# CNN-Based Analysis of Pressure Maps for Posture and User Classification

Simone Gastaldon

***Abstract*—Recent advances in smart sensing technologies have enabled the collection of high-resolution pressure maps from sensorized mattresses, offering new opportunities for posture monitoring and user identification. Accurate and non-invasive recognition of both posture and identity is challenging due to inter-subject variability and limited data per individual. In this work, I investigate several convolutional neural network (CNN) architectures to address this problem, including baseline models, networks with data augmentation, literature-inspired designs, and CNNs enhanced with Inception blocks. My results show that the inclusion of Inception modules significantly improves both posture classification and subject identification, outperforming simpler architectures while maintaining reasonable computational requirements. Analysis of the confusion matrices confirms that posture recognition is relatively straightforward, whereas subject identification remains more challenging due to the number of classes and limited training examples. These findings highlight the potential of Inception-based CNNs for practical applications in non-intrusive health monitoring and personalized care, where accurate posture tracking and user differentiation are essential.**

***Index Terms*—Pressure maps, Convolutional Neural Networks, Posture recognition, Subject identification, Smart sensing, Inception modules.**

## I. INTRODUCTION

Noninvasive monitoring of a person's daily activities and physical condition is a topic of great interest in modern research. Estimating posture during sleep and identifying the subject based on pressure maps generated by a sensorized mattress is useful for telemedicine applications, bedsore prevention, and personalized safety systems.

Convolutional neural networks (CNNs) have proven to be highly effective at deciphering intricate signals and pictures, such as pressure maps. Nevertheless, a lot of the solutions found in the literature depend on extremely big models that have a lot of parameters and high processing costs. Utilizing such models in resource-constrained real-world contexts, like embedded or edge devices, is challenging due to these limitations. Lighter models that can continue to function well even when memory and processing time are limited are required in these situations. As a result, creating small networks that can handle both subject identification and posture categorization is still a difficulty.

In this study, I suggest and assess a small, multitask CNN. Compared to more intricate networks, the model uses less computing power to estimate posture and subject identity from pressure maps at the same time. I contrasted the suggested network with versions that incorporate Inception blocks, versions based on data augmentation approaches, and versions with larger architectures. Analyzing the trade-off between generalization ability, efficiency, and accuracy is the goal. The findings demonstrate that a smaller network can outperform more complicated models in posture classification and person identification while requiring less memory and training time.

The contributions of this work are as follows:

- design and evaluation of a compact CNN for multitask analysis of postures and subjects;
- systematic comparison with more complex models and variants that include data augmentation and Inception blocks;
- analysis of the trade-off between performance and computational resources, with a focus on resource-limited scenarios.

The rest of the report is organized as follows. In Section II, I present the state of the art. In Section III, I describe the dataset and experimental setup. Section IV introduces the proposed model and the variants considered. Experimental results are reported in Section V. Finally, in Section VI, I present the conclusions.

## II. RELATED WORK

I relied on two reference articles that address the problem of subject and posture recognition using data from pressure sensors.

The main contribution of the Heydarzadeh et al. (2017) work was to demonstrate how, through image registration techniques and eigen-footprint-based dimensionality reduction, it was possible to obtain a compact representation of footprint sequences [1]. Subsequently, the use of an SVM classifier allowed them to achieve an accuracy of 97% with 10-fold validation. This study convincingly showed that features extracted from dynamic footprints contain unique and sufficiently discriminatory information for biometric identification. However, the proposed method, although effective, relies on traditional feature extraction pipelines and non-neural classification models, which do not fully exploit the potential of deep networks.

A different approach was followed by Davoodnia and Etemad (2021), who developed an end-to-end multitask deep learning-based model capable of simultaneously recognizing a subject's identity and bed posture using pressure maps [2]. The most relevant contribution of this work consists in the use of a multitask CNN with a combined cost function, capable of significantly improving performance compared to classical methods. The results obtained were extremely high: almost 100% accuracy with k-fold validation and over 99%

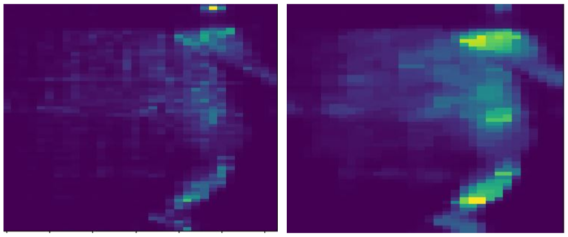

Fig. 1: Before and after preprocessing

with the leave-one-subject-out (LOSO) scheme. Despite the demonstrated effectiveness, the architecture used is complex and requires considerable computational resources, which limits its applicability in low-computing power contexts or in embedded devices.

In summary, the two studies cited above represent the main references in the field of pressure sensor analysis for recognition and classification. On the one hand, Heydarzadeh et al. validated the effectiveness of pipelines based on manual feature extraction, on the other, Davoodnia and Etemad showed how deep neural networks can achieve superior performance, but at the expense of complexity. My work continues these approaches, proposing a more compact CNN implementation, with the aim of evaluating whether it is possible to maintain a comparable level of accuracy while reducing the required resources.

## III. Processing Pipeline

Data preprocessing was aimed at ensuring the quality and consistency of recordings from two different experiments. After selecting only valid recordings and extracting subject and posture information, the data was divided into training, validation, and testing sets. The sequences were harmonized in terms of size and quality, removing potentially corrupted frames and adjusting the resolution. Filtering and normalization were then applied to improve data consistency. Finally, the recordings were synthesized and organized in a TensorFlow-compatible format, making the dataset ready for the model training and validation phases.

## IV. Signals and Features

I used pressure map data provided in a public dataset, *PmatData*, to train and test our pressure-based posture and user recognition system. The pressure data was collected using the Vista Medical FSA SoftFlex 2048 sensor. Each mattress contained 2048 one-square-inch sensors arranged on a $32 \times 64$ grid. The sensors reported values in the range 0 - 1000. Data was collected at a sampling rate of 1 Hz from 13 participants in 8 standard and 9 additional uncommon postures. The participants were aged between 19 and 34 years, with heights ranging from 170 to 186 cm and weights between 63 and 100 kg. A total of about 1800 samples was recorded for each subject.

In the first phase of preprocessing, a text file containing the complete paths of all available resources was parsed to extract only the valid recording references, systematically eliminating paths not corresponding to actual recordings. From each path I extracted the subject identifier and the posture label, and these were organized into a DataFrame that served as the structured basis for dataset management.

To ensure proper evaluation and model generalization, the dataset was divided into three subsets. A 90/10 split was applied to separate train+validation and test data, followed by a 75/15 division within the training set to obtain training and validation subsets. The splits were performed while balancing the distribution of both postures and subjects.

Given the presence of two different experiments in the dataset, a standardization process was applied. In the second experiment, recordings consisted of frame sequences. The first four and last four frames of each sequence were removed to avoid corrupted data from the start or end of recordings. In addition, the original frame resolution of $64 \times 27$ was rescaled to $64 \times 32$ to match the size of images generated in the first experiment, allowing the two sets to be used together for training.

Each recording was then filtered using a median filter, which reduces impulsive noise while preserving spatial structures. After filtering, the data was normalized using Min Max scaling, bringing values to the [0,1] range to improve numerical stability during training. An internal consistency analysis was also carried out to check for anomalous variations or discontinuities within the sequences. Finally, each sequence was reduced to a compact but informative representation by averaging all its frames.

The preprocessed data was then converted into TensorFlow `tf.data.Dataset` objects, ensuring compatibility with the training and validation pipelines of the deep learning models.

## V. Learning Framework

The proposed model is a 2D Convolutional Neural Network (CNN) with a multitask structure, designed for the simultaneous analysis of subject posture and identity. The input to the system consists of pressure maps transformed into RGB images of size $64 \times 32 \times 3$. The architecture is composed of three sequential convolutional blocks. Each block integrates a 2D convolution with a $3 \times 3$ kernel, followed by a batch normalization layer and a LeakyReLU activation function with a negative coefficient of 0.2. The progressive reduction of spatial dimensionality is achieved through MaxPooling operations with a $2 \times 2$ window. To limit overfitting, dropout mechanisms with probabilities of 20% and 40%, respectively, were introduced in the second and third blocks, distributed between the convolutional and dense sections. The output of the convolutional part is then flattened and transformed into a latent space via a fully connected layer of 128 neurons with ReLU activation. The learned representation branches into two parallel classification heads: the first dedicated to posture recognition, with softmax activation on 3 classes, and the second oriented to subject identification, with softmax on 13 identities. This multitask configuration allows for feature sharing between the two tasks and promotes joint optimization of the latent representation.

TABLE 1: Results of models

| Model | Time (s) | Mem (MB) | Acc Post. | F1 Post. | Acc Subj. | F1 Subj. |
|---|---|---|---|---|---|---|
| CNN no augm. | 694.79 | 3046.44 | 89.86% | 89.64% | 36.23% | 35.70% |
| CNN with augm. | 704.45 | 3159.80 | 85.51% | 85.64% | 30.43% | 32.71% |
| CNN Paper | 697.86 | 3509.40 | 84.06% | 84.22% | 31.88% | 28.72% |
| CNN + Inception | 714.74 | 3325.41 | 95.65% | 95.59% | 49.28% | 47.60% |

The model is optimized through a weighted combination of the two Sparse Categorical Cross-Entropy loss functions. Defining $L_p$ as the loss associated with posture classification and $L_s$ as that associated with subject identification, the overall objective function takes the form

$$L = \lambda L_p + (1 - \lambda) L_s$$

where $\lambda \in [0, 1]$ is a hyperparameter that controls the balance between the two tasks. Training was conducted using the Adam optimizer with an initial learning rate of $2 \times 10^{-5}$. To increase the stability of the learning process, a Learning Rate Scheduler was implemented that reduces the update step by 10% every five epochs. Furthermore, an Early Stopping criterion based on monitoring the validation loss was introduced, with a patience of 10 epochs and automatic restoration of the weights corresponding to the best configuration. Model regularization was ensured by the use of Dropout, distributed with different probabilities in the convolutional and fully connected layers, and Batch Normalization, which helps reduce the internal covariate shift.

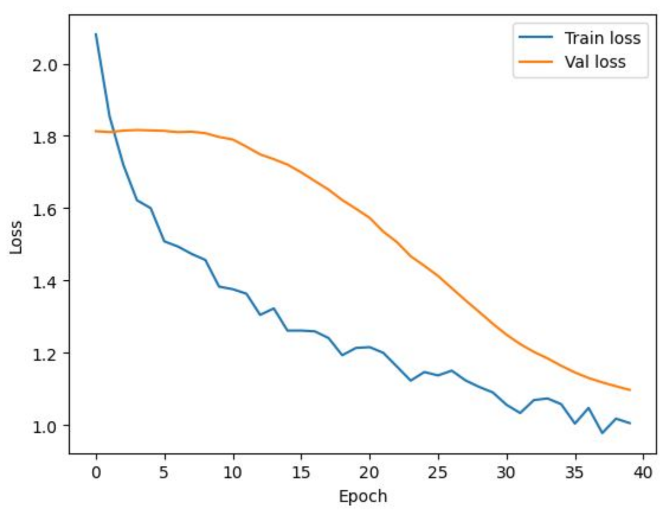


Fig. 2: Loss of base CNN no augm

The framework was evaluated using complementary validation strategies. First, a traditional split between training and validation data was used, useful for monitoring the convergence of learning curves and optimizing hyperparameters. As shown in Fig. 2, the loss curves indicate that some overfitting occurs during the first epochs, but the validation loss quickly aligns with the training loss, suggesting that the model stabilizes and generalizes well as training progresses.

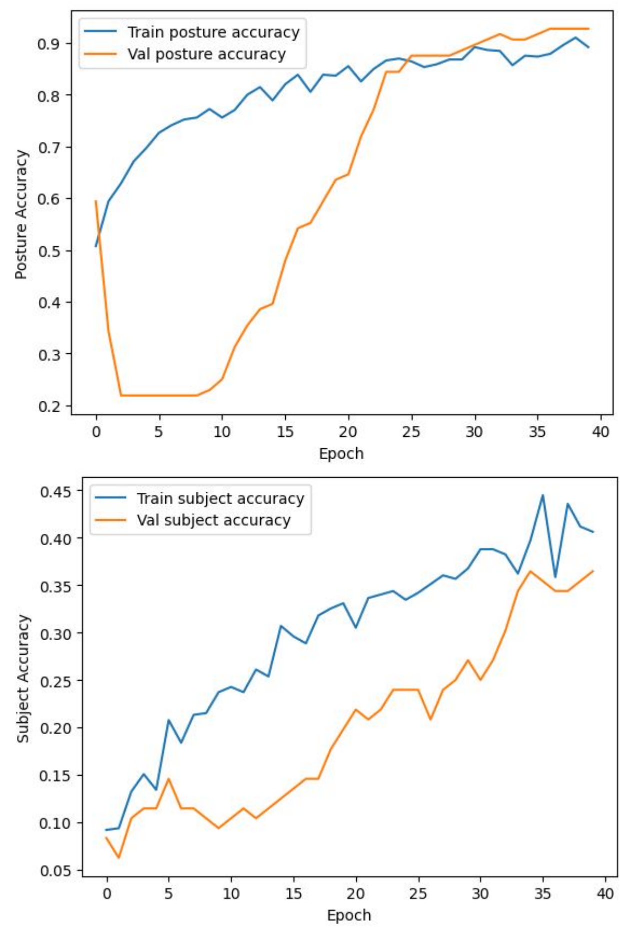


Fig. 3: Posture and Subject accuracies of base CNN no augm

Second, a stratified k-fold cross validation was implemented with k=3, providing an estimate of the generalization ability of the model and guiding the choice of the parameter $\lambda$ ($\lambda = 0.7$ found). Finally, comparative experiments with and without data augmentation were conducted to analyze the robustness of systems to variations in the pressure maps. The data augmentation techniques applied included random rotations up to $\pm 25$ degrees, horizontal and vertical translations up to 10% of the dimensions, reflections, and Gaussian noise injection with controlled variance. All these transformations were applied with a 50% probability.

To assess the impact of the architectural choices, comparative experiments were conducted with the model reported by Davoodnia and Etemad (2021) [2], which uses a traditional CNN with sequential convolutional blocks and a fully connected layer for multitask classification. Additionally, a variant was implemented in which the last convolutional block was replaced by an Inception module. This module integrates parallel branches with $1 \times 1$, $3 \times 3$, and $5 \times 5$ convolutions, preceded by $1 \times 1$ reduction layers to reduce computational complexity, as well as a max-pooling branch followed by a $1 \times 1$ convolution. The branch outputs are concatenated along the channel dimension, allowing for the capture of multi-scale spatial patterns in pressure maps. The remaining architecture, including the fully connected layer and the two softmax heads for postures and subjects, was left unchanged to isolate the effect of the Inception block. As visible in Figs. 3a and 3b, the training and validation accuracies for both posture and subject recognition show a similar trend: early overfitting in the first epochs is observed, but validation accuracy eventually reaches and stabilizes at the level of training accuracy. Training and evaluation with augmentation followed the same protocols as the baseline model, ensuring a fair comparison.

## VI. Results

Four distinct architectures were tested during the experiment: the model suggested in the literature, a CNN improved with an Inception block, the same network trained using augmentation approaches, and a baseline CNN without data augmentation. The objective was to use the pressure maps obtained from the sensorized mattress to assess the subject's identification and posture at the same time.

The baseline model showed high discriminative capacity on the first task but less performance on the second, with an accuracy of 89.86% in posture recognition and 36.23% in person identification.

The implementation of data augmentation did not result in the anticipated improvements: accuracy metrics dropped to 85.51% for posture and 30.43% for subject, with a fall in precision and F1 score.

The model based on the work of Davoodnia and Etemad [2] had a more complicated structure and used more memory, but it didn't work as well as the baseline: 84.06% accuracy for posture and 31.88% for subject. The analysis shows that the design suggested in the literature, while it worked extremely well in the original study, is not the best fit for the dataset and experimental setting used in this project.

The CNN enriched with Inception blocks achieved the most significant results, with an accuracy of 95.65% in posture classification and 49.28% in subject identification. Precision, recall, and F1-score metrics confirm a significant improvement in both tasks compared to the other models, while training times (714.74 seconds) and memory consumption (3325 MB) are in line with those of simpler architectures.

In general, the results suggest that adding Inception module makes it possible to get more robust and expressive representations of pressure maps, which makes it much easier to tell the difference between different postures and subjects. This confirms that designing the network architecture carefully has a bigger effect than preprocessing methods like data augmentation, at least in the case I looked at.

The confusion matrix analysis further confirms these results: correctly predicting posture is relatively simple, with few misclassifications among the main categories. In contrast, as seen in the two confusion matrices 4, identifying subjects is more complex; given the many classes to distinguish and relatively few examples for each subject, numerous misclassifications are observed. This underscores how the individual recognition task requires richer representations and more abundant data to achieve reliable discrimination.

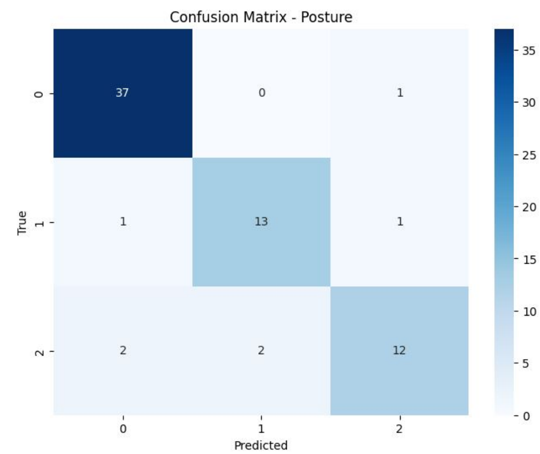


Fig. 4: Confusion matrix of posture prediction made by first CNN without augmentation

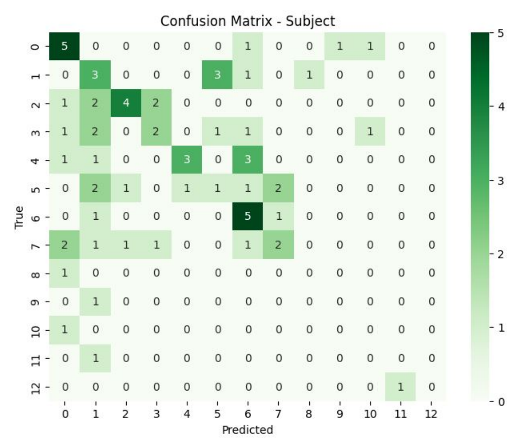


Fig. 5: Confusion matrix of subject prediction made by first CNN without augmentation

## VII. Concluding Remarks

In this study, various CNN architectures were developed and assessed to analyze pressure maps from a sensorized mattress, aiming to concurrently estimate the subject's posture and identity. The findings indicated that the implementation of Inception blocks markedly enhanced performance in both tasks, facilitating more precise differentiation between postures and participants. The examination of confusion matrices revealed that posture prediction is comparatively straightforward, however subject identification is more intricate due to the extensive number of classes and the scarcity of examples.

Observations suggest that CNNs using Inception blocks offer a viable alternative for posture monitoring and non-invasive user identification systems.

As the project progressed, I learned more about how to create convolutional networks and manage sensory data. The primary problems were that there wasn't enough data to identify subjects and that they had to choose designs that struck a compromise between speed and complexity.

For future work, it would be interesting to conduct further experiments focused on optimizing the Inception blocks, evaluating variants of their structure and parameters, in order to further improve the model's performance in both posture classification and subject identification.

## REFERENCES

[1] M. Heydarzadeh, J. Birjandtalab, M. B. Pouyan, M. Nourani, and S. Ostadabbas, "Gaits analysis using pressure image for subject identification," in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, (Orlando, FL, USA), pp. 333–336, IEEE, April 2017.

[2] V. Davoodnia and A. Etemad, "Identity and posture recognition in smart beds with deep multitask learning," *arXiv preprint arXiv:2104.02159*, April 2021.