This chapter will present the results from training and testing of the three data representations. Training results support arguments to how optimization of the models were performed as well as the generalization performance of the models. Final testing shows how the trained model will generalize unseen data, displayed in a confusion matrix, along with it's calculated type 1 and type 2 errors.

1 Initial optimization and training of the models

From initial grid search of hyperparameters, led to results that were used as an indication to initial setting of the models in order to raise or maintain performance for the different networks. The results of the grid search is shown in

2 Training optimization based on validation performance

Further optimization was performed based the development in validation validation. These result was used to evaluate the performance of the model, in terms of the under- and overfitting of the training subset. Results for this training optimization is shown as graphs in

2.1 Simple representation training results

Figure 1: Training graphs:

Classes	**Avg. Accuracy (%)**	**Avg. Sensitvity (%)**	**Avg. Specificity (%)**	**Avg. PPV (%)**	**Avg. NPV (%)**
Symptom Duration	54.56% (±12.81%)	50.68% (±0.15%)	59.55% (±0.15%)	58.33% (±0.25%)	51.33% (±0.20%)
Pain intensity	63.33% (±1.67%)	0.00% (±0.00%)	63.33% (±0.02%)	0.00% (±0.00%)	100% (±0.00%)

Table 1: My caption

2.2 Binary representation training results

Figure 2: Training graphs:

Classes	Mean Accuracy (%)	Mean Sensitvity (%)	Mean Specificity (%)	Mean PPV (%)	Mean NPV (%)
Symptom Duration	59.51% (±11.20%)	62.16% (±21.59%)	62.20% (±18.13%)	54.53% (±27.10%)	69.10% (±19.47%)
Pain intensity	65.04% (±10.83%)	48.23% (±0.28%)	71.02% (±0.12%)	44.09% (±0.26%)	78.65% (±0.14%)

Table 2: My caption

2.3 Combined representation training results

Figure 3: Training graphs:

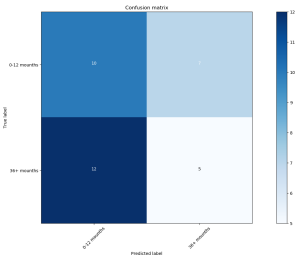
Classes	Mean Accuracy (%)	Mean Sensitvity (%)	Mean Specificity (%)	Mean PPV (%)	Mean NPV (%)
Symptom Duration	55.49% (±10.06%)	55.23% (±16.07%)	56.99% (±12.55%)	42.92% (±21.03%)	68.47% (±13.90%)
Pain intensity	65.14% (±12.87%)	37.50% (±0.35%)	67.34% (±0.15%)	22.14% (±0.20%)	92.41% (±0.07%)

Table 3: My caption

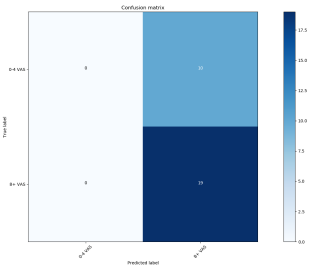
3 Testing results

These result reflect the generalization performance of the three representation models, to which the models classification capabilities to unseen data is shown in confusion matrices, from where a type 1 and type 2 error is calculated.

3.1 Simple representation training results



(a) Confusion matrix for classification of symptom duration



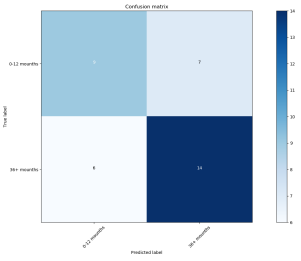
(b) Confusion matrix for classification of pain intensity

Figure 4: Confusion matrices for simple representation testing

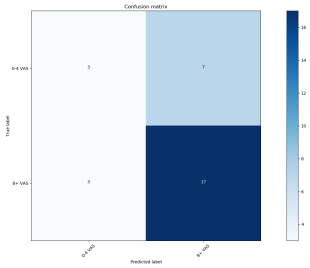
Classes	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Symptom Duration					
Pain intensity					

Table 4: My caption

3.2 Binary representation training results



(a) Confusion matrix for classification of symptom duration



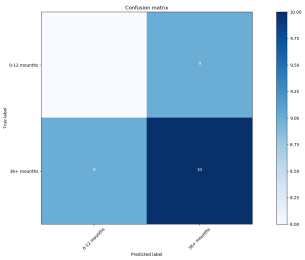
(b) Confusion matrix for classification of pain intensity

Figure 5: Confusion matrices for binary representation testing

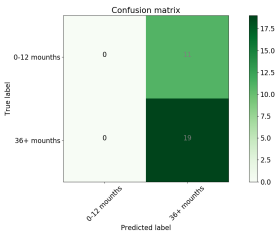
Classes	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Symptom Duration					
Pain intensity					

Table 5: My caption

3.3 Combined representation training results



(a) Confusion matrix for classification of symptom duration



(b) Confusion matrix for classification of pain intensity

Figure 6: Confusion matrices for combined representation testing

Classes	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Symptom Duration					
Pain intensity					

Table 6: My caption