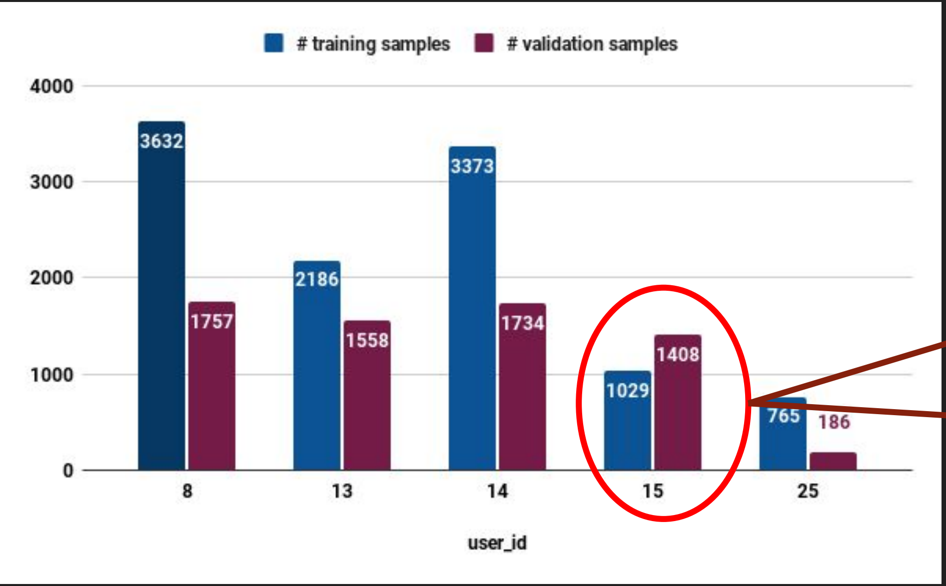
Even though user 8 had the highest records of training samples, it performed worst in terms of validation accuracy. This shows that the training and validation data is not consistent.



Model	user_id	#valid_samples	Accuracy	Precision	Recall	F1-score
LSTM	8	128	0.0703	0.5067	0.6353	0.5313
	13	74	0.3108	0.5532	0.7005	0.5958
	14	119	0.3333	0.6523	0.7582	0.6729
	15	139	0.5286	0.6472	0.7674	0.675
	25	66	0.3182	0.4758	0.5227	0.4786
	avg		**0.3158**	**0.5794**	**0.6931**	**0.6038**
BiLSTM	8	128	0.0469	0.4459	0.6505	0.4985
	13	74	0.3108	0.5491	0.7288	0.601
	14	119	0.2583	0.6768	0.7019	0.6683
	15	139	0.5143	0.6865	0.7618	0.6875
	25	66	0.303	0.4192	0.596	0.4658
	avg		**0.2875**	**0.5729**	**0.6957**	**0.5972**

Interesting Findings (cont.)

User 15 contains the second lowest number of training samples. But, we observed the best user-specific performance for it. It can be inferred that the overall dataset is very noisy because we can see the opposite of general deep learning model trends.



Model	user_id	#valid_samples	Accuracy	Precision	Recall	F1-score
LSTM	**8**	128	0.0703	0.5067	0.6353	0.5313
	13	74	0.3108	0.5532	0.7005	0.5958
	14	119	0.3333	0.6523	0.7582	0.6729
	15	139	0.5286	0.6472	0.7674	0.675
	25	66	0.3182	0.4758	0.5227	0.4786
	avg		**0.3158**	**0.5794**	**0.6931**	**0.6038**
BiLSTM	**8**	128	0.0469	0.4459	0.6505	0.4985
	13	74	0.3108	0.5491	0.7288	0.601
	14	119	0.2583	0.6768	0.7019	0.6683
	15	139	0.5143	0.6865	0.7618	0.6875
	25	66	0.303	0.4192	0.596	0.4658
	avg		**0.2875**	**0.5729**	**0.6957**	**0.5972**

Conclusion and Future Work

We employed an LSTM-based multi-label LSTM model and conduct a 2-stage training approach, i.e., user-agnostic pre-training and user-specific fine-tuning, to utilize both 3rd and 4th nurse care activity recognition challenge datasets and achieved 92% test accuracy.

Conclusion and Future Work

We employed an LSTM-based multi-label LSTM model and conduct a 2-stage training approach, i.e., user-agnostic pre-training and user-specific fine-tuning, to utilize both 3rd and 4th nurse care activity recognition challenge datasets and achieved 92% test accuracy.

The dataset, as it stands now, is very imbalanced. We plan to work more on this problem to handle the data imbalance problem for care record data

We would also like to try some more sequence models like transformer-based models to compare the results with LSTM and BiLSTM.

References

1. Sozo Inoue; Defry Hamdhana; Christina Garcia; Haru Kaneko; Nazmun Nahid; Tahera Hossain; Sayeda Shamma Alia; Paula Lago. 4th nurse care activity recognition challenge datasets, 2022.

2. Sayeda Shamma Alia, Kohei Adachi, Tahera Hossain, Nhat Tan Le, Haru Kaneko, Paula Lago, Tsuyoshi Okita, and Sozo Inoue. Summary of the third nurse care activity recognition challenge - can we do from the field data? In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, UbiComp'21, page 428–433, New York, NY, USA, 2021. Association for Computing Machinery.

3. Sozo Inoue, Paula Lago, Shingo Takeda, Alia Shamma, Farina Faiz, Nattaya Mairittha, and Tittaya Mairittha. Nurse care activity recognition challenge. IEEE Dataport, 1:4, 2019.

4. Sayeda Shamma Alia, Paula Lago, Kohei Adachi, Tahera Hossain, Hiroki Goto, Tsuyoshi Okita, and Sozo Inoue. Summary of the 2nd nurse care activity recognition challenge using lab and field data. In Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, UbiComp-ISWC '20, page 378–383, New York, NY, USA, 2020. Association for Computing Machinery.

5. Mohammad Sabik Irbaz, Abir Azad, Tanjila Alam Sathi, and Lutfun Nahar Lota. Nurse care activity recognition based on machine learning techniques using accelerometer data. In Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers, pages 402–407, 2020.

6. Sozo Inoue, Paula Lago, Tahera Hossain, Tittaya Mairittha, and Nattaya Mairittha. Integrating activity recognition and nursing care records: The system, deployment, and a verification study. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol., 3(3), sep 2019.

7. Faizul Rakib Sayem, MD Mamun Sheikh, and Md Atiqur Rahman Ahad. Feature-based method for nurse care complex activity recognition from accelerometer sensor. In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers, UbiComp '21, page 446–451, New York, NY, USA, 2021. Association for Computing Machinery.