Supplementary: Discriminative Deep Canonical Correlation Analysis for Multi-View Data

Debamita Kumar and Pradipta Maji *

S1 Important Derivations

In this section, the derivations of some of the important equations, reported in the main manuscript, are presented.

S1.1 Derivation of Equation (12)

$$\sum_{\mathbf{h}} Q(\mathbf{h}|\mathbf{v},\mathbf{y}) \ln P(\mathbf{v},\mathbf{h},\mathbf{y}) = \sum_{\mathbf{h}} Q(\mathbf{h}|\mathbf{v},\mathbf{y})\{-E(\mathbf{v},\mathbf{h},\mathbf{y}) - \ln Z\} = \sum_{\mathbf{h}} Q(\mathbf{h}|\mathbf{v},\mathbf{y})\{-E_s(\mathbf{v},\mathbf{h},\mathbf{y}) - E_c(\mathbf{h}) - \ln Z\}$$

$$= \sum_{\mathbf{h}} Q(\mathbf{h}|\mathbf{v},\mathbf{y}) \Bigg\{ \sum_{m=1}^{M}\sum_{i=1}^{V^m}\sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{p=1}^{Y}\sum_{j=1}^{H^{lm}} y_p u_{pj}^{lm} h_j^{lm} + \sum_{l=1}^{L_0-1}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{k=1}^{H^{(l+1)m}} h_j^{lm} w_{jk}^{(l+1)m} h_k^{(l+1)m}$$

$$+ \sum_{m=1}^{M}\sum_{j=1}^{H^{L_0 m}}\sum_{k=1}^{H^{(L_0+1)}} h_j^{L_0 m} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{l=(L_0+1)}^{L}\sum_{p=1}^{Y}\sum_{j=1}^{H^l} y_p u_{pj}^{l} h_j^{l} + \sum_{l=(L_0+1)}^{L-1}\sum_{j=1}^{H^l}\sum_{k=1}^{H^{(l+1)}} h_j^{l} w_{jk}^{(l+1)} h_k^{(l+1)} + \sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}} b_j^{lm} h_j^{lm}$$

$$+ \sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l} b_j^l h_j^l + \sum_{m=1}^{M}\sum_{r=(m+1)}^{M}\sum_{j=1}^{H^{L_0 m}} h_j^{L_0 m} h_j^{L_0 r} - \sum_{m=1}^{M}\sum_{j=1}^{H^{L_0 m}} \lambda_m {h_j^{L_0 m}}^2 \Bigg\} + \sum_{m=1}^{M}\sum_{i=1}^{V^m} a_i^m v_i^m + \sum_{p=1}^{Y} d_p y_p + \sum_{m=1}^{M} \lambda_m - \ln Z$$

$$= \sum_{m=1}^{M}\sum_{i=1}^{V^m}\sum_{j=1}^{H^{1m}}\sum_{h_j^{1m}\in\{0,1\}} q(h_j^{1m}|\mathbf{v},\mathbf{y}) v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{p=1}^{Y}\sum_{j=1}^{H^{lm}}\sum_{h_j^{lm}} q(h_j^{lm}|\mathbf{v},\mathbf{y}) y_p u_{pj}^{lm} h_j^{lm}$$

$$+ \sum_{l=1}^{L_0-1}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{k=1}^{H^{(l+1)m}}\sum_{h_j^{lm}}\sum_{h_k^{(l+1)m}} q(h_j^{lm}|\mathbf{v},\mathbf{y}) q(h_k^{(l+1)m}|\mathbf{v},\mathbf{y}) h_j^{lm} w_{jk}^{(l+1)m} h_k^{(l+1)m}$$

$$+ \sum_{m=1}^{M}\sum_{j=1}^{H^{L_0 m}}\sum_{k=1}^{H^{(L_0+1)}}\sum_{h_j^{L_0 m}}\sum_{h_k^{(L_0+1)}} q(h_j^{L_0 m}|\mathbf{v},\mathbf{y}) q(h_k^{(L_0+1)}|\mathbf{v},\mathbf{y}) h_j^{L_0 m} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{l=(L_0+1)}^{L}\sum_{p=1}^{Y}\sum_{j=1}^{H^l}\sum_{h_j^l} q(h_j^l|\mathbf{v},\mathbf{y}) y_p u_{pj}^l h_j^l$$

$$+ \sum_{l=(L_0+1)}^{L-1}\sum_{j=1}^{H^l}\sum_{k=1}^{H^{(l+1)}}\sum_{h_j^l}\sum_{h_k^{(l+1)}} q(h_j^l|\mathbf{v},\mathbf{y}) q(h_k^{(l+1)}|\mathbf{v},\mathbf{y}) h_j^l w_{jk}^{(l+1)} h_k^{(l+1)} + \sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{h_j^{lm}} q(h_j^{lm}|\mathbf{v},\mathbf{y}) b_j^{lm} h_j^{lm}$$

$$+ \sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}\sum_{h_j^l} q(h_j^l|\mathbf{v},\mathbf{y}) b_j^l h_j^l + \sum_{m=1}^{M}\sum_{r=(m+1)}^{M}\sum_{j=1}^{H^{L_0 m}}\sum_{h_j^{L_0 m}}\sum_{h_j^{L_0 r}} q(h_j^{L_0 m}|\mathbf{v},\mathbf{y}) q(h_j^{L_0 r}|\mathbf{v},\mathbf{y}) h_j^{L_0 m} h_j^{L_0 r}$$

$$- \sum_{m=1}^{M}\sum_{j=1}^{H^{L_0 m}}\sum_{h_j^{L_0 m}} q(h_j^{L_0 m}|\mathbf{v},\mathbf{y}) \lambda_m {h_j^{L_0 m}}^2 + \sum_{m=1}^{M}\sum_{i=1}^{V^m} a_i^m v_i^m + \sum_{p=1}^{Y} d_p y_p + \sum_{m=1}^{M} \lambda_m - \ln Z$$

$$= \sum_{m=1}^{M}\sum_{i=1}^{V^m}\sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} \mu_j^{1m} + \sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{p=1}^{Y}\sum_{j=1}^{H^{lm}} y_p u_{pj}^{lm} \mu_j^{lm} + \sum_{l=1}^{L_0-1}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{k=1}^{H^{(l+1)m}} \mu_j^{lm} w_{jk}^{(l+1)m} \mu_k^{(l+1)m}$$

*The authors are with the Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit, Indian Statistical Institute, Kolkata, India. E-mail: {debamita_r,pmaji}@isical.ac.in. Corresponding author: Pradipta Maji.

$$+\sum_{m=1}^{M}\sum_{j=1}^{H^{L_0m}}\sum_{k=1}^{H^{(L_0+1)}}\mu_j^{L_0m}w_{jk}^{(L_0+1)m}\mu_k^{(L_0+1)}+\sum_{l=(L_0+1)}^{L}\sum_{p=1}^{Y}\sum_{j=1}^{H^l}y_pu_{pj}^l\mu_j^l+\sum_{l=(L_0+1)}^{L-1}\sum_{j=1}^{H^l}\sum_{k=1}^{H^{(l+1)}}\mu_j^lw_{jk}^{(l+1)}\mu_k^{(l+1)}+\sum_{m=1}^{M}\sum_{i=1}^{V^m}a_i^mv_i^m$$

$$+\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}b_j^{lm}\mu_j^{lm}+\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}b_j^l\mu_j^l+\sum_{p=1}^{Y}d_py_p+\sum_{m=1}^{M}\sum_{r=(m+1)}^{M}\sum_{j=1}^{H^{L_0m}}\mu_j^{L_0m}\mu_j^{L_0r}+\sum_{m=1}^{M}\lambda_m\left(1-\sum_{j=1}^{H^{L_0m}}\mu_j^{L_0m}\right)-\ln Z \quad (1)$$

S1.2 Derivation of Equation (13)

$$\begin{aligned}
&\sum_{\mathbf{h}}Q(\mathbf{h}|\mathbf{v},\mathbf{y})\ln Q(\mathbf{h}|\mathbf{v},\mathbf{y})\\
&=\sum_{\mathbf{h}}\left\{\prod_{l=1}^{L_0}\prod_{m=1}^{M}\prod_{j=1}^{H^{lm}}q(h_j^{lm}|\mathbf{v},\mathbf{y})\prod_{l=(L_0+1)}^{L}\prod_{j=1}^{H^l}q(h_j^l|\mathbf{v},\mathbf{y})\left(\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\ln q(h_j^{lm}|\mathbf{v},\mathbf{y})+\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}\ln q(h_j^l|\mathbf{v},\mathbf{y})\right)\right\}\\
&=\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{h_j^{lm}\in\{0,1\}}q(h_j^{lm}|\mathbf{v},\mathbf{y})\ln q(h_j^{lm}|\mathbf{v},\mathbf{y})+\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}\sum_{h_j^l\in\{0,1\}}q(h_j^l|\mathbf{v},\mathbf{y})\ln q(h_j^l|\mathbf{v},\mathbf{y})\\
&=\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\Big\{\mu_j^{lm}\ln\mu_j^{lm}+(1-\mu_j^{lm})\ln(1-\mu_j^{lm})\Big\}+\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}\Big\{\mu_j^l\ln\mu_j^l+(1-\mu_j^l)\ln(1-\mu_j^l)\Big\} \quad (2)
\end{aligned}$$

S1.3 Derivation of Equations (14) and (15)

The variational lower bound of the log-likelihood function, corresponding to the proposed discriminative deep canonical correlation analysis (D2CCA) architecture is given by

$$\begin{aligned}
\mathcal{L}_v=&\sum_{m=1}^{M}\sum_{i=1}^{V^m}\sum_{j=1}^{H^{1m}}v_i^mw_{ij}^{1m}\mu_j^{1m}+\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{p=1}^{Y}\sum_{j=1}^{H^{lm}}y_pu_{pj}^{lm}\mu_j^{lm}+\sum_{l=1}^{L_0-1}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\sum_{k=1}^{H^{(l+1)m}}\mu_j^{lm}w_{jk}^{(l+1)m}\mu_k^{(l+1)m}\\
&+\sum_{m=1}^{M}\sum_{j=1}^{H^{L_0m}}\sum_{k=1}^{H^{(L_0+1)}}\mu_j^{L_0m}w_{jk}^{(L_0+1)m}\mu_k^{(L_0+1)}+\sum_{l=(L_0+1)}^{L}\sum_{p=1}^{Y}\sum_{j=1}^{H^l}y_pu_{pj}^l\mu_j^l+\sum_{l=(L_0+1)}^{L-1}\sum_{j=1}^{H^l}\sum_{k=1}^{H^{(l+1)}}\mu_j^lw_{jk}^{(l+1)}\mu_k^{(l+1)}+\sum_{m=1}^{M}\sum_{i=1}^{V^m}a_i^mv_i^m\\
&+\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}b_j^{lm}\mu_j^{lm}+\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}b_j^l\mu_j^l+\sum_{p=1}^{Y}d_py_p+\sum_{m=1}^{M}\sum_{r=(m+1)}^{M}\sum_{j=1}^{H^{L_0m}}\mu_j^{L_0m}\mu_j^{L_0r}+\sum_{m=1}^{M}\lambda_m\left(1-\sum_{j=1}^{H^{L_0m}}\mu_j^{L_0m}\right)\\
&-\sum_{l=1}^{L_0}\sum_{m=1}^{M}\sum_{j=1}^{H^{lm}}\Big\{\mu_j^{lm}\ln\mu_j^{lm}+(1-\mu_j^{lm})\ln(1-\mu_j^{lm})\Big\}-\sum_{l=(L_0+1)}^{L}\sum_{j=1}^{H^l}\Big\{\mu_j^l\ln\mu_j^l+(1-\mu_j^l)\ln(1-\mu_j^l)\Big\}-\ln Z. \quad (3)
\end{aligned}$$

So, in order to obtain equation (14), the following steps are required to be followed.

$$\begin{aligned}
&\frac{\partial\mathcal{L}_v}{\partial\mu_j^{lm}}=0\\
&\Rightarrow\sum_{k=1}^{H^{(l-1)m}}\mu_k^{(l-1)m}w_{kj}^{lm}+\sum_{k=1}^{H^{(l+1)m}}w_{jk}^{(l+1)m}\mu_k^{(l+1)m}+\sum_{p=1}^{Y}y_pu_{pj}^{lm}+b_j^{lm}-\ln\mu_j^{lm}-1+\ln(1-\mu_j^{lm})+1=0\\
&\Rightarrow\ln\left(\frac{1-\mu_j^{lm}}{\mu_j^{lm}}\right)=-\sum_{k=1}^{H^{(l-1)m}}\mu_k^{(l-1)m}w_{kj}^{lm}-\sum_{k=1}^{H^{(l+1)m}}w_{jk}^{(l+1)m}\mu_k^{(l+1)m}-\sum_{p=1}^{Y}y_pu_{pj}^{lm}-b_j^{lm}\\
&\Rightarrow\frac{1}{\mu_j^{lm}}-1=exp\left\{-\sum_{k=1}^{H^{(l-1)m}}\mu_k^{(l-1)m}w_{kj}^{lm}-\sum_{k=1}^{H^{(l+1)m}}w_{jk}^{(l+1)m}\mu_k^{(l+1)m}-\sum_{p=1}^{Y}y_pu_{pj}^{lm}-b_j^{lm}\right\}\\
&\Rightarrow\mu_j^{lm}=\sigma\left(\sum_{k=1}^{H^{(l-1)m}}\mu_k^{(l-1)m}w_{kj}^{lm}+\sum_{k=1}^{H^{(l+1)m}}w_{jk}^{(l+1)m}\mu_k^{(l+1)m}+\sum_{p=1}^{Y}y_pu_{pj}^{lm}+b_j^{lm}\right), \quad (4)
\end{aligned}$$

where $\sigma(x)=\dfrac{1}{1+e^{-x}}$ is the sigmoid function.

Similarly, in order to obtain equation (15), the following steps are required to be followed.

$$\frac{\partial \mathcal{L}_v}{\partial \mu_j^{L_0 m}} = 0$$

$$\Rightarrow \sum_{k=1}^{H^{(L_0-1)m}} \mu_k^{(L_0-1)m} w_{kj}^{L_0 m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} \mu_k^{L_0+1} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0 m} + b_j^{L_0 m} + \sum_{r\neq m=1}^{M} \mu_j^{L_0 r} - \lambda_m$$

$$- \ln \mu_j^{L_0 m} - 1 + \ln(1 - \mu_j^{L_0 m}) + 1 = 0$$

$$\Rightarrow \ln\left(\frac{1-\mu_j^{L_0 m}}{\mu_j^{L_0 m}}\right) = - \sum_{k=1}^{H^{(L_0-1)m}} \mu_k^{(L_0-1)m} w_{kj}^{L_0 m} - \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} \mu_k^{L_0+1} - \sum_{p=1}^{Y} y_p u_{pj}^{L_0 m} - b_j^{L_0 m} - \sum_{r\neq m=1}^{M} \mu_j^{L_0 r} + \lambda_m$$

$$\Rightarrow \frac{1}{\mu_j^{L_0 m}} - 1 = exp\left\{ - \sum_{k=1}^{H^{(L_0-1)m}} \mu_k^{(L_0-1)m} w_{kj}^{L_0 m} - \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} \mu_k^{L_0+1} - \sum_{p=1}^{Y} y_p u_{pj}^{L_0 m} - b_j^{L_0 m} - \sum_{r\neq m=1}^{M} \mu_j^{L_0 r} + \lambda_m \right\}$$

$$\Rightarrow \mu_j^{L_0 m} = \sigma\left(\sum_{k=1}^{H^{(L_0-1)m}} \mu_k^{(L_0-1)m} w_{kj}^{L_0 m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} \mu_k^{L_0+1} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0 m} + b_j^{L_0 m} + \sum_{r\neq m=1}^{M} \mu_j^{L_0 r} - \lambda_m \right). \tag{5}$$

S1.4 Derivation of Equation (18)

$$P(\mathbf{h}^{1m}|\mathbf{v}^m, \mathbf{y}, \mathbf{h}^{2m}) = \frac{P(\mathbf{h}^{1m}, \mathbf{v}^m, \mathbf{y}, \mathbf{h}^{2m})}{\sum_{\mathbf{h}^{1m}} P(\mathbf{h}^{1m}, \mathbf{v}^m, \mathbf{y}, \mathbf{h}^{2m})} = \frac{\sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} P(\mathbf{h}^{1m}, \mathbf{v}^m, \mathbf{y}, \mathbf{h}^{2m})}{\sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} \sum_{\mathbf{h}^{1m}} P(\mathbf{h}^{1m}, \mathbf{v}^m, \mathbf{y}, \mathbf{h}^{2m})}$$

$$= \frac{\left(\frac{1}{Z}\right) \sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} e^{\left\{ \sum_{i=1}^{V^m} \sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{p=1}^{Y} \sum_{j=1}^{H^{1m}} y_p u_{pj}^{1m} h_j^{1m} + \sum_{j=1}^{H^{1m}} \sum_{k=1}^{H^{2m}} h_j^{1m} w_{jk}^{2m} h_k^{2m} + \sum_{j=1}^{H^{1m}} h_j^{1m} b_j^{1m} + X_{-\mathbf{h}^{1m}} \right\}}}{\left(\frac{1}{Z}\right) \sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} \sum_{\mathbf{h}^{1m}} e^{\left\{ \sum_{i=1}^{V^m} \sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{p=1}^{Y} \sum_{j=1}^{H^{1m}} y_p u_{pj}^{1m} h_j^{1m} + \sum_{j=1}^{H^{1m}} \sum_{k=1}^{H^{2m}} h_j^{1m} w_{jk}^{2m} h_k^{2m} + \sum_{j=1}^{H^{1m}} h_j^{1m} b_j^{1m} + X_{-\mathbf{h}^{1m}} \right\}}}$$

$$= \frac{\left(\sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} e^{(X_{-\mathbf{h}^{1m}})} \right) e^{\left\{ \sum_{i=1}^{V^m} \sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{p=1}^{Y} \sum_{j=1}^{H^{1m}} y_p u_{pj}^{1m} h_j^{1m} + \sum_{j=1}^{H^{1m}} \sum_{k=1}^{H^{2m}} h_j^{1m} w_{jk}^{2m} h_k^{2m} + \sum_{j=1}^{H^{1m}} h_j^{1m} b_j^{1m} \right\}}}{\left(\sum_{\mathbf{v}^{-m}, \mathbf{h}^{-1m}, \mathbf{h}^{-2m}, \mathbf{h}^l} e^{(X_{-\mathbf{h}^{1m}})} \right) \sum_{\mathbf{h}^{1m}} e^{\left\{ \sum_{i=1}^{V^m} \sum_{j=1}^{H^{1m}} v_i^m w_{ij}^{1m} h_j^{1m} + \sum_{p=1}^{Y} \sum_{j=1}^{H^{1m}} y_p u_{pj}^{1m} h_j^{1m} + \sum_{j=1}^{H^{1m}} \sum_{k=1}^{H^{2m}} h_j^{1m} w_{jk}^{2m} h_k^{2m} + \sum_{j=1}^{H^{1m}} h_j^{1m} b_j^{1m} \right\}}}$$

$$= \frac{e^{\left\{ \sum_{j=1}^{H^{1m}} h_j^{1m} \left(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m} \right) \right\}}}{\sum_{\mathbf{h}^{1m}} e^{\left\{ \sum_{j=1}^{H^{1m}} h_j^{1m} \left(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m} \right) \right\}}} = \frac{\prod_{j=1}^{H^{1m}} e^{h_j^{1m} \left(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m} \right)}}{\prod_{j=1}^{H^{1m}} \sum_{\mathbf{h}^{1m}} e^{h_j^{1m} \left(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m} \right)}} \tag{6}$$

Here, $X_{-\mathbf{h}^{1m}} = \sum_{l=2}^{L_0} \sum_{m=1}^{M} \sum_{p=1}^{Y} \sum_{j=1}^{H^{lm}} y_p u_{pj}^{lm} h_j^{lm} + \sum_{l=2}^{L_0-1} \sum_{m=1}^{M} \sum_{j=1}^{H^{lm}} \sum_{k=1}^{H^{(l+1)m}} h_j^{lm} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{m=1}^{M} \sum_{j=1}^{H^{L_0 m}} \sum_{k=1}^{H^{(L_0+1)}} h_j^{L_0 m} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)}$

$$+ \sum_{l=(L_0+1)}^{L} \sum_{p=1}^{Y} \sum_{j=1}^{H^l} y_p u_{pj}^l h_j^l + \sum_{l=(L_0+1)}^{L-1} \sum_{j=1}^{H^l} \sum_{k=1}^{H^{(l+1)}} h_j^l w_{jk}^{(l+1)} h_k^{(l+1)} + \sum_{l=2}^{L_0} \sum_{m=1}^{M} \sum_{j=1}^{H^{lm}} b_j^{lm} h_j^{lm} + \sum_{l=(L_0+1)}^{L} \sum_{j=1}^{H^l} b_j^l h_j^l + \sum_{m=1}^{M} \sum_{i=1}^{V^m} a_i^m v_i^m$$

$$+ \sum_{p=1}^{Y} d_p y_p + \sum_{m=1}^{M} \sum_{r=(m+1)}^{M} \sum_{j=1}^{H^{L_0 m}} h_j^{L_0 m} h_j^{L_0 r} + \sum_{m=1}^{M} \lambda_m \left(1 - \sum_{j=1}^{H^{L_0 m}} (h_j^{L_0 m})^2 \right) - \ln Z.$$

S1.5 Derivation of Equation (19)

From (6), we get, $P(\mathbf{h}^{1m}|\mathbf{v}^m,\mathbf{y},\mathbf{h}^{2m}) = \dfrac{\prod_{j=1}^{H^{1m}} e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}{\prod_{j=1}^{H^{1m}} \sum_{\mathbf{h}^{1m}} e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}.$

Since, the hidden nodes in a particular layer is independent of each other, we can write

$$\prod_{j=1}^{H^{1m}} P(h_j^{1m}|\mathbf{v}^m,\mathbf{y},\mathbf{h}^{2m}) = \frac{\prod_{j=1}^{H^{1m}} e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}{\prod_{j=1}^{H^{1m}} \sum_{\mathbf{h}^{1m}} e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}.$$

Hence, $P(h_j^{1m}|\mathbf{v}^m,\mathbf{y},\mathbf{h}^{2m}) = \dfrac{e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}{\sum_{\mathbf{h}^{1m}} e^{h_j^{1m}(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m})}}$

$$= \frac{e^{\left\{\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m}\right\}}}{1 + e^{\left\{\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m}\right\}}} = \sigma\left(\sum_{i=1}^{V^m} v_i^m w_{ij}^{1m} + \sum_{p=1}^{Y} y_p u_{pj}^{1m} + \sum_{k=1}^{H^{2m}} w_{jk}^{2m} h_k^{2m} + b_j^{1m}\right). \qquad (7)$$

S1.6 Derivation of Equation (20)

$$P(h_j^{lm}|\mathbf{h}^{(l-1)m},\mathbf{h}^{(l+1)m},\mathbf{y}) = \frac{e^{h_j^{lm}\left(\sum_{k=1}^{H^{(l-1)m}} h_k^{(l-1)m} w_{kj}^{lm} + \sum_{k=1}^{H^{(l+1)m}} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{p=1}^{Y} y_p u_{pj}^{lm} + b_j^{lm}\right)}}{\sum_{\mathbf{h}^{lm}} e^{h_j^{lm}\left(\sum_{k=1}^{H^{(l-1)m}} h_k^{(l-1)m} w_{kj}^{lm} + \sum_{k=1}^{H^{(l+1)m}} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{p=1}^{Y} y_p u_{pj}^{lm} + b_j^{lm}\right)}}$$

$$= \frac{e^{\left\{\sum_{k=1}^{H^{(l-1)m}} h_k^{(l-1)m} w_{kj}^{lm} + \sum_{k=1}^{H^{(l+1)m}} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{p=1}^{Y} y_p u_{pj}^{lm} + b_j^{lm}\right\}}}{1 + e^{\left\{\sum_{k=1}^{H^{(l-1)m}} h_k^{(l-1)m} w_{kj}^{lm} + \sum_{k=1}^{H^{(l+1)m}} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{p=1}^{Y} y_p u_{pj}^{lm} + b_j^{lm}\right\}}}$$

$$= \sigma\left(\sum_{k=1}^{H^{(l-1)m}} h_k^{(l-1)m} w_{kj}^{lm} + \sum_{k=1}^{H^{(l+1)m}} w_{jk}^{(l+1)m} h_k^{(l+1)m} + \sum_{p=1}^{Y} y_p u_{pj}^{lm} + b_j^{lm}\right) \qquad (8)$$

S1.7 Derivation of Equation (21)

$$P(h_j^{L_0m}|\mathbf{h}^{(L_0-1)m},\mathbf{h}^{L_0+1},\mathbf{h}^{L_0r},\mathbf{y}) = \frac{e^{h_j^{L_0m}\left(\sum_{k=1}^{H^{(L_0-1)m}} h_k^{(L_0-1)m} w_{kj}^{L_0m} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0m} + b_j^{L_0m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{r\neq m=1}^{M} h_j^{L_0r} - \lambda_m\right)}}{\sum_{\mathbf{h}^{L_0m}} e^{h_j^{L_0m}\left(\sum_{k=1}^{H^{(L_0-1)m}} h_k^{(L_0-1)m} w_{kj}^{L_0m} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0m} + b_j^{L_0m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{r\neq m=1}^{M} h_j^{L_0r} - \lambda_m\right)}}$$

$$= \frac{e^{\left\{\sum_{k=1}^{H^{(L_0-1)m}} h_k^{(L_0-1)m} w_{kj}^{L_0m} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0m} + b_j^{L_0m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{r\neq m=1}^{M} h_j^{L_0r} - \lambda_m\right\}}}{1 + e^{\left\{\sum_{k=1}^{H^{(L_0-1)m}} h_k^{(L_0-1)m} w_{kj}^{L_0m} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0m} + b_j^{L_0m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{r\neq m=1}^{M} h_j^{L_0r} - \lambda_m\right\}}}$$

$$= \sigma\left(\sum_{k=1}^{H^{(L_0-1)m}} h_k^{(L_0-1)m} w_{kj}^{L_0m} + \sum_{p=1}^{Y} y_p u_{pj}^{L_0m} + b_j^{L_0m} + \sum_{k=1}^{H^{(L_0+1)}} w_{jk}^{(L_0+1)m} h_k^{(L_0+1)} + \sum_{r\neq m=1}^{M} h_j^{L_0r} - \lambda_m\right) \qquad (9)$$

S2 Learning Algorithm of Proposed D2CCA Model

The learning algorithm of the proposed D2CCA model is outlined in Algorithm 1.

Algorithm 1: Learning of D2CCA Architecture.

Input: Set of N training data vectors $\{\mathbf{v}^m\}_{n=1}^N$ for M modalities along with the corresponding set of class labels $\{\mathbf{y}\}_{n=1}^N$, number of persistent chains (S), number of epochs (τ), learning rate (η), weight decay (ρ), momentum (ζ), and number of Gibbs steps (α).
Output: Final parameter set $\boldsymbol{\theta}^\tau$ of the architecture.

Perform greedy layer-wise pretraining to initialize the set of model parameters, $\boldsymbol{\theta}^0$.
Randomly initialize S Markov chains $\{\tilde{\mathbf{v}}^{m^0}, \tilde{\mathbf{y}}^0, \tilde{\mathbf{h}}^{1m^0}, \cdots, \tilde{\mathbf{h}}^{L_0 m^0}, \tilde{\mathbf{h}}^{(L_0+1)^0}, \cdots, \tilde{\mathbf{h}}^{L^0}\}_{s=1}^S$.
for each epoch $t = 0$ to τ **do**
    // Variational inference
    for each training sample $n = 1$ to N **do**
        (i) Run mean field updates using (14) - (16) until convergence.
        (ii) Save the obtained mean field parameter ($\boldsymbol{\mu}$) for the corresponding training sample, $\boldsymbol{\mu}_n = \boldsymbol{\mu}$.
    end for
    // Stochastic approximation
    for each persistent chain $s = 1$ to S **do**
        Run the chain for α-steps and sample the state $\{\tilde{\mathbf{v}}^{m^{t+1}}, \tilde{\mathbf{y}}^{t+1}, \tilde{\mathbf{h}}^{1m^{t+1}}, \cdots, \tilde{\mathbf{h}}^{L_0 m^{t+1}}, \tilde{\mathbf{h}}^{(L_0+1)^{t+1}}, \cdots, \tilde{\mathbf{h}}^{L^{t+1}}\}$ from $\{\tilde{\mathbf{v}}^{m^t}, \tilde{\mathbf{y}}^t, \tilde{\mathbf{h}}^{1m^t}, \cdots, \tilde{\mathbf{h}}^{L_0 m^t}, \tilde{\mathbf{h}}^{(L_0+1)^t}, \cdots, \tilde{\mathbf{h}}^{L^t}\}$ using (19) - (25).
    end for
    Update the parameters of the model from $\boldsymbol{\theta}^t$ to $\boldsymbol{\theta}^{(t+1)}$ using (27).
end for

S3 Performance Analysis for Pair of Modalities

In the existing literature, several approaches are found which compute linear canonical correlation analysis (CCA) on non-linear transformations of the given input pair of views. In this section, the performance of these existing methods is compared with that of the proposed D2CCA architecture on the five real-life cancer data sets. Four representative pairs of modalities are chosen for each of the data sets and given as input to both state-of-the-art methods as well as the proposed architecture. The state-of-the-art approaches include randomized canonical correlation analysis (rCCA) [2], deep canonical correlation analysis (DCCA) [1], and deep canonically correlated autoencoders (DCCAE) [3]. While rCCA is a non-deep approach, both DCCA and DCCAE are based on deep architectures. In case of rCCA and DCCA, 50 features are extracted at final layer and 10 features are extracted at output layer for DCCAE, as suggested in the respective papers. The obtained features are then applied to the input of the SVM for classification purpose. However, in case of the proposed architecture, 10 features are considered for the joint representation, and the subtypes of the carcinoma samples are predicted from the architecture itself. So, the cancer subtypes are identified with the help of the given input pair of views only and the corresponding results are reported in Table S1 for both training-testing and 10-fold CV.

From the results reported in Table S1, it can be observed that the existing methods can efficiently identify subtypes of carcinoma samples for certain pairs of views, but fail to provide similar results for some other pairs. For example, the existing methods can identify over 90% of the samples correctly for input views mDNA and microRNA on KIDNEY data set, but they have assigned all the kidney carcinoma samples to one particular subtype for the input pair Protein and CNS. Hence, the performance of these methods depends on the given input pair of modalities. However, the proposed method is not only able to achieve similar results for the given input pairs of views, but also attain highest classification accuracy on all the five data sets for both training-testing and 10-fold CV. Statistical significance analysis demonstrates that out of total 120 cases, the proposed approach achieves significantly better p-values for 110 cases, and better but not significant p-values for the rest 10 cases.

S4 Statistical Significance Analysis of MDDBM architecture

In the current study, multimodal discriminative deep Boltzmann machine (MDDBM) is proposed by incorporating supervised information of sample categories into exiting multimodal deep Boltzmann machine (MDBM). In this section, the effectiveness of the proposed MDDBM architecture is studied with reference to MDBM approach on five real-life cancer data sets.

Table S1: Comparative Performance Analysis for Pair of Modalities on Omics Data Sets

Data Sets	Different Metrics		rCCA	DCCA	DCCAE	D2CCA	rCCA	DCCA	DCCAE	D2CCA	rCCA	DCCA	DCCAE	D2CCA	rCCA	DCCA	DCCAE	D2CCA
			mDNA and microRNA				Protein and microRNA				Protein and RNA				microRNA and RNA			
CESC	Train-Test		44.23	46.15	51.92	**65.38**	51.92	44.23	44.23	**57.69**	51.92	44.23	44.23	**63.46**	42.31	42.31	46.15	**65.38**
	10-fold CV	Mean	50.00	49.17	53.33	**69.17**	53.33	49.17	50.83	**66.67**	39.17	44.17	44.17	**70.00**	46.67	46.67	45.83	**59.17**
		Median	50.00	50.00	50.00	**66.67**	50.00	50.00	58.33	**66.67**	41.67	45.83	41.67	**70.83**	50.00	50.00	50.00	**58.33**
		StdDev	**3.93**	7.30	8.96	13.64	9.78	8.29	15.44	**7.86**	**5.62**	9.66	6.86	9.78	9.78	9.78	10.58	**2.64**
		Paired-t:p	1.04E-03	2.96E-03	2.82E-03	-	2.28E-03	4.57E-04	6.88E-03	-	1.31E-05	9.19E-06	1.15E-04	-	1.50E-03	1.50E-03	1.56E-03	-
		Wilcoxon:p	2.46E-03	2.45E-03	8.23E-03	-	1.08E-02	2.52E-03	8.13E-03	-	2.45E-03	2.53E-03	2.46E-03	-	2.32E-03	2.32E-03	1.03E-02	-
			mDNA and microRNA				Protein and microRNA				Protein and RNA				microRNA and RNA			
CRC	Train-Test		73.85	73.85	73.85	**77.69**	73.85	73.85	73.85	**78.46**	73.85	73.85	73.85	**75.38**	73.85	73.85	73.85	**80.00**
	10-fold CV	Mean	74.07	74.07	74.07	**82.22**	74.07	74.07	74.07	**81.85**	74.07	74.07	74.07	**77.41**	74.07	74.07	74.07	**81.11**
		Median	74.07	74.07	74.07	**81.48**	74.07	74.07	74.07	**81.48**	74.07	74.07	74.07	**77.78**	74.07	74.07	74.07	**81.48**
		StdDev	0.00	0.00	0.00	**5.47**	0.00	0.00	0.00	**3.68**	0.00	0.00	0.00	**2.10**	0.00	0.00	0.00	**5.64**
		Paired-t:p	5.49E-04	5.49E-04	5.49E-04	-	4.54E-05	4.54E-05	4.54E-05	-	3.63E-04	3.63E-04	3.63E-04	-	1.70E-03	1.70E-03	1.70E-03	-
		Wilcoxon:p	3.26E-03	3.26E-03	3.26E-03	-	3.10E-03	3.10E-03	3.10E-03	-	4.73E-03	4.73E-03	4.73E-03	-	6.06E-03	6.06E-03	6.06E-03	-
			mDNA and microRNA				Protein and CNS				microRNA and RNA				RNA and CNS			
KIDNEY	Train-Test		96.71	96.05	94.74	**98.03**	69.08	69.08	69.08	**73.03**	**98.03**	94.08	94.08	**98.03**	69.08	88.82	88.82	**97.37**
	10-fold CV	Mean	97.42	95.81	97.10	**99.03**	67.74	67.74	67.74	**81.94**	97.42	95.48	90.00	**98.71**	67.74	69.35	67.74	**81.61**
		Median	96.77	96.77	96.77	**100.00**	67.74	67.74	67.74	**80.65**	96.77	96.77	91.94	**100.00**	67.74	67.74	67.74	**77.42**
		StdDev	2.54	2.66	2.82	**1.56**	0.00	0.00	0.00	**4.61**	2.54	3.47	6.17	**1.67**	0.00	**4.09**	0.00	12.64
		Paired-t:p	7.48E-03	4.23E-03	4.06E-02	-	2.24E-06	2.24E-06	2.24E-06	-	5.19E-02	1.15E-02	1.45E-03	-	3.52E-03	9.49E-03	3.52E-03	-
		Wilcoxon:p	2.17E-02	9.83E-03	8.74E-02	-	2.38E-03	2.38E-03	2.38E-03	-	1.70E-01	9.78E-03	2.47E-03	-	2.47E-03	8.23E-03	2.47E-03	-
			mDNA and microRNA				Protein and CNS				microRNA and RNA				RNA and CNS			
LGG	Train-Test		53.23	62.90	64.52	**74.73**	48.39	48.39	48.39	**67.74**	48.39	48.39	48.39	**65.59**	48.39	45.16	48.39	**65.05**
	10-fold CV	Mean	47.37	66.32	75.79	**77.89**	47.37	47.37	49.74	**72.11**	47.37	47.37	47.37	**50.26**	47.37	47.37	47.37	**48.42**
		Median	47.37	65.79	**77.63**	**77.63**	47.37	47.37	51.32	**72.37**	47.37	47.37	47.37	**50.00**	**47.37**	**47.37**	**47.37**	**47.37**
		StdDev	0.00	8.30	12.33	**4.51**	0.00	**2.15**	4.20	3.96	0.00	0.00	0.00	**3.81**	0.00	0.00	0.00	**1.36**
		Paired-t:p	2.48E-09	2.30E-03	3.02E-01	-	5.09E-09	7.68E-09	2.97E-07	-	1.99E-02	1.99E-02	1.99E-02	-	1.84E-02	1.84E-02	1.84E-02	-
		Wilcoxon:p	2.36E-03	1.41E-02	2.88E-01	-	2.45E-03	2.50E-03	2.52E-03	-	7.88E-03	7.88E-03	7.88E-03	-	1.93E-03	1.93E-03	1.93E-03	-
			mDNA and microRNA				Protein and CNS				microRNA and RNA				RNA and CNS			
LUNG	Train-Test		93.41	94.14	**94.87**	**94.87**	57.14	57.14	57.14	**60.07**	94.14	87.55	75.09	**94.87**	57.14	94.14	57.14	**95.24**
	10-fold CV	Mean	94.82	93.75	94.11	**96.43**	57.14	57.14	56.96	**66.96**	95.18	89.64	70.18	**96.07**	57.14	68.39	69.46	**82.14**
		Median	**96.43**	94.64	95.54	**96.43**	57.14	57.14	57.14	**66.96**	95.54	89.29	60.71	**97.32**	57.14	57.14	60.71	**93.75**
		StdDev	3.81	4.23	4.13	**2.79**	0.00	0.00	**0.56**	4.39	**2.53**	5.82	16.30	3.35	0.00	18.35	**14.57**	19.32
		Paired-t:p	9.36E-03	8.71E-04	3.52E-02	-	2.94E-05	2.94E-05	2.92E-05	-	1.06E-01	1.26E-04	3.03E-04	-	1.36E-03	6.12E-02	7.25E-02	-
		Wilcoxon:p	2.50E-02	3.38E-03	1.06E-01	-	2.49E-03	2.49E-03	2.50E-03	-	1.93E-01	3.79E-03	3.44E-03	-	2.45E-03	2.96E-02	4.62E-02	-

Table S2: Effectiveness of MDDBM Architecture on Omics Data Sets

Data Sets	MDBM	
	Paired-t:p	Wilcoxon:p
CESC	3.87E-02	2.34E-02
CRC	7.35E-02	1.20E-01
KIDNEY	1.19E-04	3.68E-03
LGG	2.38E-06	2.45E-03
LUNG	1.77E-03	3.46E-03

Two statistical significance tests are performed and the corresponding p-values, computed using paired-t (one-tailed) and Wilcoxon signed-rank (one-tailed) tests, with 95% confidence level, are presented in Table S2. Statistical significance analysis demonstrates that out of total 10 cases, the proposed MDDBM architecture achieves significantly better p-values for 8 cases and better but not significant p-values in remaining 2 cases. Thus, from the results reported in Table S2, it can be observed that incorporating supervised information in the hidden representations improves discriminative ability of the MDDBM architecture as compared to the unsupervised counterpart.

References

[1] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep Canonical Correlation Analysis, 2013.

[2] D. Lopez-Paz, S. Sra, A. Smola, Z. Ghahramani, and B. Schölkopf. Randomized Nonlinear Component Analysis, 2014.

[3] W. Wang, R. Arora, K. Livescu, and J. Bilmes. On Deep Multi-View Representation Learning, 2015.