**……**

**GBM-AA/EA-Protein-OS-3YR**

**PRAD-AA/EA-mRNA-PFI-2YR**

**3YR**

**4YR**

**5YR**

**2YR**

**1YR**

**DFI**

**OS**

**DSS**

**PFI**

**Pan-Gyn**

**……**

**GBM**

**……**

**PARD**

**……**

**ACC**

**mRNA Feature**

**Protein Feature**

**Data Preparation**

**TCGA Data of AA and EA**

**Data Preparation**

**2 Feature Types**

**40 Cancer Types**

**Machine Learning input**

**4 Clinical Outcome**

**Endpoints**

**5 Event time thresholds**

**1600 machine learning**

**tasks**

**Model Training**

**Result Comparison**

**Comparison of three Machine Learning schemes**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Machine learning scheme** | **Experiment** | **Training data ethnic composition** | **Testing data ethnic composition** | **AUROC** | |
| **Median** | **Mean** |
| Mixture learning | Mixture 0 | AA+EA | AA+EA | 0.71 | 0.72 |
| Mixture 1 | EA | 0.71 | 0.73 |
| Mixture 2 | AA | 0.68 | 0.67 |
| Independent learning | Independent 1 | EA | EA | 0.70 | 0.71 |
| Independent 2 | AA | AA | 0.59 | 0.58 |
| Transfer learning | Transfer learning | EA (source domain) AA (target domain) | AA | 0.70 | 0.69 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Machine learning scheme** | **Experiment** | **Training data ethnic composition** | **Testing data ethnic composition** | **Training data ethnic composition** | **Testing data ethnic composition** |
| Mixture learning | Mixture 0 | EA+ASA | EA+ASA | EA+NAT-A | EA+NAT-A |
| Mixture 1 | EA | EA |
| Mixture 2 | ASA | NAT-A |
| Independent learning | Independent 1 | EA | EA | EA | EA |
| Independent 2 | ASA | ASA | NAT-A | NAT-A |
| Transfer learning | Transfer learning | EA (source domain) ASA (target domain) | ASA | EA (source domain) NAT-A (target domain) | NAT-A |

**New target domain (ASA/NAT-A)**

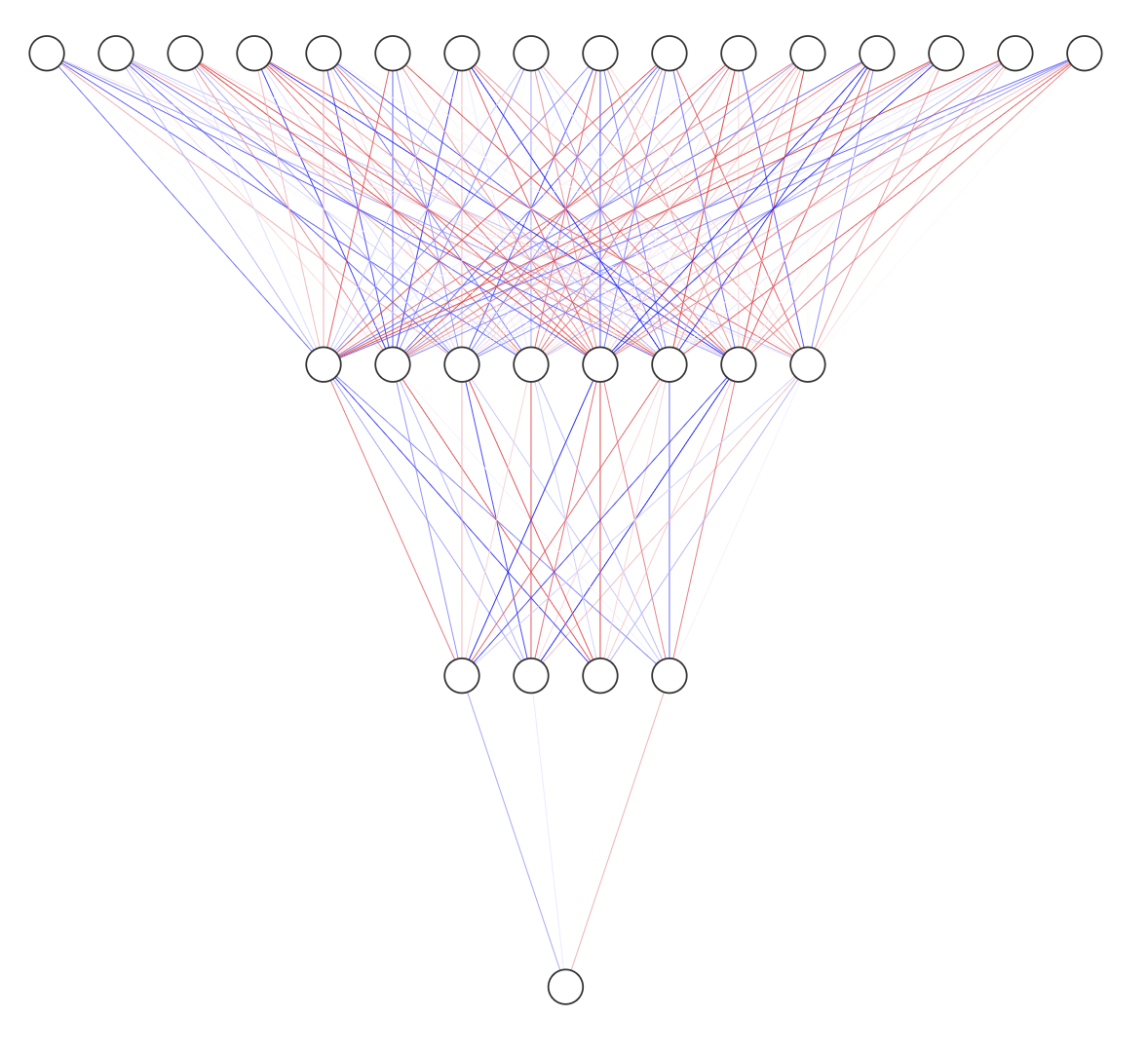
**Fine-tune model**

**Baseline model**

**Original domain**

**Use new feature combine**

**Baseline experiments**



**Input layer (389 nodes)**

**Full connection layer (128 nodes)**

**.**

**.**

**.**

**Dropout layer**

**.**

**. Hidden layer**

**.**

**Full connection layer (64 nodes)**

**.**

**.**

**.**

**Dropout layer**

**Output layer**

|  |  |
| --- | --- |
| **Cancer category** | **Cancer types** |
| GBMLGG | GBM, LGG |
| COADREAD | COAD, READ |
| KIPAN | KIRC, KICH, KIRP |
| STES | ESCA, STAD |
| PanGI | COAD, STAD, READ, ESCA |
| PanGyn | OV, CESC, USC, UCEC |
| PanSCCs | LUSC, HNSC, ESCA, CESC, BLCA |
| PanPan | ACC, BLCA, BRCA, CESC, COAD, DLBC, ESCA, GBM, HNSC, KICH, KIRC, KIRP, LAML, LGG, LIHC, LUAD, LUSC, MESO, OV, PAAD, PCPG, PRAD, READ, SARC, SKCM, STAD, TGCT, THCA, THYM, UCEC, UCS, UVM |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | AUROC | |
| Feature Combine | Data Distribution | Baseline | Fine-tune model |
| SARC-mRNA-OS-2YR | EA->AA | 0.54 | 0.68 |
| BLCA-Protein-PFI-4YR | EA->AA | 0.79 | 0.86 |
| UCEC-Protein-DFI-3YR | EA->AA | 0.59 | 0.63 |
| SARC-mRNA-DSS-3YR | EA->AA | 0.63 | 0.68 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | AUROC | |
| Feature Combine | Data Distribution | Baseline | Fine-tune model |
| BRCA-mRNA-PFI-3YR | EA->ASA | 0.60 | 0.64 |
| STAD-mRNA-OS-1YR | EA->ASA | 0.61 | 0.64 |
| UCEC-mRNA-DSS-4YR | EA->ASA | 0.65 | 0.63 |
| CESE-mRNA-OS-2YR | EA->NAT-A | 0.56 | 0.61 |
| CESE-mRNA-DSS-1YR | EA->NAT-A | 0.44 | 0.51 |
| CESE-mRNA-DSS-2YR | EA->NAT-A | 0.54 | 0.6 |
| PanGyn-mRNA-OS-2YR | EA->NAT-A | 0.42 | 0.51 |