"Deep Sleep with Smart Beds" Subject and Posture Classification with Deep Neural Networks

Elena Camuffo[†], Daniele Orsuti[‡]

***Abstract*—The quality of our life is highly influenced by the quality of our sleep. The monitoring of individuals' sleep can ensure their well-being and allow the prevention of diseases, like pressure ulcers and obstructive sleep apnea.**
Sleep condition data can be gathered by means of smart beds and other technologies. In this work, a public dataset of pressure map images is exploited to provide a mean able to classify postures and subjects.
The task is carried out with a multi-branch Convolutional Neural Network (CNN), a deep learning model inspired by the Inception module. The model is proved to outperform the state-of-art, obtaining an accuracy of near 100% over the three main postures (supine, left and right), and 91%, considering an extended set of 17 postures. This experiment was carried out following a leave-one-subject-out (LOSO) validation scheme, to further investigate the robustness of the model. Moreover, a test is performed on the joint subject and posture recognition, using a k-fold cross validation scheme, obtaining an accuracy higher than 99%.
However, the proposed CNN is not enough to exploit the temporal correlation of frames in the sequences of images provided. Therefore, a recurrent architecture is introduced.

The Convolutional Long Short-Term Memory (LSTM) model proposed can achieve an accuracy near 86%, despite its simplicity and its limited number of parameters. It is probably the most promising for future researches, as the main purpose of this scenario is to work with frame sequences and exploit the temporal correlation of patients' sleep data.

***Index Terms*—Smart Beds, Sleep Posture Monitoring, Deep Learning, Convolutional Neural Networks, Recurrent Neural Networks.**

I. Introduction

A sufficient amount of quality sleep is essential to ensure the physical and mental well-being of an individual; a night of poor sleep can make a person feel fatigue on the next day and long-term sleep disorders will even induce a range of health problems. Numerous studies in the literature have shown that sleep position is highly related to sleep quality. For instance, supine posture is associated with obstructive sleep apnea syndrome [1], which can cause breath pauses overnight. Moreover, bed pressure ulcers are another serious health disease caused by remaining a long period of time in the same posture. Nowadays more than 2.5 million people in the United States develop pressure ulcers every year [2] with a total annual care cost of over $11 billion [3]. An ulcer may develop for example during a multi hour surgery or in post-surgery recovery in bed-bound patients. Instead of lengthy and expensive treatment, regularly turning the person can effectively prevent the development of this category of diseases. Therefore, monitoring the patients' posture change over time and alerting the caregivers if a repositioning is needed, can guarantee a healthy permanence in hospitals.

The main focus of this work is on in-bed posture and subject classification using electronic pressure mapping systems. Many in-bed posture detection techniques exist, and they are categorized depending on the employed technology such as video cameras [4], wearable sensors [5], or pressure mattresses [6] [7]. Among the aforementioned approaches, pressure based pose detection systems avoid problems like occlusion, lightening variations and users' privacy.
In addition to classification of postures, smart beds can enable automated recognition and identification of users, which extends their applications for security and authentication purposes, as well as personalization of smart-home experiences.
Since the pressure mat provides a two-dimensional array of pressure values, posture detection can be considered as an image processing task.

Despite the number of existing frameworks to deal with pressure maps data, the accuracy which is currently offered ranges in between 80% to 90%. The difficulty arises from the fact that the person-specific factors such as height, weight and body shape can cause a high variability in pressure images even if restricted to a specific posture. Moreover, the pressure mapping systems limit the resolution of the acquired image.
These issues can be addressed by applying effective preprocessing techniques and deep neural networks architectures as it is presented in this work. A public pressure map dataset, PmatData [6] has been used to train and test a deep multi-branch CNN. The proposed framework is validated incorporating a leave-one-subject-out validation scheme as well as a k-fold cross validation, in order to compare the results with previous studies on the same dataset [6] [7]. In addition, a novel approach is presented: it consists in learning spatiotemporal features of consecutive frames with a convolutional LSTM unit based neural network.

[†]Department of Information Engineering, University of Padova, email: elena.camuffo@studenti.unipd.it
[‡]Department of Information Engineering, University of Padova, email: daniele.orsuti@studenti.unipd.it
Special thanks to Prof. Michele Rossi.

Fig. 1: Different postures present in the dataset.

The rest of this paper is organized as follows. In section II the state-of-art works are described. Section III introduces the processing pipeline and then a description of the dataset and the preprocessing techniques is provided in section IV. Finally, in section V the learning framework is given, and the results are reported in section VI. Final conclusions are drawn in section VII.

II. Related Work

As this paper is mainly concerned with Deep Neural Network models, the higher impact on this project concern the whole setup and data manipulation of M. Heydarzadeh, M. Nourani and S. Ostadabbas in [7] and the deep model of Yan-Ying Li et al. in [4]. Other approaches, like the work of G. Matar and G. Kaddoum employed a Feed Forward Neural Network (FFNN) for the classification task in [6]; Mehrdad Heydarzadeh et al. used an autoencoder to extract a relevant features' representation [8] and Zhou Tianyu et al. mixed a CNN model with features manually extracted [9].

Up to now, researchers have applied different image processing techniques to solve in-bed posture and subjects' classification. In [6] it is possible to identify the first attempt to classify subjects using bed posture data. Manually extracted features were fed to a dense network, pre-trained by incorporating a restricted Boltzmann machine. The authors published their experimental dataset which is used in this paper, too. However, they focused on subject classification in just three standard postures: *right*, *supine* and *left*. Similarly, [10] proposed a FFNN using as input HoG+LBP features extracted from the pressure images for identifying 4 standard postures. Most of the existent studies only focused on the identification of few standard postures. But legs and arms positions are also important factors for achieving quality sleep and to avoid spinal alignment problems in the long term.

This is why in [7] a deep CNN is used to classify subject and postures from a single frame of data. Both tasks are accomplished simultaneously using a combined loss function. They obtained high accuracy in identifying 17 in-bed postures adopting a LOSO validation scheme, i.e. their model performs well even on data considerably different from the training ones.

In this work it is proposed an extension of the latter approach which aims at extending its range of applicability. Temporal features from consecutive in-bed posture frames are furtherly taken into account.

III. Processing Pipeline

The subject and posture classification task is carried out by exploiting the following processing pipeline:

1) Pressure images extraction from *.txt* files and preprocessing.
2) Training of a deep learning model for classification.

The preprocessing is needed to improve the body shape detection efficiency, as the original raw data are pressure maps, thus limited resolution images subject to artifacts.

Several preprocessing techniques have been tested. The selected one consists of two filtering steps (median and thresholding) and a histogram equalization. This preprocessing combination is shown to have a reasonable impact on the outcome of the training, as it leads the models to improve their performances and reduce their convergence time.

Also a number of different models have been tested in order to accurately solve the classification problem.

The most promising one, has been shown to be a multi-branch CNN architecture, which roughly recalls the structure of the Inception module [4]. This model presents in addition some peculiarities proper of the state-of-art CNN model of [7], like two SoftMax activation output layers, placed in parallel, which allow to perform posture and subject classification at the same time. For this reason, the chosen loss function is modeled as

$$\mathcal{L} = \lambda \mathcal{L}_{subject} + (1 - \lambda)\mathcal{L}_{posture}, \tag{1}$$

where $\mathcal{L}_{(\cdot)}$ indicates the categorical cross-entropy loss function.

The multi-branch architecture is exploited to perform first a posture classification with a LOSO validation scheme and then a joint posture and subject identification, using a 10-fold cross validation scheme. The subject identification is not possible with the LOSO scheme as one subject at a time is not included in the training.

On the other hand, being this model a CNN, it does not consider the temporal correlation existing between pressure images, indeed each posture presents bunches of sequential acquired images. Therefore, a Convolutional LSTM model is put beside the first one, to discuss a comparison of performances, and it is proved to be effective too, holding a lower number of parameters.

IV. Signals and Features

Before going deeper into technical details, an overview of the datasets and the preprocessing techniques applied is given.

A. Dataset

The experiment is based on a dataset of raw data provided in *.txt* format files. The raw data are pressure map matrices, collected by means of smart beds. Specifically, a *Vista Medical FSA SoftFlex 2048* system is used to acquire sequences of 64x32 pressure map matrices, with a sampling rate of $1Hz$. The pressure map matrices are the raw data collected, ideally reporting numbers in range of $[0 - 10\,000]$ for each sensor (in practice the highest value reached is 4095). The sequences are held for approximately 2 minutes each for a total of around

120 per subject (but some sequences are shorter than others). The experiment shows 13 subjects involved and 17 postures, each belonging to one of the main three: *supine*, *left* and *right*.

The overall number of frames was of 20 024, but a margin of 3 frames at the beginning, and 3 frames at the end was removed together with some corrupted black frames, reducing the dataset to 18 681 elements.

For the 10-fold experiment, a 10% of the dataset is retained for testing, and a 10% of the training set is used for validation.

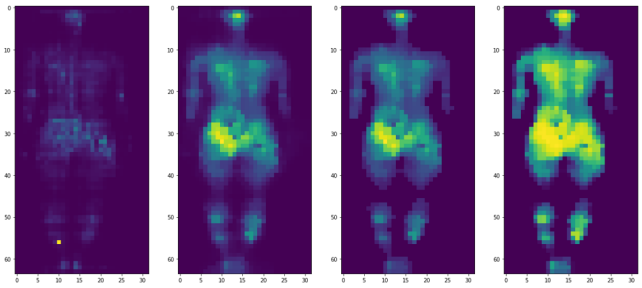

Fig. 2: Preprocessing flow visualization (supine posture). Note the difference between (b) and (c) in the knees' area.

B. Preprocessing

The preprocessing stage is involved with three steps, performed before splitting the dataset and train the Network. They consist of:

1) Median filtering with a 3x3 kernel.
2) Threshold filtering [11], applying:

$$h(x) = \begin{cases} x & \text{if } x > T \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

   where $T = 15$, and x is the intensity of the pixel.
3) Histogram equalization (over the single channel of the pressure map images).

The Median filter, used also by [7], is needed to reduce the noise caused by the occasional malfunctioning of pressure sensors, which manifests in artifacts on the image. In addition it contributes to smooth the map and reveal the body shape. The Threshold filter, instead, is used in combination with the histogram equalization step, to perform the equalization along the body shape while not affecting the background.

The application of this sequence of preprocessing steps makes the images more distinguishable even from a human eye (fig. 2). In addition, the joint application of these 3 steps leads to an improvement in the training of the model itself, rather than applying only a 3x3 median filter, as in [7].

V. Learning Framework

As discussed above, the models used consist of a multi-branch CNN and a convolutional LSTM architectures.

A. Multi-branch CNN architecture

Figure 3 shows the block diagram representing the structure of the implemented network. Inspired by the Inception module, three branches has been stacked together. The main block consists of Conv-BatchNorm-MaxPool-LeakyReLU and it is replaced by Conv-BatchNorm-LeakyReLU where MaxPooling was not applicable. Each branch includes a combination of the above main blocks terminated by fully connected layers. For each branch, different kernel sizes have been used (3×3, 5×5, 7×7). As a result the multi-branch structure introduces a parallel multi-scale analysis capable to catch multiple size patterns in pressure image data. All the features of the three branches are then concatenated into one dimensional vector of size 1024. Finally, the classification is achieved by feeding the outcome to two dense layers (with dropout rate of 50%), followed by two SoftMax used to allow simultaneous classification of subjects and postures. Each convolutional block has been followed by an increasing dropout rate of 10%, 20%, 30%. The dropout layers allow the network to become less sensitive to the specific weights of neurons. This results in a network capable of better generalization and which is less likely to overfit the training data. *L2* regularization loss was also employed for similar purposes using a coefficient of $\sigma = 0.004$.

At the training stage the following hyper-parameters are selected: batch size equal to 64, number of epochs equal to 40, Adam optimizer with variable learning rate minimizing the mixed categorical cross-entropy loss (equation 1). The learning rate, is set to the initial value of 2×10^{-3} and decayed with a rate of 0.95 every 10 epochs.

As discussed in section III, two validation schemes are adopted to validate the proposed method, k-fold and LOSO. In the k-fold cross-validation the hyperparameter λ is set to 0.5. On the other hand, in the LOSO scheme, subject classification is not possible, since evaluation needs to be done on the test subject. Therefore, in the latter case λ is set to 0, since no improvement is shown in minimizing both losses.

Data Augmentation– In order to test the network generalization abilities, a data augmentation operation is performed. Due to the limited hardware capacities, each frame is augmented of a factor 3, resulting in a total number of $56\,043$ images. Table 1 reports the geometric transformations randomly applied to the pressure maps. In this case, the training is performed with 50 epochs instead of 40.

Relative weight	Transformation
50%	Rotation of 180^o
20%	Translation by up to $\pm 10\%$ along x
20%	Translation by up to $\pm 10\%$ along y
20%	Rotation by up to $\pm 10^o$

TABLE 1: Geometric transformations used for data augmentation.

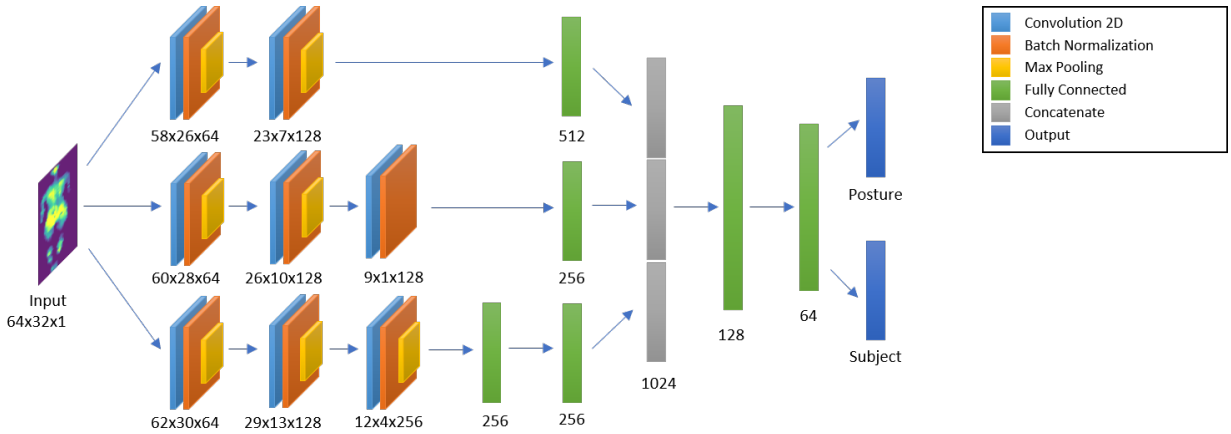

Fig. 3: Multi-branch CNN architecture scheme.

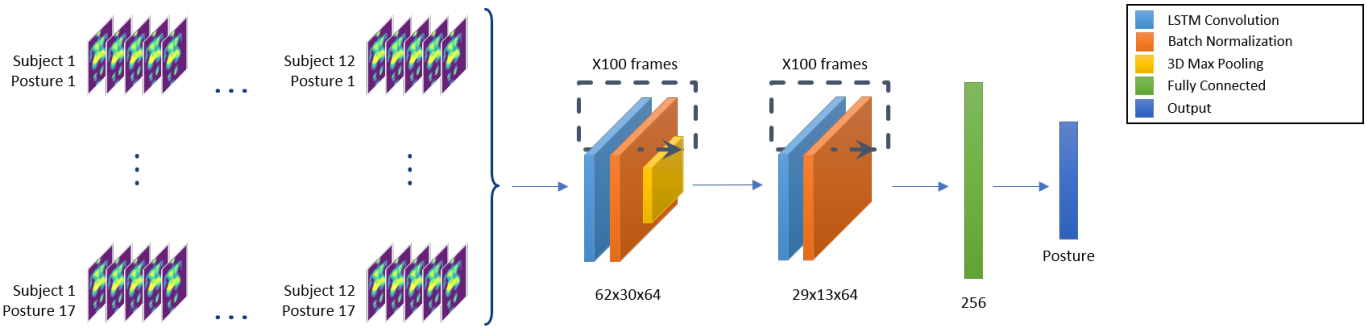

Fig. 4: Convolutional LSTM architecture scheme.

B. Convolutional LSTM architecture

Figure 4 illustrates the novel proposed framework to deal with subsequent frames of in-bed postures. It consists of 2 stacked ConvLSTM layers with batch-normalization, tanh activation fuction and 3D MaxPooling after the first one. The layers have 16 (3×3) and 16 (3×3) filters, respectively. The second ConvLSTM layer removes the temporal dimension and its output is fed to a dense layer with ReLU activation function, followed by a SoftMax activation.

In this architecture, pressure maps data belonging to the same posture and subject, have been treated as sequential data and not as isolated frames. The subdivision in bunches has been performed as shown in figure 4.

In such cases an interesting approach is to use model based on LSTM cells. Here, previous outputs are allowed to be used as input while having hidden states. Therefore holding information on previous data, the network is able to reuse those information about just seen frames to make decisions. However, LSTMs can not directly learn spatio-temporal features from a sequence of images. This limitation is overcome replacing the Hadamard product of the original LSTM with the convolution operation (ConvLSTM) [12]. As a result data that flow through the ConvLSTM cells keep the image input dimension instead of being just a 1D feature vector.

Ideally, the cell state will not be reset until the entire time sequence is fed to the network, but the training data are provided to neural network in batches with sizes restricted by the GPU memory capacity. When the input data consist of sequential images, it is important to determine both the number of frames in each input sample (the bigger, the more long-term dependencies are captured) and samples in each batch (the more, the better the model is able to generalize and avoid overfitting).

In PmatData dataset, sequences length ranges from 60 to 207 frames. However only one reaches 207 and the average length is about 100. Therefore the network is fed with temporal sequences that are 100 frames long with a batch size of 32. The sequences shorter than 100 have been padded by repeating the last frame and the longer ones have been cropped. Also other settings have been tested. For instance, subsampling the frame sequences by a factor of 5 and training with batch size of 8, reduces training time and selects frames which are temporally spread limiting their similarity. The network is trained for 40 epochs, using Adam optimizer with learning rate of 10^{-3}, decayed of 0.95 every 10 epochs. Dropout layers are inserted after each recurrent block with increasing rate of 10% and 20%, and after the dense layer with rate 30%. The model is tested using LOSO validation scheme.

LOSO posture identification over 17 postures

Model	Preprocessing	Augmentation	Hyper-parameters	Accuracy	Precision	Recall	F1 score
CNN ref. [7]	m [7]	no	$\lambda = 0.2$	85.1	84.5	85.8	82.2
CNN ref. [7]	m, t, e	no	–	88.9	86.7	88.7	85.8
Multi-branch CNN	m, t, e	no	–	**91.0**	**89.2**	**91.1**	**89.0**
Multi-branch CNN	m, t, e	no	no reg.	86.3	86.6	86.3	82.8
Multi-branch CNN	m, t, e	no	1 final dense	89.4	87.2	89.4	87.0
Multi-branch CNN	m, t, e	3x	–	90.4	90.9	90.3	90.5
Convolutional LSTM	m, t, e	no	–	**85.6**	**85.7**	**85.6**	**85.4**
Convolutional LSTM	m, t, e	no	subsampling (1/5)	82.0	83.1	82.0	82.1

TABLE 2: Comparison of models, preprocessing techniques and hyper-parameters settings. In the table, metrics are in % and m, t, e stand for median, threshold and equalization, respectively. If not specified, $\lambda = 0$ and dropout/regularization are set as in section V.

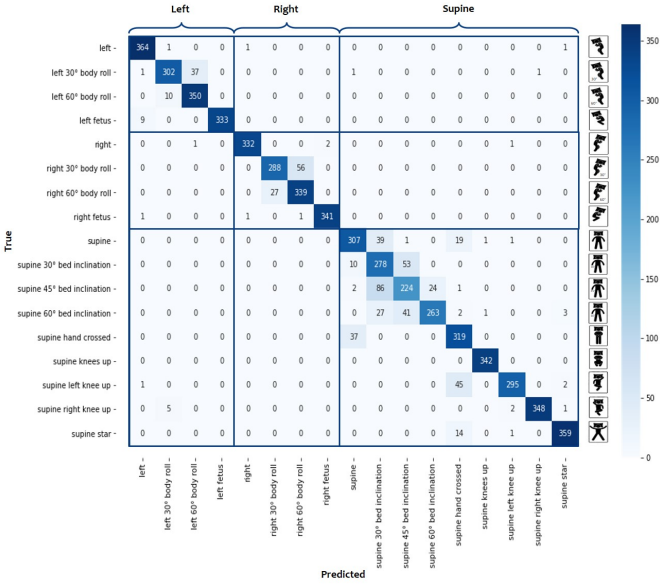

Fig. 5: Average Confusion matrix obtained using LOSO validation scheme on multi-branch model and data augmentation (6^{th} row of table 2); images are preprocessed as in sec. IV.

10-fold subject identification

Model	Supine	Left	Right
FFNN ref. [6]	85.5	82.3	80.4
CNN ref. [7]	99.9	100	100
Multi-branch CNN	100	99.9	99.9

TABLE 3: Comparison between references data and accuracies (in %) found using multi-branch architecture for subject identification with 10-fold cross validation scheme.

VI. Results

Considering the multi-branch CNN model, the experiment with 10-fold cross validation, results in a simultaneous posture and subject recognition accuracy of near 100% over 17 postures. In addition, table 3 presents the results of subject identification obtained training the model separately over the three standard postures. The values of accuracies reached are very high and close to the results obtained with the CNN model in [7]. This method, however, does not allow to generalize well. Indeed, frames belonging to the same sequence of a selected subject may appear in different sets (train, test, validation), but they result to be very similar one another, since the individual is lying in the same position all the time long; consequently the classifier is led to recognize the images' position because very close to the ones it has already seen, but it is not able to generalize, when new images are tested (e.g. if a new subject is added).

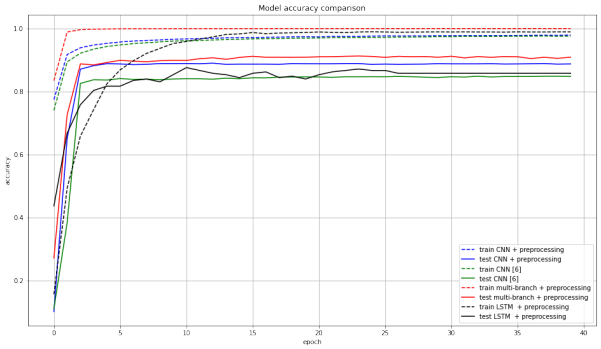

Fig. 6: Accuracy curves for training and test sets as a function of models and preprocessing techniques. LOSO validation scheme has been used.

LOSO posture identification

Model	Augm.	Supine	Left	Right
CNN ref. [7]	no	99.0	99.7	100
Multi-branch CNN	no	100	100	100
Multi-branch CNN	3x	99.7	99.2	99.7
Convolutional LSTM	no	98.1	94.3	98.3

TABLE 4: Comparison between references data and accuracies (in %) found using multi-branch architecture for posture identification with LOSO validation scheme.

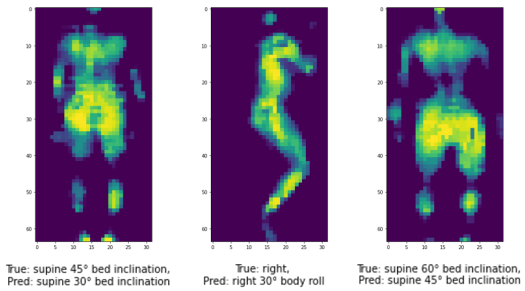

Fig. 7: Example of errors committed by the multi-branch model in the classification.

This is why LOSO scheme is a way more appropriate validation procedure to prove the robustness of the model. In this procedure, the training set is composed of all subjects except one, used for testing. This way the network is tested on images it has never seen before and this verifies its learning capabilities. Table 4 illustrates the model performances in the classification of the three main postures evaluated using LOSO cross-validation. It can be observed that there is not drop in accuracy with respect to the 10-fold cross validation confirming the robustness of the proposed model. The LOSO validation scheme has then been extended to the classification of all 17 available postures. Figure 6 shows that the proposed multi-branch CNN and the applied preprocessing techniques lead to an improvement in accuracy compared to the CNN of [7]. It can be observed that the multi-branch CNN model quickly converges to a steady state and reaches the highest training/test accuracies (red curves). The average accuracy reached in the test set reads 91% after around 5 to 10 epochs, while in the training set an accuracy of 100% is reached after only 2-3 epochs, outperforming [7], whose training accuracy convergence is slower and never reaches values higher than 98% (green curves). Table 2 reports the outcomes of several experiments led on the dataset, exploiting different models, different configurations of the networks and hyper-parameters.

The level of accuracy is confirmed also by the average confusion matrix of figure 5, which is divided in sectors corresponding to the three main postures, and reports almost all zero values in the inter-class sections of the matrix. As stated above, misclassified sub-postures fall all within the correct main postures and the committed errors concern sub-postures with subtle differences (figure 7). For instance, variations of supine postures where bed inclination varies of 15^o between sub-postures, result hard to be classified even for a human eye.
Finally, data augmentation confirms the validity of the model reaching an accuracy higher than 90% over 17 postures.

Considering the results obtained using the recurrent architecture, they read almost 97% on average over the three main postures and a reasonable 86% over all the 17 postures. The metrics of the experiments led are reported in table 2, while in figure 6 the outgoing of training and test curves is shown (black curves). The average test accuracy is, as expected, lower than the other methods, partly due to the lower number of parameters this model holds, with respect to the multi-branch CNN (almost $1/4$). Moreover, as stated above, subsequent frames of the same posture present low variance. Thus, in this case, there are not significant temporal features which can be extracted and further contribute to a more accurate classification of the considered posture. However, this is a good starting point for future developments, as it involves also temporal features. With a powerful hardware capability, the model could be largely improved and integrated with additional features, or more meaningful ones.

VII. Concluding Remarks

A. Conclusions

From this work emerges that a multi-branch CNN model using different kernel sizes, is able to learn an enhanced feature representation of the input samples, and accomplish both posture and identity recognition tasks.
The impact of the preprocessing significantly affects the final outcome, increasing the accuracy of the model of near 4%. But the model has to face some limitations deriving from the dataset: quite large, but holding pressure data, thus limited resolution images. Results from data augmentation and LOSO validation scheme show that the method can be extended even to larger datasets while maintaining an accurate classification capability. Finally, reasonably good performances are also achieved by the ConvLSTM model, even if there is still room for improvement.

B. Future works

The ability of ConvLSTMs based neural networks to extract spatio-temporal features from a sequence of images is appealing and is worth to spend some time to further develop the recurrent model presented above. Extending the pressure dataset including temporal signals such as heart rate, body parts temperature and respiration rate may help to achieve even more accurate systems and deeper understand sleeping posture effects on an individual.

C. What we have learned

The major difficulty we have encountered lies in the great similarity among the images of the dataset. They resulted too much similar to perform a good analysis on the frame sequences and to make a classic pseudo-random division of the overall initial dataset. Also selecting the right model, keeping safe this restriction, was not so easy.

The most important thing we have learned doing this project, concerns how to build state-of-art model architectures and components, from scratch, referring only to written papers. The drafting of this paper was also really helpful, to improve our writing and also reading abilities for future works.

References

[1] R. D. Cartwright, "Effect of sleep position on sleep apnea severity," *Sleep*, vol. 7, no. 2, pp. 110–114.

[2] L. Soban, S. Hempel, B. Munjas, J. Miles, and L. Rubenstein, "Preventing pressure ulcers in hospitals: A systematic review of nurse-focused quality improvement interventions.," , vol. 37, no. 6, pp. 245–252.

[3] E. Haesler, "National Pressure Ulcer Advisory Panel, European Pressure Ulcer Advisory Panel and Pan Pacific Pressure Injury Alliance," in *Prevention and Treatment of Pressure Ulcers: Quick Reference Guide*, 2014.

[4] Y.-Y. Li, Y.-J. Lei, L. C. ling Chen, and Y.-P. Hung, "Sleep Posture Classification with Multi-Stream CNN Using Vertical Distance Map," Jan. 2018.

[5] P. Jeng and C. W. D. Wang, L.and Hu, "A Wrist Sensor Sleep Posture Monitoring System: An Automatic Labeling Approach.," *Preprints*, 2019.

[6] M. B. Pouyan, J. Birjandtalab, M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "A Pressure Map Dataset for Posture and Subject Analytics," in *IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, (The University of Texas at Dallas, Richardson, TX, USA), Feb. 2017.

[7] V. Davoodnia and A. Etemad, "Identity and Posture Recognition in Smart Beds with Deep Multitask Learning," in *IEEE International Conference on Systems, Man and Cybernetics (SMC)*, (Department of Electrical and Computer Engineering, Queen's University, Kingston, ON, Canada), Oct. 2019.

[8] M. Heydarzadeh, M. Nourani, and S. Ostadabbas, "In-bed posture classification using deep autoencoders," in *38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, (Orlando, FL, USA), Aug. 2016.

[9] Z. Tianyu, M. Zhenjiang, and Z. Jianhu, "Combining CNN with Hand-Crafted Features for Image Classification," Aug. 2018.

[10] G. Matar and G. Kaddoum, "Artificial Neural Network for in-Bed Posture Classification Using Bed-Sheet Pressure Sensors," *IEEE journal of biomedical and health informatics*, vol. 24, Jan. 2020.

[11] X. Xu, F. Lin, A. Wang, C. Song, Y. Hu1, and W. Xu, "On-bed Sleep Posture Recognition Based on Body-Earth Mover's Distance," Oct. 2015.

[12] X. Shi, Z. Chen, H. Wang, D. Yeung, W. Wong, and W. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," *CoRR*, vol. abs/1506.04214, 2015.