Credibility-aware Reliable Multi-Modal Fusion Using Probabilistic Circuits

Saurabh Mathur[1]*, Sahil Sidheekh[1]*, Pranuthi Tenali[1]*, Erik Blasch[2], Kristian Kersting[3], Sriraam Natarajan[1]

[1]The University of Texas at Dallas
[2] Air Force Research Laboratory
[3] TU Darmstadt and Hessian Center for AI (hessian.AI)
SaurabhSanjay.Mathur@utdallas.edu, Sahil.Sidheekh@utdallas.edu, Pranuthi.Tenali@utdallas.edu, erik.blasch.1@us.af.mil, kersting@cs.tu-darmstadt.de, Sriraam.Natarajan@utdallas.edu

Abstract

We consider the problem of late multi-modal fusion for discriminative learning. Motivated by multi-source domains that require understanding the reliability of each data source, we explore the notion of credibility in the context of multi-modal fusion. We propose a combination function that uses probabilistic circuits (PCs) to combine predictive distributions over individual modalities. We also define a probabilistic measure to evaluate the credibility of each modality via inference queries over the PC. Our experimental evaluation demonstrates that our fusion method can reliably infer credibility while maintaining competitive performance with the state of the art.

Introduction

Real-world decision-making requires reasoning reliably by utilizing the diverse modalities of data sources that are available. While such multi-modal data offers a rich representation and potentially multiple views of the underlying phenomena (for example, images vs blood tests in a clinical setting), it also makes learning and inference more challenging. Raw data from different sources is often noisy, incomplete, and inconsistent. This heterogeneity poses a significant obstacle to effective data fusion and analysis.

Multi-modal fusion techniques (Baltrušaitis, Ahuja, and Morency 2018) have emerged as a promising approach to combine information from multiple sources to enhance performance on discriminative learning tasks. These techniques aim to extract and integrate complementary information from different modalities, leading to more robust and reliable results. However, a crucial aspect that often remains overlooked in multimodal fusion is *explicit modeling of the credibility* of the information sources. In many applications, such as sensor fusion (Khaleghi et al. 2013), medical diagnosis (Kline et al. 2022), and financial analysis (Sawhney et al. 2020), the quality and reliability of the information sources vary significantly. Distinguishing reliable sources from non-reliable sources is essential for making accurate and informed decisions. Multimodal fusion methods often assume that all sources are equally credible, which can lead to suboptimal performance or even erroneous conclusions.

Credibility aware methods in the context of late multi-modal fusion have previously used weighted average (Rogova and Nimier 2004), discounting factors (Elouedi, Mellouli, and Smets 2004a) and Bayesian networks (Wright and Laskey 2006). This results in models of credibility that are either too simple (as in the case of weighted averages and discounting factors) to model complex dependencies or too complex to perform tractable inference (as in the case of Bayesian networks). We focus on **multi-modal discriminative learning and propose a late fusion method that uses Probabilistic Circuits (PCs)** (Choi, Vergari, and Van den Broeck 2020), to effectively combine the predictive distributions over individual modalities. PCs are a class of generative models that are expressive enough to model complex distributions while tractable for exact inference. Using the tractability of PCs, we define a probabilistic measure for assessing the credibility. We also experimentally validate the efficacy of PCs in modeling complex interactions between modalities and reliably estimating their credibility.

We will begin with a concise overview of essential background and relevant works. Following this, we formulate the problem at hand and our PC based fusion method, along with the architectural details and methodology for assessing credibility. We then experimentally evaluate the effectiveness of our method and finally conclude by summarizing our findings, contributions and future work.

Background

Multi-modal fusion Multi-modal fusion (Baltrušaitis, Ahuja, and Morency 2018) is the integration of information from diverse sources or modalities. This field harnesses the potential of combining data of various types, like text, images, and audio, to improve decision-making, pattern recognition, and predictive modeling. There are two broad approaches to multi-modal fusion in the discriminative learning setting, namely, early fusion and late fusion.

Early fusion approaches fuse information from multiple sources by combining the features before making predictions. A simple way to achieve this would be to combine raw modality features via concatenation or pooling via operations such as average, min, max, etc. (Baltrušaitis, Ahuja,

*These authors contributed equally.

and Morency 2018). In more complex deep learning models, early fusion is typically achieved by learning joint feature spaces (Gadzicki, Khamsehashari, and Zetzsche 2020). Apart from the curse of dimensionality, feature aggregation results in the loss of information about source-specific distributions (Schulte and Routley 2014). This makes it difficult to infer the credibility of input sources.

On the other hand, late fusion approaches combine the information from multiple sources by making predictions on each source and then combining the predictions. Combining rules (Natarajan et al. 2005; Manhaeve et al. 2018) like weighted mean (Shutova, Kiela, and Maillard 2016) and Noisy-OR (Tian et al. 2020) are commonly used for late fusion. While these combining rules allow explicit modeling of the credibility of each source, they assume independence of the influence of each source on the target. Late fusion in deep learning models is implemented via additional feedforward layers (Glodek et al. 2011; Ramirez, Baltrušaitis, and Morency 2011). This allows them to model complex correlations and influences of the sources on the target. However, this also makes it difficult to model the credibility of each source since neural network layers are opaque.

Credibility Combining information from multiple, heterogeneous sources requires information fusion systems to account for the credibility of each modality's contribution (De Villiers et al. 2018). Credibility, as distinct from reliability, deals with the information's truthfulness, while reliability relates to the source's consistency (Blasch et al. 2013). While human experts might estimate their information's credibility (self-confidence), automated sources require external evaluation (Blasch et al. 2014). We approach the problem of accounting for source reliability in multimodal fusion from the perspective of the credibility of the information provided by the source. Prior works have used source-reliability coefficients learned using domain and contextual information (Nimier 1998; Fabre, Appriou, and Briottet 2001). In the absence of such information, an alternate approach involves learning these coefficients from training data. This is achieved by minimizing the distance between a vector of beliefs resulting from fusion and a target vector from the training set (Rogova and Kasturi 2001; Elouedi, Mellouli, and Smets 2004b). Another method for establishing reliability, by using training data, is based on *separatability*, wherein the average statistical separability of information classes in each source is considered (Benediktsson, Swain, and Ersoy 1990). This category of methods i.e. learning coefficients from training data, proves useful in establishing the relative reliability of classifiers.

Probabilistic circuits (PCs) (Choi, Vergari, and Van den Broeck 2020) are a class of generative models that represent the joint distribution over a set of random variables (say $\mathbf{X}$) using computational graphs that comprise sum and product nodes as internal nodes, and simple tractable distributions at the leaves. Formally, a PC $\mathcal{M}$ is defined as the tuple $(G = (V, E), \theta)$ where the Directed Acyclic Graph G represents the computational graph structure and θ is the set of learnable parameters. The distribution induced by the PC M having root node n is given as

$$P_n(\mathbf{X}=\mathbf{x}) = \begin{cases} \sum_{c\in\mathbf{ch}(n)} w_c P_c(\mathbf{X}=\mathbf{x}) & n \in \text{Sum} \\ \prod_{c\in\mathbf{ch}(n)} P_c(\mathbf{X}_{\mathbf{sc}(c)} = \mathbf{x}_{\mathbf{sc}(c)}) & n \in \text{ Product} \\ \psi_n(\mathbf{X}=\mathbf{x}) & n \in \text{Leaf} \end{cases}$$

where $\mathbf{ch}(n)$ gives the children of node n, $\mathbf{sc}(n)$ gives the scope of node n and ψ_n is the probability density (or mass) function associated with the leaf node n.

The key advantage of PCs is that they admit tractable and often linear time inference for a variety of probabilistic queries under mild assumptions about the structure of G. In this work, we consider a subclass of PCs that are *smooth* and *decomposable* (typically called sum-product networks (Poon and Domingos 2011)). A PC satisfies smoothness if the scope of each sum node is identical to the scope of each of its children. It satisfies decomposability if, for each product node, all the children have disjoint scopes. Smoothness and decomposability allow us to tractably infer marginal and conditional distributions from the learned joint.

The structure of PCs can be learned recursively via greedy heuristics (Gens and Pedro 2013; Rooshenas and Lowd 2014; Dang, Vergari, and Van den Broeck 2020), or by latent-space decomposition (Adel, Balduzzi, and Ghodsi 2015). However, structure learning can be costly for large-scale data, and recent approaches rely on random and tensorized structures that resemble deep neural models (Mauro et al. 2017; Peharz et al. 2020a,b; Sidheekh, Kersting, and Natarajan 2023) to achieve state-of-the-art performance.

Methodology

We focus on the multi-modal discriminative learning setting, where multiple experts are trained for each modality, and their predictive distributions are combined using a function f to obtain the final output.

Given: A dataset $\mathcal{D} = \{(\mathbf{x}_1^i, \mathbf{x}_2^i \ldots \mathbf{x}_m^i, y^i)\}_{i=1}^N$ comprising N data points, each with information from m different modalities, i.e. each $\mathbf{x}_j^i \in \mathbb{R}^{d_j}$ where d_j denotes the feature dimension corresponding to modality j.

To do: Learn a discriminative model $\mathcal{M}$ parameterized by $\{\theta, \phi = \{\phi_i\}_{i=1}^m\}$ that approximates the multimodal predictive distribution as

$$\begin{aligned} P(Y|\mathbf{X}_1, \ldots, \mathbf{X}_m) &\approx \mathcal{M}_{\theta,\phi}(\mathbf{X}_1, \ldots, \mathbf{X}_m) \\ &= \mathcal{M}_\theta(\mathcal{M}_{\phi_1}(\mathbf{X}_1), \ldots, \mathcal{M}_{\phi_m}(\mathbf{X}_m)) \end{aligned}$$

where $\mathcal{M}_\theta$ denotes the fusion function, and $\mathcal{M}_{\phi_i}$(or $\mathcal{M}_i$) denotes the unimodal predictor corresponding to modality i.

Figure 1 presents the general late-fusion architecture. Here, $\mathcal{M}_1, \ldots, \mathcal{M}_m$ are probabilistic unimodal discriminative models corresponding to each of the m modalitities. Each model $\mathcal{M}_j$ induces a distribution over the target Y conditioned on modality j and can be implemented by any differentiable probabilistic classifier such as a Multilayer perceptron (MLP). Let this distribution be $\mathbf{p}_j = P(Y \mid \mathbf{X}_j = \mathbf{x}_j)$. Late-fusion methods combine information from multiple modalities by defining a combining function over these probability distributions as the function $\mathcal{M}_\theta(\mathbf{p}_1, \ldots, \mathbf{p}_m)$.

We propose a probabilistic method for combining the unimodal predictions by employing probabilistic circuits. The

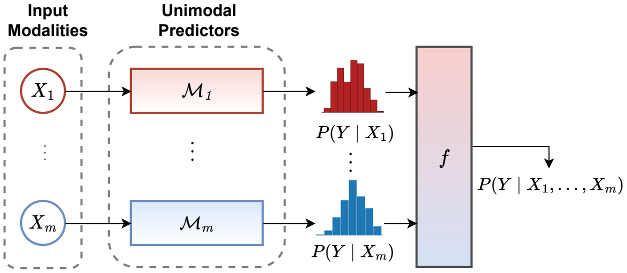

Figure 1: Late multi-modal fusion architecture for discriminative learning

resulting model can explicitly model complex correlations between the influence of each source on the target while still being able to reason about the credibility of each source, as we elaborate below.

Combining Unimodal Predictions with PCs

We define the combining rule f using a PC $\mathcal{M}_\theta$ that models the joint over the unimodal probability distributions and the target Y, i.e., $P_{\mathcal{M}_\theta}(Y, \mathbf{p}_1, \ldots, \mathbf{p}_m)$. We use categorical distributions at the leaf to model the target Y and Dirichlet distributions to model the unimodal predictive distributions $\mathbf{p}_1, \ldots, \mathbf{p}_m$ at the leaves. Since PCs are differentiable computational graphs, learning can be done in an end-to-end manner via backpropagation. The resulting late fusion method allows for two kinds of inference - predictive inference and credibility assessment.

Predictive Inference. Given a multi-modal example, $(\mathbf{x}_1, \ldots, \mathbf{x}_m)$, we can perform predictive inference over target Y in two steps

1. Compute $\mathbf{p}_j = P(Y \mid \mathbf{X}_j = \mathbf{x}_j) = \mathcal{M}_{\phi_j}$ for each modality $j = 1, \ldots, m$ using the unimodal predictors.
2. Infer the multimodal predictive distribution over Y given the unimodal distributions $\mathbf{p}_1, \ldots, \mathbf{p}_m$ by performing conditional inference over the PC ($\mathcal{M}_\theta$) as

$$P_{\mathcal{M}_\theta}(Y \mid \mathbf{p}_1, \ldots, \mathbf{p}_m) = \frac{P_{\mathcal{M}_\theta}(Y, \mathbf{p}_1, \ldots, \mathbf{p}_m)}{P_{\mathcal{M}_\theta}(\mathbf{p}_1, \ldots, \mathbf{p}_m)}$$

Credibility Assessment. In line with prior work on active feature elicitation (Natarajan et al. 2018; Das et al. 2023), we define credibility as the relative amount of information contributed by a modality to the multi-modal predictive distribution over the target Y. More specifically, we define the credibility of a modality in terms of the divergence between the conditional probability distribution excluding that modality and the conditional distribution including all modalities, i.e.

$$C_j = \delta(P(Y \mid \mathbf{X} \setminus \mathbf{X}_j) \mid\mid P(Y \mid \mathbf{X})) \qquad (1)$$

where δ is a divergence measure, such as the KL-Divergence, and $\mathbf{X} = \cup_i \mathbf{X}_i$. To facilitate comparison across modalities, we define the relative credibility score μ as

$$\mu_j = \frac{C_j}{\sum_j C_j}.$$

Note that $0 \leq \mu_j \leq 1 \forall j$ and $\sum_j \mu_j = 1$, and is therefore a normalized and probabilistic measure for assessing the credibility of modality j.

It must be mentioned that unlike neural methods, the tractability offered by a PC for conditional and marginal inference allows us to define probabilistic measures for the credibility of each modality. The resulting approach is also more robust as a PC can naturally handle missing modalities in the input by tractably marginalizing out the corresponding unimodal probability distribution in step 2.

Empirical Evaluation

To experimentally validate the utility of the proposed approach, we consider the AV-MNIST dataset, which is a benchmark dataset designed for multimodal fusion. It comprises two data modalities: images of dimension 28×28 depicting digits from 0 to 9, and their corresponding audio represented as spectograms of dimension 112×112. We consider the discriminative learning task of identifying the digit based on the multi-modal input. Following (Vielzeuf et al. 2018), we used deep neural models with the LeNet architecture to encode the input data and make predictions for each modality. Specifically, we processed the image input through a 4-layer convolutional neural network with filter sizes $[5, 3, 3, 3]$. Similarly, the audio input was encoded using a 6-layer convolutional neural network with filter sizes $[5, 3, 3, 3, 3, 3]$. The encoding obtained for each modality was processed through feedforward neural networks comprising of 1 hidden layer with 64 neurons to obtain the unimodal predictions. To facilitate seamless integration with the neural models, we use the deep parameterization proposed by (Peharz et al. 2020a) to implement our PC-based combination function. We also implemented 3 basline combination functions as elaborated below for comparison:

1. **Weighted Mean** combination function that defines the multimodal predictive distribution as: $P(Y|X_1, X_2, \ldots, X_m) = \sum_{i=1}^{m} w_i P(Y|X_i)$ where w_i are learnable weights such that $0 \leq w_i \leq 1$ and $\sum_{i=1}^{m} w_i = 1$. The constraints on the weights ensure that the combination function outputs valid distribution.

2. **Noisy-Or** combination function that defines the multimodal predictive distribution as:
$P(Y|X_1, X_2, \ldots, X_m) = 1 - \prod_{i=1}^{m}(1 - P(Y|X_i))$.

3. **Multi Layer Perceptron (MLP)** combination function that maps the vector of unimodal predictions $[P(Y|X_i)]_{i=1}^{m}$ to the multimodal predictive distribution $P(Y|X_1, X_2, \ldots, X_m)$ using a feedforward neural network having 1 hidden layer with 64 neurons.

For each fusion method, we use the same backbone architecture to obtain the unimodal predictions. We train all models end to end via gradient descent and backpropagation to minimize the cross-entropy loss between the targets and predictions, using an Adam optimizer with a learning rate of 0.001 and batch size of 128.

Overall, we aim to answer the following research questions empirically:

Fusion Model	Test Performance				
	Accuracy	Precision	Recall	F1Score	AUROC
MLP	72.57 ± 0.46	72.61 ± 0.43	72.42 ± 0.64	72.23 ± 0.96	96.26 ± 0.05
Weighted Mean	66.28 ± 2.05	66.45 ± 1.95	66.10 ± 2.20	65.80 ± 2.56	95.27 ± 0.04
Noisy-OR	68.88 ± 0.32	68.87 ± 0.31	68.70 ± 0.35	68.41 ± 0.83	94.48 ± 0.11
Probabilistic Circuit (ours)	72.45 ± 0.41	72.54 ± 0.51	72.30 ± 0.42	72.13 ± 0.63	96.39 ± 0.07

Table 1: Mean test performance of late fusion methods on the AV-MNIST dataset, ± standard deviation across 3 trials.

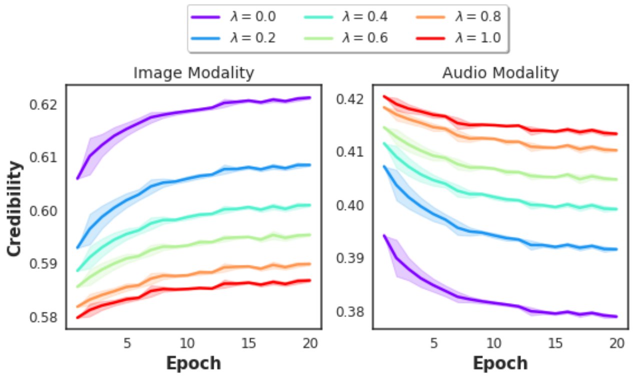

Figure 2: **Mean Validation Relative Credibility** obtained using a PC for the two modalities of the AV-MNIST dataset across training epochs. Varying degrees of noise (controlled by λ) are introduced into the audio modality. The shaded region represents the standard deviation across 3 trials.

(Q1) Can a PC-based combining rule efficiently capture intricate dependencies between modalities to achieve performance at par with existing methods?

(Q2) Can the tractability of PCs be used to reliably infer credibility scores for each source modality?

Performance Benchmarking

Table 1 summarizes the test-set performance of the baseline models and our PC model on the AV-MNIST dataset in terms of the classification metrics - Accuracy, Precision, Recall, F1-Score and AUC-ROC, after training for 50 epochs. We observe that our PC based combination function not only outperforms simple probabilistic baselines such - Weighted Mean and Noisy-Or on all performance metrics, but also achieves performance similar to that of an MLP based fusion method. Thus, the PC based late fusion method is expressive enough to capture intricate dependencies between unimodal predictive distributions.

Credibility Evaluation

We aim to evaluate whether our PC based late fusion method can reliably compute credibility of each modality. To this end, we design the following experiment. We consider the AV-MNIST dataset and a PC-based fusion model trained over it for 30 epochs. We now introduce varying degrees of noise into one of the modalities (say i), keeping others fixed, and train the PC to maximize the joint predictive likelihood. More specifically, we define

$$\tilde{P}(Y|X_i) = \lambda P(Y|X_i) + (1-\lambda)N$$

where $N \sim \text{Dir}(\alpha)$ is a noisy probability vector sampled from a Dirichlet distribution with parameters α, and $0 \leq \lambda \leq 1$. $\tilde{P}(Y|X_i)$ is thus a convex combination of two probability distribution and is therefore a valid distribution. λ controls the amount of information retained in $\tilde{P}$ from the unimodal predictive distribution.

Note that as $\lambda \rightarrow 0$, $\tilde{P}(Y|X_i) \rightarrow N$, and thus has less predictive information about modality i. Thus, the credibility score should ideally decrease for modality i and increase for the other modalities. Figure 2 shows how the mean relative credibility modeled by PC over the validation set varies as it is trained over the noisy unimodal distributions with noise introduced into the audio modality, for varying values of λ. As expected, we can see that the credibility of the audio modality decreases as training progresses, while that of the image modality increases. Further, we can also observe that the decrease in credibility increases as $\lambda \rightarrow 0$. To demonstrate this correlation more evidently, we plot the Mean Relative Credibility outputted by the trained PC for each modality on the test set, for the two settings where noise is introduced into one of image/audio modalities in Figure 3. We can clearly see that in both settings, the credibility score of the noisy modality decreases as $\lambda \rightarrow 0$, while that of the non-noisy modality increases. Thus, the credibility score outputted by the PC is a reliable measure that is reflective of the information contributed by each modality to the final predictive distribution.

By averaging the credibility of each modality over all datapoints, we have so far looked at a *global measure*, and the image modality seem to have higher global credibility than audio for AV-MNIST (see $\lambda = 1$). However, the credibility of each modality may differ locally for individual datapoints, which can also be evaluated efficiently using the PC.

Conclusion

We considered the problem of late multi-modal fusion in the discriminative learning setting. We developed a probabilistic circuit-based combination function for late-fusion that is expressive enough to model complex interactions, robust to missing modalities, and capable of making reliable and credibility-aware predictions. Our experiments demonstrate that the proposed approach is competitive with the state-of-the-art while allowing for a principled way to infer the credi-

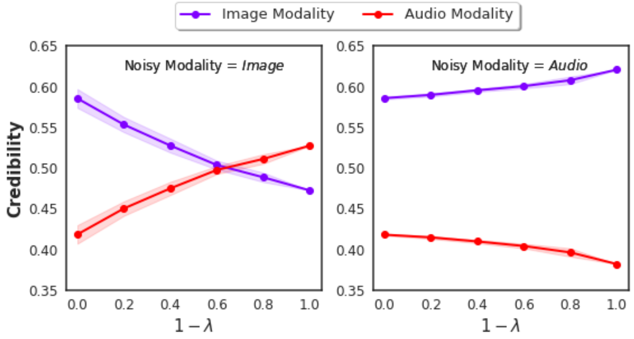

Figure 3: **Mean Test Relative Credibility** outputted by a PC for the two modalities of the AV-MNIST dataset across varying degrees of noise (controlled by λ) introduced into each modality. The shaded region represents the standard deviation across 3 independent trials.

bility of each modality. Future work includes scaling experimental evaluation to domains with more sources and extending the framework to allow subgroup-specific credibilities.

Acknowledgements

The authors acknowledge the support by AFOSR award FA9550-23-1-0239 and ARO award W911NF2010224.

References

Adel, T.; Balduzzi, D.; and Ghodsi, A. 2015. Learning the Structure of Sum-Product Networks via an SVD-based Algorithm. In *Conference on Uncertainty in Artificial Intelligence*.

Baltrušaitis, T.; Ahuja, C.; and Morency, L.-P. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence*, 41(2): 423–443.

Benediktsson, J.; Swain, P.; and Ersoy, O. 1990. Neural Network Approaches Versus Statistical Methods In Classification Of Multisource Remote Sensing Data. *IEEE Transactions on Geoscience and Remote Sensing*, 28(4): 540–552.

Blasch, E.; Jøsang, A.; Dezert, J.; Costa, P. C.; and Jousselme, A.-L. 2014. URREF self-confidence in information fusion trust. In *17th International Conference on Information Fusion (FUSION)*, 1–8. IEEE.

Blasch, E.; Laskey, K. B.; Jousselme, A.-L.; Dragos, V.; Costa, P. C.; and Dezert, J. 2013. URREF reliability versus credibility in information fusion (STANAG 2511). In *Proceedings of the 16th International Conference on Information Fusion*, 1600–1607. IEEE.

Choi, Y.; Vergari, A.; and Van den Broeck, G. 2020. Lecture Notes: Probabilistic Circuits: Representation and Inference.

Dang, M.; Vergari, A.; and Van den Broeck, G. 2020. Strudel: Learning Structured-Decomposable Probabilistic Circuits. In *Proceedings of the 10th International Conference on Probabilistic Graphical Models*, volume 138 of *Proceedings of Machine Learning Research*, 137–148. PMLR.

Das, S.; Ramanan, N.; Kunapuli, G.; Radivojac, P.; and Natarajan, S. 2023. Active feature elicitation: An unified framework. *Frontiers in Artificial Intelligence*, 6: 1029943.

De Villiers, J.; Pavlin, G.; Jousselme, A.; Maskell, S.; de Waal, A.; Laskey, K.; Blasch, E.; and Costa, P. 2018. Uncertainty representation and evaluation for modeling and decision-making in information fusion. *Journal for Advances in Information Fusion*, 13(2): 198–215.

Elouedi, Z.; Mellouli, K.; and Smets, P. 2004a. Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1): 782–787.

Elouedi, Z.; Mellouli, K.; and Smets, P. 2004b. Assessing sensor reliability for multisensor data fusion within the transferable belief model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 34(1): 782–787.

Fabre, S.; Appriou, A.; and Briottet, X. 2001. Presentation and description of two classification methods using data fusion based on sensor management. *Inf. Fusion*, 2: 49–71.

Gadzicki, K.; Khamsehashari, R.; and Zetzsche, C. 2020. Early vs late fusion in multimodal convolutional neural networks. In *2020 IEEE 23rd international conference on information fusion (FUSION)*, 1–6. IEEE.

Gens, R.; and Pedro, D. 2013. Learning the structure of sum-product networks. In *International conference on machine learning*, 873–880. PMLR.

Glodek, M.; Tschechne, S.; Layher, G.; Schels, M.; Brosch, T.; Scherer, S.; Kächele, M.; Schmidt, M.; Neumann, H.; Palm, G.; et al. 2011. Multiple classifier systems for the classification of audio-visual emotional states. In *Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part II*, 359–368. Springer.

Khaleghi, B.; Khamis, A.; Karray, F. O.; and Razavi, S. N. 2013. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1): 28–44.

Kline, A.; Wang, H.; Li, Y.; Dennis, S.; Hutch, M.; Xu, Z.; Wang, F.; Cheng, F.; and Luo, Y. 2022. Multimodal machine learning in precision health: A scoping review. *npj Digital Medicine*, 5(1): 171.

Manhaeve, R.; Dumancic, S.; Kimmig, A.; Demeester, T.; and De Raedt, L. 2018. DeepProbLog: Neural Probabilistic Logic Programming. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Mauro, N. D.; Vergari, A.; Basile, T. M. A.; and Esposito, F. 2017. Fast and Accurate Density Estimation with Extremely Randomized Cutset Networks. In *ECML/PKDD (1)*, volume 10534 of *Lecture Notes in Computer Science*, 203–219. Springer.

Natarajan, S.; Das, S.; Ramanan, N.; Kunapuli, G.; and Radivojac, P. 2018. On Whom Should I Perform this Lab Test Next? An Active Feature Elicitation Approach. In *IJCAI*, 3498–3505.

Natarajan, S.; Tadepalli, P.; Altendorf, E.; Dietterich, T. G.; Fern, A.; and Restificar, A. 2005. Learning first-order probabilistic models with combining rules. In *Proceedings of the 22nd international conference on Machine learning*, 609–616.

Nimier, V. 1998. Supervised multisensor tracking algorithm. In *9th European Signal Processing Conference (EUSIPCO 1998)*, 1–4.

Peharz, R.; Lang, S.; Vergari, A.; Stelzner, K.; Molina, A.; Trapp, M.; den Broeck, G. V.; Kersting, K.; and Ghahramani, Z. 2020a. Einsum Networks: Fast and Scalable Learning of Tractable Probabilistic Circuits. In *ICML*.

Peharz, R.; Vergari, A.; Stelzner, K.; Molina, A.; Shao, X.; Trapp, M.; Kersting, K.; and Ghahramani, Z. 2020b. Random Sum-Product Networks: A Simple and Effective Approach to Probabilistic Deep Learning. In *UAI*.

Poon, H.; and Domingos, P. 2011. Sum-product networks: A new deep architecture. In *UAI*.

Ramirez, G. A.; Baltrušaitis, T.; and Morency, L.-P. 2011. Modeling latent discriminative dynamic of multi-dimensional affective signals. In *Affective Computing and Intelligent Interaction: Fourth International Conference, ACII 2011, Memphis, TN, USA, October 9–12, 2011, Proceedings, Part II*, 396–406. Springer.

Rogova, G.; and Kasturi, J. 2001. Reinforcement learning neural network for distributed decision making. In *Proc. of the Forth Conf. on Information Fusion*.

Rogova, G. L.; and Nimier, V. 2004. Reliability in information fusion: literature survey. In *Proceedings of the seventh international conference on information fusion*, volume 2, 1158–1165.

Rooshenas, A.; and Lowd, D. 2014. Learning Sum-Product Networks with Direct and Indirect Variable Interactions. In *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, 710–718. Bejing, China: PMLR.

Sawhney, R.; Mathur, P.; Mangal, A.; Khanna, P.; Shah, R. R.; and Zimmermann, R. 2020. Multimodal multi-task financial risk forecasting. In *Proceedings of the 28th ACM international conference on multimedia*, 456–465.

Schulte, O.; and Routley, K. 2014. Aggregating predictions vs. aggregating features for relational classification. In *2014 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)*, 121–128. IEEE.

Shutova, E.; Kiela, D.; and Maillard, J. 2016. Black holes and white rabbits: Metaphor identification with visual features. In *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Human language technologies*, 160–170.

Sidheekh, S.; Kersting, K.; and Natarajan, S. 2023. Probabilistic Flow Circuits: Towards Unified Deep Models for Tractable Probabilistic Inference. In *The 39th Conference on Uncertainty in Artificial Intelligence*.

Tian, J.; Cheung, W.; Glaser, N.; Liu, Y.-C.; and Kira, Z. 2020. UNO: Uncertainty-aware noisy-or multimodal fusion for unanticipated input degradation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 5716–5723. IEEE.

Vielzeuf, V.; Lechervy, A.; Pateux, S.; and Jurie, F. 2018. Centralnet: a multilayer approach for multimodal fusion. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 0–0.

Wright, E. J.; and Laskey, K. B. 2006. Credibility models for multi-source fusion. In *2006 9th International Conference on Information Fusion*, 1–7. IEEE.