<https://iopscience.iop.org/article/10.1088/1742-6596/892/1/012016/pdf>

<https://iopscience.iop.org/article/10.1088/1742-6596/892/1/012016>

One of the main issues of having small datasets is that of not having a precise prediction model, especially since the information gained in the learning process is fragile and unreliable. Therefore, the prediction uncertainty is largely affected by the lack of information obtained from the small dataset and leads to poor learning performances. This is due to the fact that small datasets are not able to provide enough information, and this leads to gaps between samples, this information gaps make the learning difficult to predict. [?]

~~To obtain a precise prediction model, a large sample set of data is required for the learning process. Otherwise, the model built is considered unreliable and the information gained is fragile~~

~~mall dataset is one of the factors affecting inaccurate prediction model.~~

So, if a small sample is used as training sample of a model, it might significantly affect the prediction uncertainty because of lack of information [2]. Consequently, the knowledge gained from small sample of data is considered unreliable and imprecise for learning system [3]. In addition, in the context of computational learning theory, small sample size gives a major effect to learning performances and is found as one of encountered in machine learning and data mining. This is because insufficient data size of training dataset is liable to poor performances of learning[4].

The main reason why small dataset cannot provide enough information is that gaps are existed between samples. The gaps between sample and observations are called the information gaps, which cause most of the learning tools are difficult to predict

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Autistic Spectrum Disorder Screening Data for Adolescent Data Set** | | | | | | |
| **Neural Network** | | | | | | |
| **Encoder Size** | **Neurons** | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 25 | 28 | 66.67% | 77.55% | 80% | 90% | 97% |
| 30 | 63.77% | 63.27% | 72% | 80% | 98% |
| 30 | 28 | 63.77% | 63.27% | 72% | 80% | 98% |
| 30 | 63.77% | 63.27% | 72% | 80% | 98% |
| **SVM Classifier** | | | | | | |
| **Encoder Size** | Kernel | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 25 | Linear | 94.20% | 89.80% | 97% | 80% | 98% |
| 30 | 97.10% | 91.84% | 96% | 96% | 97% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Breast Cancer Coimbra Data Set** | | | | | | |
| **Neural Network** | | | | | | |
| **Encoder Size** | **Neurons** | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 4 | 28 | 50% | 46.55% | 41.38% | 41.67% | 50% |
| 30 | 50% | 53.45% | 58.62% | 58.33% | 53.45% |
| 5 | 28 | 50% | 53.45% | 58.62% | 58.33% | 50% |
| 30 | 48.78% | 46.55% | 41.38% | 41.67% | 50% |
| **SVM Classifier** | | | | | | |
| **Encoder Size** | Kernel | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 4 | Linear | 54.29% | 63.79% | 44.83% | 41.67% | 66.67% |
| 5 | 60% | 55.17% | 75.86% | 33.33% | 33.33% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Heart Disease Data Set** | | | | | | |
| **Neural Network** | | | | | | |
| **Encoder Size** | **Neurons** | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 6 | 28 | 52.40% | 49.66% | 52% | 53.33% | 66.67% |
| 30 | 52.00% | 50.00% | 52% | 53.00% | 66.67% |
| 7 | 28 | 52.40% | 49.66% | 52.33% | 53.33% | 66.00% |
| 30 | 3% | 3% | 4% | 4% | 4% |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Autistic Spectrum Disorder Screening Data for Adolescent Data Set** | | | | | | |
| **Neural Network** | | | | | | |
| **Encoder Size** | **Neurons** | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 25 | 28 | 66.67% | 77.55% | 80% | 90% | 97% |
| 30 | 63.77% | 63.27% | 72% | 80% | 98% |
| 30 | 28 | 63.77% | 63.27% | 72% | 80% | 98% |
| 30 | 63.77% | 63.27% | 72% | 80% | 98% |
| **SVM Classifier** | | | | | | |
| **Encoder Size** | Kernel | **30% training** | **50% training** | **75% training** | **90% training** | **98% training** |
| 25 | Linear | 94.20% | 89.80% | 97% | 80% | 98% |
| 30 | 97.10% | 91.84% | 96% | 96% | 97% |





