Applied Machine Learning for Digital Health
Nurse Care Activity Recognition Sensor Comparison with CNN

Samik Real
Data Engineering
Hasso Plattner Institut
Prof.-Dr.-Helmert-Straße 2-3
14482 Potsdam, Germany
Email: samik.real@student.hpi.de

Orhan Konak
Digital Health
Hasso Plattner Institut
Prof.-Dr.-Helmert-Straße 2-3
14482 Potsdam, Germany
Email: orhan.konak@hpi.de

***Abstract*—Physical activity recognition has been studied vastly, while Nurse Care Activity Recognition is a field with scarce research. The reason is due to the complexity of the compounded tasks involved where the activity involves the subject and the receiver, unlike Human Activity Recognition that only involves the subject. Another reason for the low popularity is the difficulty and scarcity of public datasets available. Nurse Care Activity Recognition is such an important field of study because it allows nurses to alocate more time to patient care instead of activity logging. Current research in this field uses traditional machine learning techniques and different sensor types. This paper aims to asses the importance of different sensor types and promote the use of more complex deep learning models. It also achieved comparable results to the 2019 Nurse Care Activity Recognition Challenge without using the sensor with the most information. This was achieved by improving the CNN architecture and obtaining interesting and counter-logic results regarding correlation between adjacent sensor readings.**

1. Introduction

With the increase of technology available to the public such as smartphones and wearable electronics, *ubiquitous sensing* and related fields have gained popularity. One of these fields with an increase in demand is Human Activity Recognition (HAR), which is the automatic identification of physical activities performed by humans [1] [2]. There are two fields of HAR : Vision Based HAR, which uses microphones and cameras for body movement recording; and Sensor Based HAR, which uses sensors to infer body movement [3]. This paper focuses on Sensor Based HAR, specifically on *Nurse Care Activity Recognition* and the influence of different sensor modalities.

Nurses need to document all activities for patient treatment registry and legal reasons, taking up time that could be spent on caring for patients [4]. Applying HAR to Nurse Care Activity Recognition aims to document nurse activities automatically and allow nurses to allocate more time to patient care.

Nurse Care Activity Recognition consists of identifying physical activities, similarly to HAR, but with more complex and compound tasks since the activities are performed to patients, thus making intra-class variability dependant on subject and patient receiving a treatment [5].

Identifying nurse care activities and human physical activities with the use of sensors poses a typical time-series classification problem, where the sensors are one dimensional and data points obtained from these sensor readings are highly-correlated with neighbouring data points [6]. Feature engineering is extremely important for the success of classification problems and usually requires expert knowledge to construct them adequately, which is why automated feature extraction is so appealing for this kind of tasks [7].

Convolutional Neural Networks (CNN or 2-D CNN) gained popularity in the computer vision field thanks to the outperformance of CNN models in comparison to previous models, making them state of the art for computer vision tasks since 2012 [8]. The reason for which CNNs outperform other Neural Network models is because CNNs share weights in each layer, making it harder for the model to overfit on training samples and taking into account spatial neighbour points in data samples, unlike Multi-layer-perceptrons. An added benefit of CNNs are their capability to extract complex features while training which proved to be of great success in image classification [9]. But the benefits from CNNs are not limited to the computer vision and image classification field. With the introduction of 1-D CNNs in 2015 [10], they have become state of the art in time series classification tasks as well [11]–[13] where they shine for the same reasons 2-D CNNs had success.

The contribute of this paper to the scarce Nurse Care Activity Recognition research field is two-fold. Firstly it aims to assess the importance of different sensor modalities on a public nurse activity dataset [14]. And secondly it promotes the usage of more complex deep learning models in this field, such as different CNN architectures, by showing comparable results.

The next sections of this paper are organized as follows: section 2 contextualizes this work in relation to previous research in Nurse Care Activity Recognition. Section 3 describes the experiments utilized to assess the importance of different sensor modalities. Section 4 showcases the results obtained from the experiments and addresses the limitations of this research. Section 5 analyses and explains the findings and in section 6 conclusions are established as well as presenting some future work.

2. Related Work

Previous work on Nurse Care Activity Recognition is scarce due to the increased complexity and the compound nature of the tasks, as well as the increased difficulty of acquiring data due to privacy concerns [5]. Nonetheless there has been a few early research papers on this topic with promising results.

There was a work published in 2006 where a Bluetooth accelerometer sensor, *B-pack*, was proposed to recollect data from nurses and for classification purposes. These sensors were placed on the chest, waist, left and right arms [15]. The authors validated their data recollection with an actual nurse who performed the 16 different tasks. It is not clear how many data samples were recollected for each task. Different features such as mean and maximum or minimum values were extracted. The classification accuracy using one nearest neighbour algorithm was 0.9680. This work showed the possibility of classifying tasks performed by care professionals with data from accelerometers, although it lacked inter-subject testing.

Abbate and colleagues [16] proposed in 2012 a fall detection system using accelerometer data from smartphones. The work was validated with 86 samples from seven subjects, obtaining an accuracy of 0.8140 with a neural network. Another similar work is presented in 2015 by De [17] where he proposes a similar method to Abbate but for general activities of daily living. Both of these works were focused toward patients with Alzheimer's and dementia and showed that activity recognition applied to medical health care is possible.

In 2015 Inoue submitted a paper where nurse activities were classified in a real hospital environment [18]. A total of 5700 activity instances were recollected from 22 nurses performing 25 different tasks. Accelerometer sensors were attached to the nurses' wrist, back and front chest pocket. Posterior probabilities were extracted from time steps in the data. A Bayesian network was used as the classifier and obtained an accuracy of 0.8096. This work showcased nurse care activity classification in a functioning hospital with accelerometer data.

In 2019 a global Nurse Care Recognition Challenge was held [5] to promote research in this field. This challenge consisted of classifying six different activities that were performed by eight subjects. The winner of this challenge used a K-NN classifier with an accuracy of 0.8020 [19].

Research in this field is scarce and machine learning approaches tend to be traditional, which is why the present paper extends the Nurse Care Recognition Challenge from 2019 by proposing a Convolutional Neural Network architecture and assessing the importance of different sensors used in the challenge.

3. Methods

This section describes the experiment, the dataset used, as well as the data preprocessing steps and the model selection.

3.1. Experiment Description

This paper builds on top of the 2019 Nurse Care Recognition Challenge [5] to assess the importance of the three different sensors used: motion capture, accelerometer and meditag sensors. Due to the inconvenient and restricting sensor location on the body necessary for the motion capture sensor, it was not used in this experiment. Omitting the motion capture sensor enables the experiment to be translated into a real hospital environment. Each classifier model proposed is trained on data from three modalities: accelerometer, meditag and accelerometer + meditag. The results of this experiment is later compared to the results from the challenge.

3.2. Dataset

The dataset used in the 2019 Nurse Care Challenge [14] and for this paper consists of nurse activity data from three sensors: accelerometer, meditag and motion capture sensors. The accelerometer sensor is from a smartphone located in the front pocket of the nurse, measuring acceleration in three dimensions with a sampling rate of 4Hz. The meditag sensor is a Bluetooth localization sensor meassuring the position x and y in meters and the air pressure in mHg with a sampling rate of 20Hz. The motion capture sensor measures the position of 29 markers in the body in 3 dimensions with a sampling rate of 100Hz for each marker. The data was recollected in a controlled and simulated environment in Kyushu Institute of Technology, Japan. Eight different nurses perform 6 activites. These activities are coded with the following number IDs, Vital signs measurements: 0, Blood collection: 1, Blood Glucose Measurement: 2, Indwelling drip retention and connection: 3, Oral care: 4 and Diaper exchange and cleaning of area: 5. Some sensor readings were noted to be inconsistent. The data was already divided into Train and Test sets by the Challenge authorities.

A slightly modified dataset was used for this experiment, removing samples with windows smaller or greater than 1 minute. Only five samples were removed with this criteria resulting in 189 Train samples and 80 Test samples. These samples were not subdivided into Train Validation and Test due to the low amount of instances. The distribution of the samples for Train and Test can be seen in figure 1.

3.3. Data Preprocessing

Data sample sensor vectors are first downsampled with linear interpolation to 4 Hz sampling rate each similar to

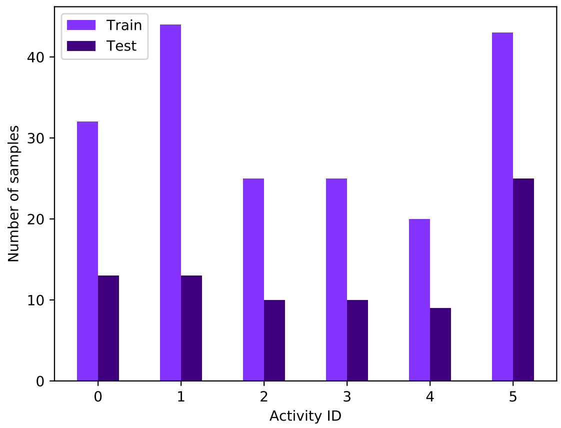

Figure 1. Number of samples of each activity class for Train and Test sets.

[19]. It was also normalized to have zero mean and unit variance. The data was first divided into small window sizes of 2 seconds with different overlap percentage as literature suggests with HAR and time series data [20], but it was empirically discovered that using 1 minute long windows and no overlap yielded the best results for this dataset.

3.4. Model Selection

With the increase in popularity of HAR, many researchers have started using more complex models instead of the traditional and less complex models of machine learning such as KNN and Bayesian Netowrks. The use of CNN models in time series data has shown promising results to the point where it has become state of the art in some fields such as EEG data classification [3], [11]–[13]. But this is not the case for Nurse Care Activity Recognition where traditional and less complex machine learning models are being used [15]–[18]. This can be due to the complexity of the compound tasks found in nurse care, making automatic feature extraction with Convolutional Neural Networks more challenging.

Apart from the contribution of assessing the importance of different sensors, this paper also aims to promote the usage of more complex deep learning models in this field and show comparable results. For this reason the CNN classifier type was chosen, as well as a KNN classifier for comparison to a traditionally simple machine learning algorithm. A KNN classifier also achieved the best performance in the 2019 Nurse Care Challenge [5], [19]. To be able to generalize the findings of the importance of different sensor modalities, different CNN architectures were used for the experiment. All CNN models were trained for 300 epochs with Train and Test datasets, since cross validation yielded low results, as well as to abide by the Challenge rules. Stochastic gradient descent was used with different learning rates, seen in table 1.

3.4.1. Sequential CNN. This architecture has a sequential form where the input of each layer corresponds to the output of the previous layer and is compatible with all three modalities. This means that it has 1 input of 6 vectors, in the case of using it with accelerometer + meditag modality, or 1 input of 3 vectors when using it with only 1 sensor modality. Features from each sensor are extracted together since the convolution is applied to all data vectors jointly. The Sequential CNN has a total of 5 convolutional layers and one fully connected layer of 100 neurons at the end. Each convolutional layer has a kernel size of 3. Batch normalization and PReLu activation function is applied after each layer. A visualization of the architecture can be seen in figure 3.

3.4.2. Two-headed CNN. This architecture is only compatible with accelerometer + meditag modality and the idea is to extract specific features from the accelerometer and meditag sensor individually instead of extracting features with both sensors together as in 3.4.1. It has 2 inputs with 3 vectors in each input. After the first convolution in each head, a dropout and a maxpooling layer is added which is then concatenated together with the other input head. After this the architecture is exactly the same as in 3.4.1.

3.4.3. Three-headed CNN. This architecture has three inputs with a single vector in each head and is only compatible with accelerometer and meditag modalities. The idea of this architecture is to extract individual features from each vector. The hypothesis of this architecture is that there is no added information or correlation between each vector of a sensor. The layer setup is similar to 3.4.2 but with 3 inputs of 1 vector instead. The architecture can be seen in figure 5.

3.4.4. Six-headed CNN. This architecture is similar to 3.4.3, where features are extracted from each single vector individually but only compatible with accelerometer + meditag modality. This architecture is smaller as well with a total of 3 convolutional layers and 1 fully connected layer. It can be visualized in figure 6.

4. Results

A summary of the results of training different model architectures on different sensor modalities can be seen in table 1. The different KNN architectures were trained on raw data, unlike [19] where features were constructed and an ensemble model of multiple KNNs was created. The best architecture for the accelerometer modality and for the meditag modality is the Three-headed CNN with a Test accuracy of 0.40 and 0.2875 respectively. The best model for the accelerometer + meditag modality is the Two-headed CNN with a Test accuracy of 0.4625, which is the best accuracy obtained with any modality.

The accuracy and loss history from the Two-headed CNN model can be seen in figures 7 and 8 respectively. Both graphs show that improvement throughout the

TABLE 1. ACCURACY RESULTS OF DIFFERENT CLASSIFICATION MODELS TRAINED ON DIFFERENT SENSOR MODALITIES

Model Architecture	Learning Rate	Sensor Modality	Training Accuracy	Test Accuracy
Sequential CNN	0.066	Accelerometer	0.3519	0.3775
Three-headed CNN	0.066	Accelerometer	0.3862	0.40
3-NN	NA	Accelerometer	0.6614	0.3125
Sequential CNN	0.050	Meditag	0.6481	0.2483
Three-headed CNN	0.050	Meditag	0.6508	0.2875
1-NN	NA	Meditag	1.0	0.1875
Sequential CNN	0.065	Accelerometer + Meditag	0.4113	0.45
Two-headed CNN	0.060	Accelerometer + Meditag	0.4127	0.4625
Six-headed CNN	0.065	Accelerometer + Meditag	0.4127	0.45
1-NN	NA	Accelerometer + Meditag	1.0	0.2625

300 epochs is small or nonexistent. Figure 2 shows the accuracy per class of this model, where it can be seen that classification for Activity class ID 3 and 4 is zero while the accuracy for activity 5 is 1.0.

TABLE 2. RESULTS FROM THE 2019 NURSE CARE RECOGNITION CHALLENGE [5]

Model	Sensor Modality	Training Acc	Test Acc
KNN	MoCap + Meditag	0.6611	0.8020
Spatio-temp Graph	MoCap	0.5700	0.6460
CNN	All 3 modalities	1.0	0.4650
Random Forest	Accelerometer	0.6000	0.4310

The best model from this experiment in table 1 is the Two-headed CNN with Test accuracy of 0.4625, which can be compared to the CNN from the 2019 Challenge in table 2 with Test accuracy of 0.4650 [5]. But the CNN model from table 2 uses all sensor modalities, while the Two-headed CNN uses accelerometer + medicap modalities.

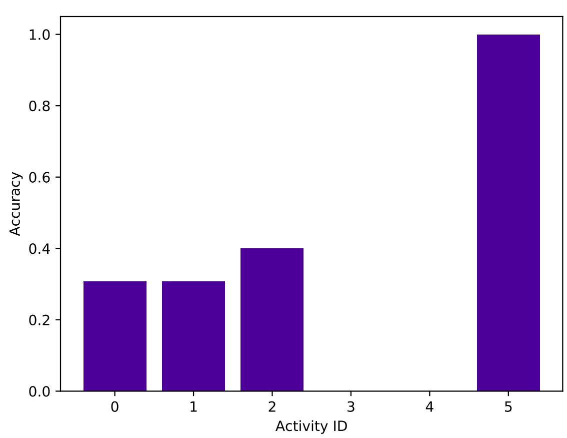

Figure 2. Accuracy by activity class for the Two-headed CNN model trained on accelerometer + meditag sensor modality. Activity ID encodings are Vital signs measurements: 0, Blood collection: 1, Blood Glucose Measurement: 2, Indwelling drip retention and connection: 3, Oral care: 4 and Diaper exchange and cleaning of area: 5. The accuracy of this model is 0.4625.

5. Discussion

The experiment used Train/Test splits instead of cross validation due to low Test accuracy, which shows that the model can not capture the underlying characteristics from the Train set to generalize to the Test set. A possible explanation for obtaining low accuracy when using cross validation is that all activities from one particular class go to the Test split in most splits or key samples that are included in the original Train set are not present in the cross validated Train samples and therefore the model can not learn from these important data samples.

On table 1 it can be seen that the modality with the best Test accuracy is the accelerometer + meditag modality, being the accelerator the second best and the meditag modality the last. This means that the sensor with the most information is the accelerometer, but the meditag sensor also adds information since the accuracy improves when adding both sensors together.

From the different model architectures described in 3.4, all architectures that divided the feature extraction in some way outperformed the Sequential CNN that grouped all data vectors together. When using only one sensor, it was best to separate all three dimensions so features were extracted individually. But when using two sensors it was best to separate the sensors but group the dimensions from each sensor ending up with 2 inputs of 3 vectors each as seen with the Two-headed CNN. This means that the intra-sensor information when using only one sensor adds noise and that there is no vector correlation; but when using 2 sensors the inter-sensor information obtained from features is better than from intra-sensor information. In other words it is best to separate all vectors when using one sensor and to separate sensors but join vectors from the sensor when using 2 sensors. This might be counter-intuitive since a naive thinking might be that sensors are correlated and that more meaningful information is obtained when features are extracted from data together. But this is not the case as stated previously.

The difficulty for the model to overfit shown in figure 7 and 8 shows that the model can not extract meaningful information. This could be due to the usage of a simple model, but this theory was discarded when more complex models were tested. This means that the raw data from the sensors might not provide enough for the CNN classifier. This is corroborated with the results from figure 2 where some classes were not learned at all. Further testing with time-series transformations, such as Fourier transforms, might be beneficial and necessary to increase the model's performance. [21].

The benefit of discarding the motion capture sensor is that it makes the experiment easily translatable to a real hospital environment since the motion capture sensors difficult nurses to perform their job because of the number of sensors necessary and their location on the body. This comes with the cost of reduced accuracy, since the best performing models from the 2019 Challenge use the motion capture modality as seen in table 2. But the results from the experiment in this paper also shows that a good CNN architecture can compensate the lack of the motion capture sensor and obtain comparable results as seen with the Two-headed CNN from this paper and the CNN model from the 2019 Challenge in table 2.

6. Conclusion

In this paper various CNN architectures were trained on different sensor modalities and their importance assessed. The results of the experiment from this paper shows that data vectors do not necessarily have meaningful information between them, contrary to what one might think. This paper also showed that it is possible to apply CNNs to Nurse Care Activity Recognition with non hindering sensors, given that a good architecture is chosen. Although some improvements might be made when applying time-series processing techniques such as Fourier transforms.

Future works will include adding motion capture modality to the experiment to further assess the importance of this sensor, add more classification models such as ensemble methods increase the performance of complex classifiers and allow them to compete against the simple machine learning approaches being used and applying time-series transformation techniques to the raw data such as Fourier transforms and other types of feature extraction techniques to allow models to extract more information and patterns necessary to improve model performance, which would promote the use of deep learning models which have become state of the art in so many fields..

References

[1] O. D. Lara and M. A. Labrador, "A survey on human activity recognition using wearable sensors," *IEEE communications surveys & tutorials*, vol. 15, no. 3, pp. 1192–1209, 2012.

[2] C. A. Ronao and S.-B. Cho, "Human activity recognition with smartphone sensors using deep learning neural networks," *Expert systems with applications*, vol. 59, pp. 235–244, 2016.

[3] A. Bevilacqua, K. MacDonald, A. Rangarej, V. Widjaya, B. Caulfield, and T. Kechadi, "Human activity recognition with convolutional neural networks," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2018, pp. 541–552.

[4] M. T. Sum and M. A. Chebor, "Documentation: historical perspectives, purposes, benefits and challenges as faced by nurses," *Int J Hum Soc Sci*, vol. 3, no. 16, pp. 236–40, 2013.

[5] P. Lago, S. S. Alia, S. Takeda, T. Mairittha, N. Mairittha, F. Faiz, Y. Nishimura, K. Adachi, T. Okita, F. Charpillet *et al.*, "Nurse care activity recognition challenge: summary and results," in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 746–751.

[6] Y. LeCun, Y. Bengio *et al.*, "Convolutional networks for images, speech, and time series," *The handbook of brain theory and neural networks*, vol. 3361, no. 10, p. 1995, 1995.

[7] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[9] D. C. Ciresan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Convolutional neural network committees for handwritten character classification," in *2011 International Conference on Document Analysis and Recognition*. IEEE, 2011, pp. 1135–1139.

[10] S. Kiranyaz, T. Ince, R. Hamila, and M. Gabbouj, "Convolutional neural networks for patient-specific ecg classification," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 2608–2611.

[11] C. Jobanputra, J. Bavishi, and N. Doshi, "Human activity recognition: A survey," *Procedia Computer Science*, vol. 155, pp. 698 – 703, 2019, the 16th International Conference on mobile systems and pervasive computing (MobiSPC 2019),The 14th International Conference on future networks and communications (FNC-2019), the 9th International Conference on sustainable energy information technology. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1877050919310166

[12] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (eeg) classification tasks: a review," *Journal of neural engineering*, vol. 16, no. 3, p. 031001, 2019.

[13] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

[14] "Nurse care activity recognition challenge," 2019, last accessed: Aug. 16, 2020. [Online]. Available: http://dx.doi.org/10.21227/2cvj-bs21

[15] R. Ohmura, F. Naya, H. Noma, and K. Kogure, "B-pack: a bluetooth-based wearable sensing device for nursing activity recognition," in *2006 1st international symposium on wireless pervasive computing*. IEEE, 2006, pp. 6–pp.

[16] S. Abbate, M. Avvenuti, F. Bonatesta, G. Cola, P. Corsini, and A. Vecchio, "A smartphone-based fall detection system," *Pervasive and Mobile Computing*, vol. 8, no. 6, pp. 883–899, 2012.

[17] D. De, P. Bharti, S. K. Das, and S. Chellappan, "Multimodal wearable sensing for fine-grained activity recognition in healthcare," *IEEE Internet Computing*, vol. 19, no. 5, pp. 26–35, 2015.

[18] S. Inoue, N. Ueda, Y. Nohara, and N. Nakashima, "Mobile activity recognition for a whole day: recognizing real nursing activities with big dataset," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 1269–1280.

[19] M. E. Kadir, P. S. Akash, S. Sharmin, A. A. Ali, and M. Shoyaib, "Can a simple approach identify complex nurse care activity?" in *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers*, 2019, pp. 736–740.

[20] O. Banos, J.-M. Galvez, M. Damas, H. Pomares, and I. Rojas, "Window size impact in human activity recognition," *Sensors*, vol. 14, no. 4, pp. 6474–6499, 2014.

[21] D. R. Brillinger, *Time series: data analysis and theory*. SIAM, 2001.

Appendix

The architectures of the CNN models used in this paper can be found in this section. Figure 3 corresponds to the sequential CNN architecture. This architecture was trained on all sensor modalities: accelerometer, meditag and the combination of both sensors. Figure 4 illustrates the two-headed CNN architecture trained on only one modality: accelerometer + meditag. Figure 5 depicts the Three-headed CNN architecture trained on accelerometer modality and meditag modality. And figure 6 corresponds to the six-headed CNN model trained on the accelerometer + meditag modality. Accuracy and loss history from the Two-headed CNN can be seen in figures 7 and 8 respectively.

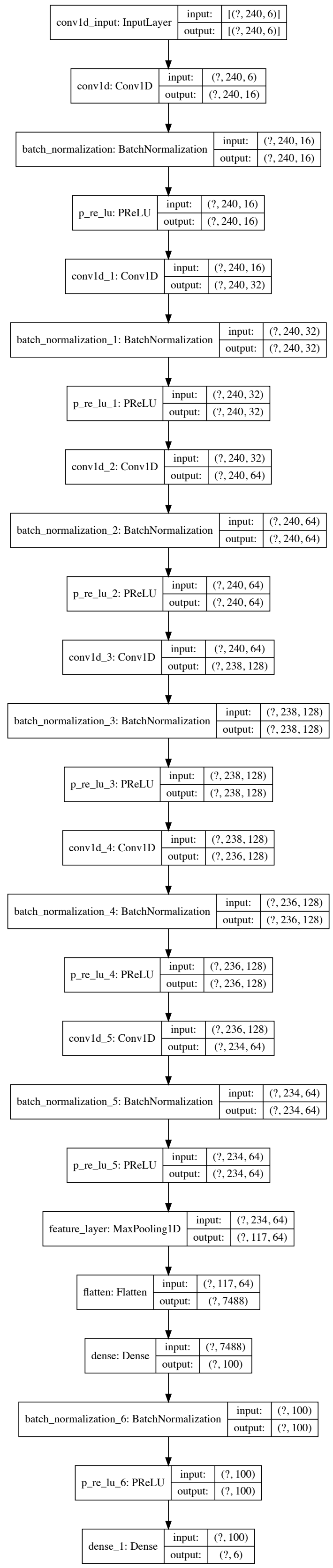

Figure 3. Sequential CNN model architecture with one input

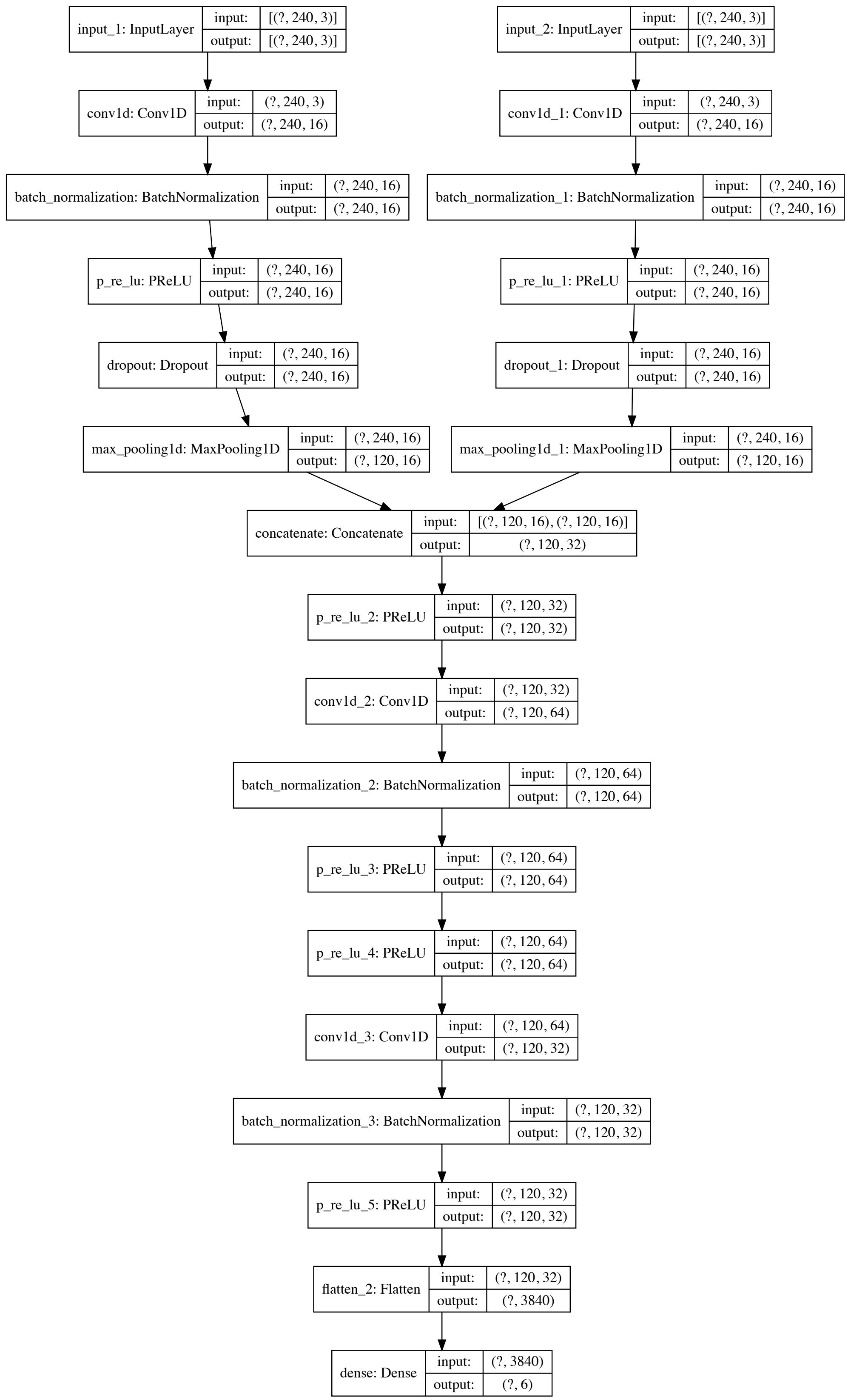

Figure 4. Two-headed CNN model architecture

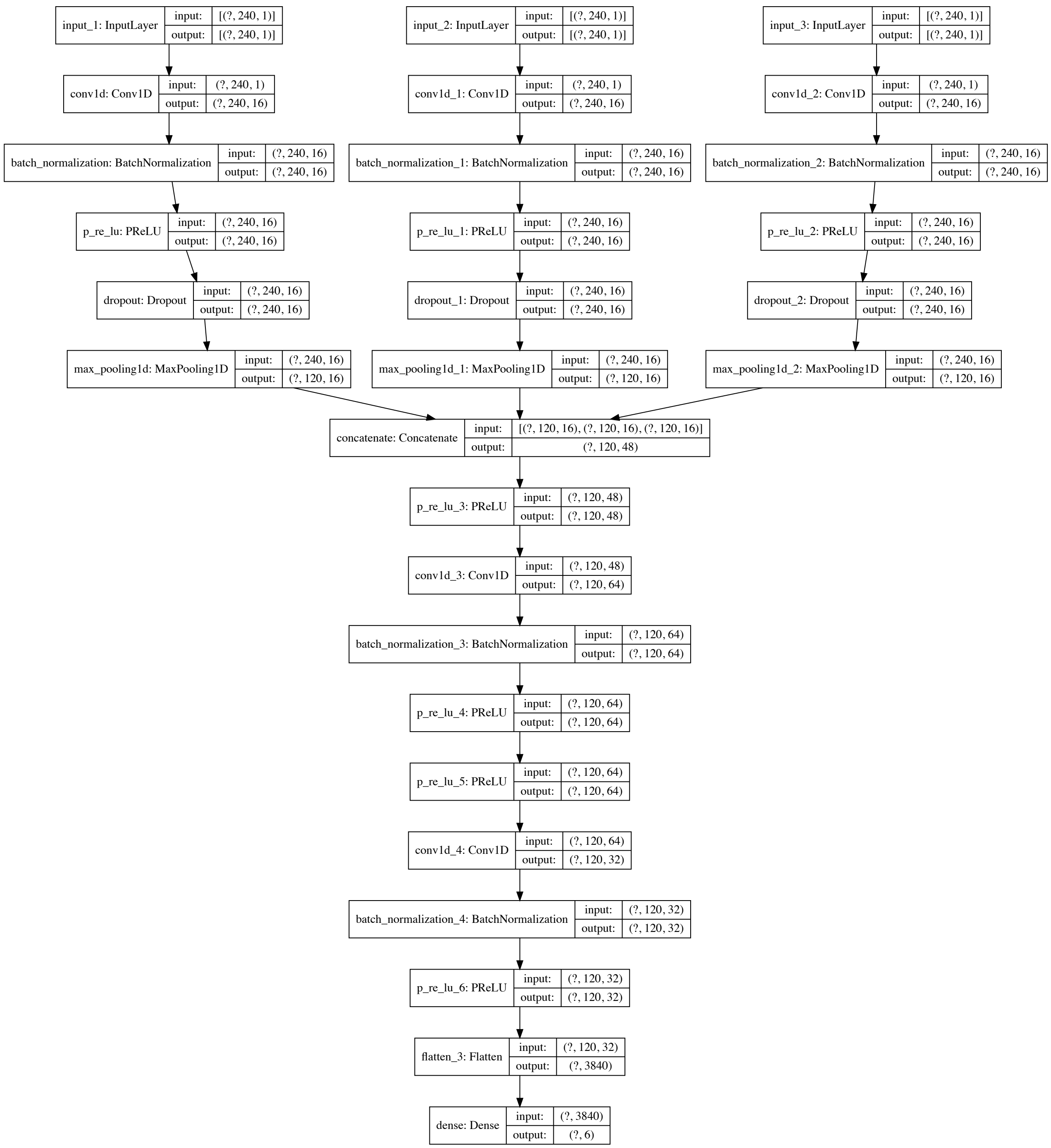

Figure 5. Three-headed CNN model architecture

Figure 6. Six-headed CNN model architecture

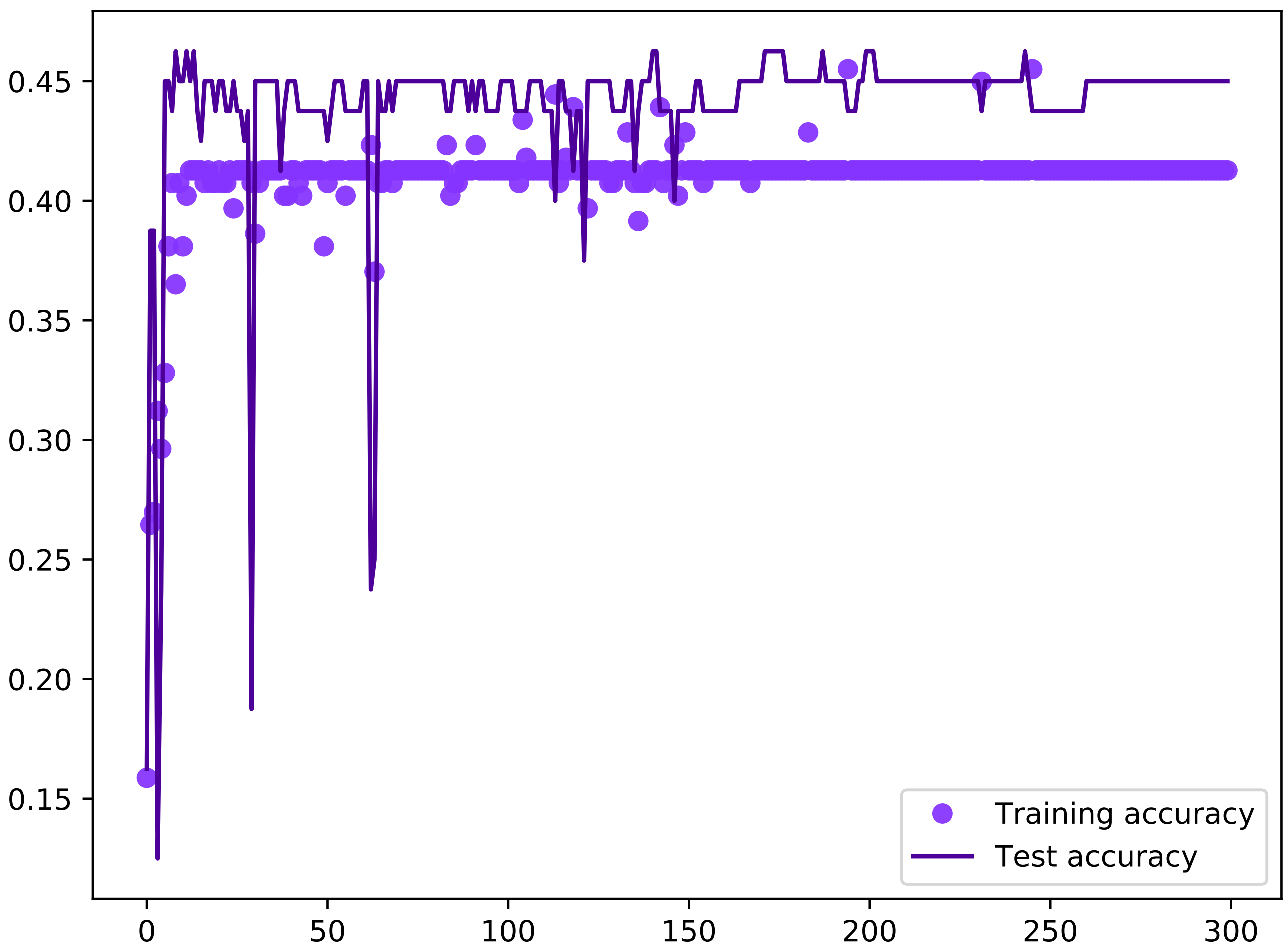

Figure 7. Training and test accuracy for the Two-headed CNN model trained on accelerometer and meditag sensor modality.

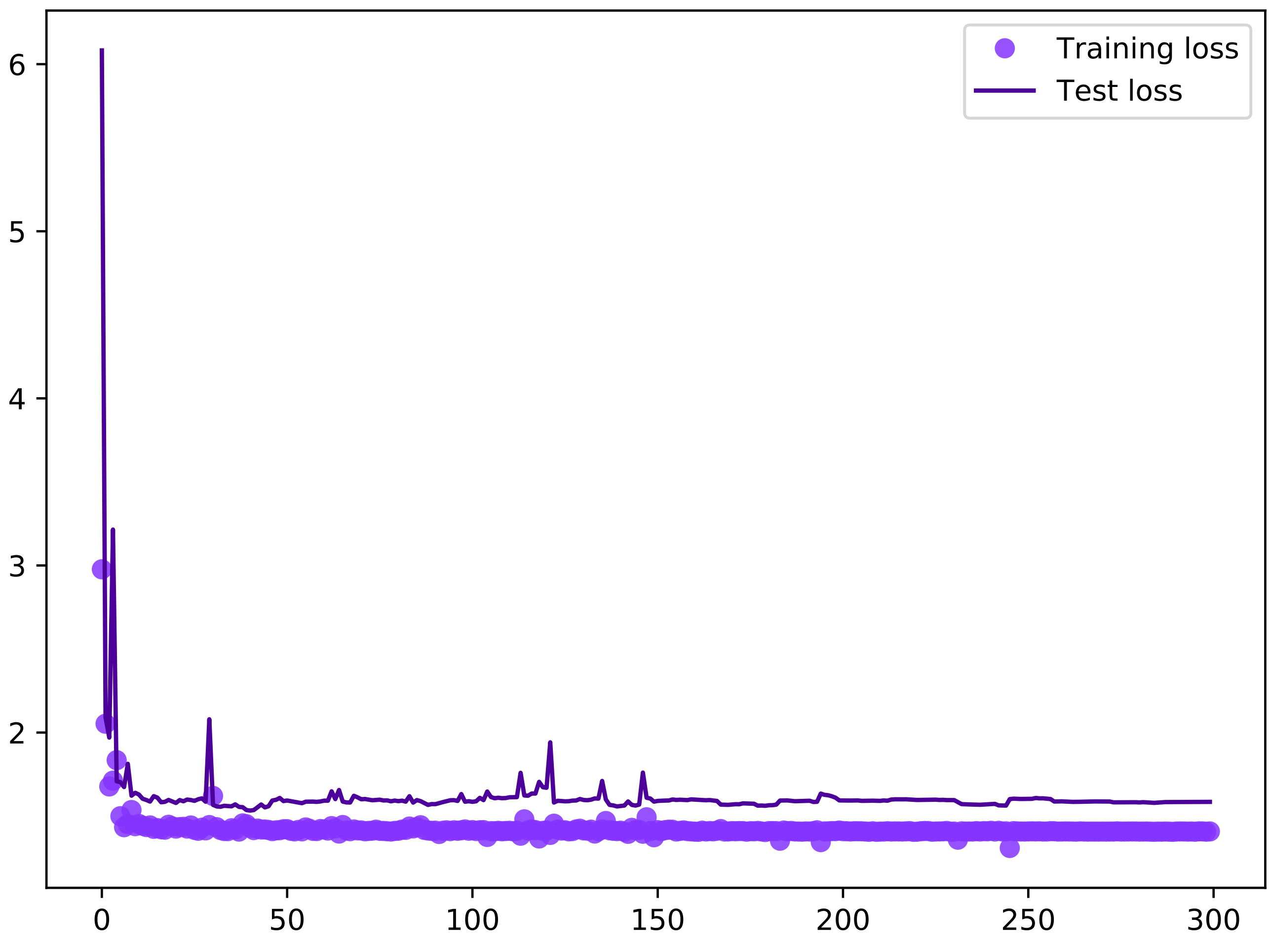

Figure 8. Training and test loss for the Two-headed CNN model trained on accelerometer and meditag sensor modality.