A Hybrid Modeling Framework for Crop Prediction Tasks via Dynamic Parameter Calibration and Multi-Task Learning

William Solow[1] , **Paola Pesantez-Cabrera**[2] , **Markus Keller**[2] , **Lav Khot**[2] , **Sandhya Saisubramanian**[1] , **Alan Fern**[1]
[1]Oregon State University
[2]Washington State University
{soloww, sandhya.sai, afern}@oregonstate.edu, {p.pesantezcabrera, mkeller, lav.khot}@wsu.edu

Abstract

Accurate prediction of crop states (e.g., phenology stages and cold hardiness) is essential for timely farm management decisions such as irrigation, fertilization, and canopy management to optimize crop yield and quality. While traditional biophysical models can be used for season-long predictions, they lack the precision required for site-specific management. Deep learning methods are a compelling alternative, but can produce biologically unrealistic predictions and require large-scale data. We propose a hybrid modeling approach that uses a neural network to parameterize a differentiable biophysical model and leverages multi-task learning for efficient data sharing across crop cultivars in data limited settings. By predicting the parameters of the biophysical model, our approach improves the prediction accuracy while preserving biological realism. Empirical evaluation using real-world and synthetic datasets demonstrates that our method improves prediction accuracy by 60% for phenology and 40% for cold hardiness compared to deployed biophysical models. Project site: https://tinyurl.com/DMC-MTL-Site.

1 Introduction

Accurate forecasts of crop states are critical for growers to schedule time-sensitive farm operations such as cold stress mitigation, pruning, fertilization, irrigation, and harvesting for specialty crops [Keller *et al.*, 2016; Milani and Cawley, 2024; Rogiers *et al.*, 2022]. However, accurate crop state prediction is challenging due to (1) limited availability of historical data for per-cultivar calibration and sparse observations during each growing season [Zapata *et al.*, 2017], and (2) the need to accurately model complex relationships between daily weather features and crop states [Guralnick *et al.*, 2024]. Existing approaches to this crop state prediction problem typically fall into two categories: mechanistic biophysical models and data-driven deep learning approaches.

Historically, biophysical models have been used to model a variety of crop states, such as phenology (timing of each developmental stage) and cold-hardiness (tolerance to low temperatures). Phenology is modeled by the Growing Degree Day (GDD) biophysical model based on daily accumulated heat units [Parker *et al.*, 2013]. Cold hardiness models predict the lethal bud temperature as a function of air temperature and phenological stage [Ferguson *et al.*, 2011]. Despite research supporting that crop states depend on exogenous weather features [Greer *et al.*, 2006], most specialty crop models only use air temperature as input, limiting their expressiveness [Badeck *et al.*, 2004]. Further, current biophysical models do not capture temporal nuances, such as how varied chilling hours in winter can change dormancy release and phenological development [Keller and Tarara, 2010]. These limitations affect their ability to produce accurate medium range (7-14 day) forecasts of the crop state [Reynolds, 2022].

Deep learning offers a compelling alternative to biophysical modeling due to its ability to model complex, nonlinear, and temporal relationships between weather variables and crop physiological states. However, purely data-driven models typically require large datasets and often produce *biologically unrealistic* predictions that violate known physiological constraints on plant development, such as predicting bud break after flowering. This makes them unsuitable for actionable medium range forecasts [Saxena *et al.*, 2023a].

To address the limitations of both deep learning and biophysical modeling for crop state forecasting tasks, we present a hybrid approach that uses a recurrent network to *dynamically refine the parameters* of a differentiable biophysical crop model, based on daily weather (Figure 1). For example, given historical weather data and phenology observations, the network is trained to assign the base temperature for heat accumulation in the GDD model such that the calibrated GDD model can accurately predict various phenological stages at deployment time. To address limited per-cultivar data, we use multi-task learning [Caruana, 1997] to share information across cultivars and improve prediction accuracy.

While hybrid modeling techniques have been successfully applied to a variety of physical processes [Jia *et al.*, 2021; Cai *et al.*, 2021], they remain underexplored for crop state tasks. Prior hybrid methods often approximate portions of the biophysical model [Van Bree *et al.*, 2025] or predict residuals [Vijayshankar *et al.*, 2021]. In contrast, our approach predicts parameters of the biophysical model, thereby producing biologically realistic and more accurate predictions.

Scope, Contributions, and Track Relevance. This work is motivated by the operational needs of our primary

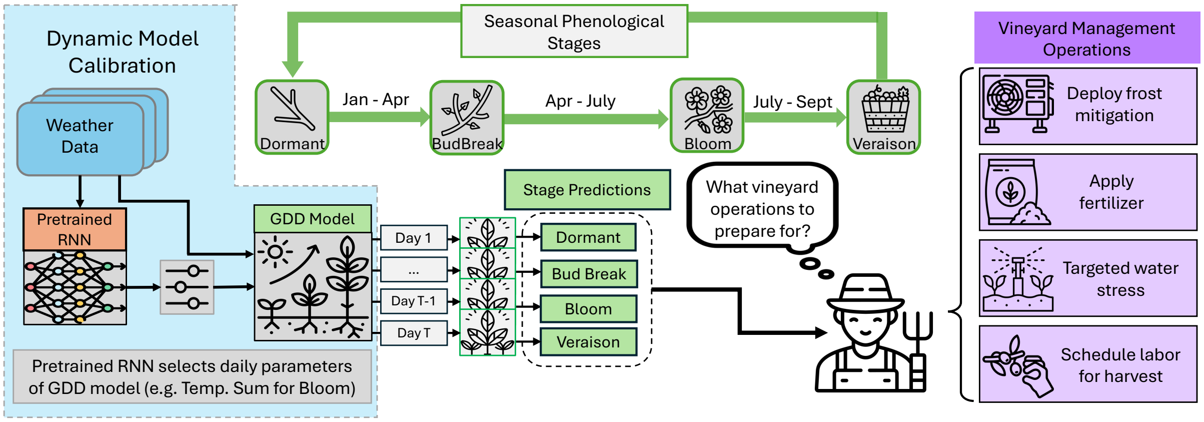

Figure 1: Overview of our proposed method using phenology prediction as an example. In this case, seasonal phenological stages guide vineyard management operations. Our Dynamic Model Calibration via Multi-Task Learning (DMC-MTL) approach uses a pretrained neural network and biophysical model to produce high fidelity and biologically realistic phenology state forecasts. The pretrained RNN produces daily parameter predictions (Base Temperature, Temperature Sum for Bud Break, etc.) of the Growing Degree Day (GDD) phenology model which handles the daily stage prediction.

stakeholders—specialty crop growers in the Pacific Northwest (PNW) of the United States who face increasing risks from weather variability and climate stress [Reynolds, 2022]. In this region, over $10 billion in annual specialty-crop production depends on timely and accurate crop state forecasts, including phenology and cold hardiness [Knowling *et al.*, 2021]. Our team includes AI experts and agricultural domain experts who played key roles in data collection, selection of biophysical models, and system deployment.

While our framework is broadly applicable to other crops, the paper focuses on wine grapes as a representative specialty crop and studies two critical crop states: *phenology* and *cold hardiness*. We train and validate our models using grapevine data collected in the PNW region between 1988-2025.

Our primary contributions are: (1) presenting a novel hybrid approach for accurate crop state forecasting by refining parameters of the biophysical model, conditioned on the weather features; (2) formulating the crop state prediction problem as a multi-task learning problem that leverages data efficiently across grape cultivars; (3) presenting an in-season adaptation variant of our model; and (4) empirical evaluations using real-world and synthetic datasets that demonstrate our approach's robustness and increased accuracy over state-of-the-art biophysical baselines, deep learning approaches, and hybrid models. Our model has been *deployed* on AgWeatherNet [WSU, 2025] which has over 26000 registered growers.

2 Background and Related Work

Hybrid Modeling of Biophysical Processes. Hybrid modeling combines deep learning and mechanistic modeling to obtain increasingly accurate and interpretable models of biophysical processes [Willard *et al.*, 2023]. Physics-Informed Neural Networks (PINNs; [Raissi *et al.*, 2019]) encode biological constraints in the form of partial differential equations into neural networks, enabling more accurate and biologically realistic predictions [Karpatne *et al.*, 2017; Karniadakis *et al.*, 2021]. In addition to PINNs, residual error hybrid models use a deep learning model to predict the difference between the true observation and the biophysical model prediction [de Mattos Neto *et al.*, 2022]. Process replacement hybrid models [Wang *et al.*, 2023] replace a poorly understood aspects of the biophysical model with a neural network with wide applications to the natural sciences [Feng *et al.*, 2022; Shen *et al.*, 2023]. Table 1 summarizes the characteristics of these models along different desiderate for crop state forecasting tasks. Another related work is that of Unagar *et al.* (2021) that use reinforcement learning for parameter calibration of a lithium battery. However, their problem setting assumes that the next true state of the system is known, which is untrue in our problem setting where medium range forecasts are needed. We adopt a supervised learning approach and train a network to dynamically modulate the parameters of the biophysical model in response to input features, enabling medium horizon forecasts.

Multi-Task Learning. When a set of tasks share similar structure, the multi-task learning [Zhang and Yang, 2018] framework can be used to efficiently aggregate data across tasks. Hard parameter sharing methods share a common set of hidden layers with unique prediction heads while soft parameter sharing methods learn unique models but regularize weight updates to keep models similar [Ruder, 2017]. Shared embedding spaces [Caruana, 1997] and task-specific embeddings [Changpinyo *et al.*, 2018] are viable methods for encoding task-specific information. We empirically evaluate these approaches in data-limited setting across genetically diverse grape cultivars treated as separate tasks, assessing prediction accuracy for all cultivars.

Machine Learning for Crop State Forecasts. Saxena *et al.* (2023a) applied multi-task learning to grape bud break prediction using a classification model. However, this model made erroneous predictions (e.g., predicting the onset of dormancy after bud break) that were inconsistent with biological processes. Saxena *et al.* (2023b) framed the grape cold hardiness prediction problem as a multi-task learning problem and

Modeling Approach	Biologically Realistic	Exogenous Features	Temporal Info.
Biophysical Model	✓	✗	✗
Deep Learning	✗	✓	✓
PINN	Sometimes	✓	✓
Residual	Sometimes	✓	✓
Process Replacement	✓	✗	✗
DMC-MTL (Ours)	✓	✓	✓

Table 1: Modeling approaches for crop state forecasting tasks evaluated along different desiderata. PINN, Residual, Process Replacement, and DMC-MTL are all hybrid models. *Biologically realistic* means the model predictions reflect biological laws. *Exogenous features* means that the model output can be conditioned on additional weather features. *Temporal info.* indicates that the model can leverage historical input in addition to what the biophysical model uses.

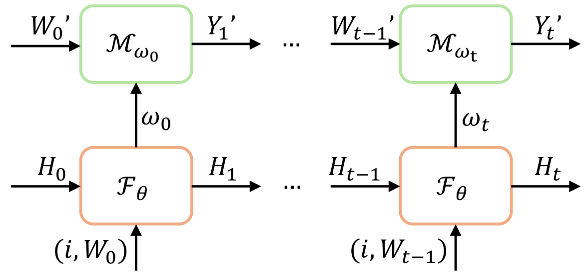

Figure 2: The network architecture of our approach DMC-MTL. The multi-task RNN ($\mathcal{F}_\theta$) sequentially embeds cultivar i.d. i and concatenates it with the daily weather features W_t to predict a parameterization ω_t of the biophysical model $\mathcal{M}$. Using the weather input to the biophysical model W_t' and the daily parameterization ω_t, crop state forecasts $Y_t' \ldots Y_{t+k}'$ can be generated.

used a recurrent neural network (RNN) to improve prediction accuracy over the deployed Ferguson biophysical model for cold hardiness [Ferguson *et al.*, 2014], demonstrating efficacy of multi-task learning to leverage data across cultivars. Van Bree *et al.* (2025) proposed a process replacement hybrid model for bloom date in cherry trees by approximating the temperature response function in the GDD model with a neural network. However, their method did not consider the effect of exogenous weather features on phenology nor the temporal variation of heat accumulation. In contrast, our method leverages a RNN to both encode temporal and exogenous weather information to avoid these limitations, and demonstrates state of the art performance in predicting grape phenology and matches grape cold hardiness while producing biologically realistic forecasts.

3 Dynamic Model Calibration

Our problem setting and approach, **D**ynamic **M**odel **C**alibration with **M**ulti-**T**ask **L**earning (DMC-MTL), is inspired by the following observations and hypotheses: (1) crop state prediction tasks are similar across cultivars, motivating our multi-task approach; (2) the parameters of crop state prediction models are hypothesized to vary over time based on historical weather in addition to the daily average temperature; and (3) crop state forecasting requires biologically realistic and accurate predictions.

Problem Formulation. We formulate the problem of estimating dynamic parameters of a biophysical model as a time series supervised learning problem and adopt the multi-task setting. Let $\mathcal{M}_\omega$ denote the biophysical model with parameters ω. Let D_i be the set of observed weather and daily crop states for each crop cultivar i. Let $S_{i,k}$ be the k-th season in D_i with $S_{i,k} = \{W_0, Y_0, \ldots, W_T, Y_T\}$ where W_t is the observed weather feature vector and Y_t is the observed crop state on day t. Given $W_t' \subset W_t$ as input (the biophysical model input is lower dimensional than the total number of observed weather features), $\mathcal{M}_\omega$ predicts a crop state Y_t'. We train a multi-task recurrent neural network model $\mathcal{F}_\theta$ that takes W_t and cultivar ID i as input, and outputs daily parameters ω_t of $\mathcal{M}$. The resulting parameterized model $\mathcal{M}_{\omega_t}$, along with W_t', is used to generate crop state predictions Y_t'. Given time series input $S_{i,k}$, we use $\mathcal{F}_\theta$ and $\mathcal{M}$ to obtain a sequence of parameter estimates $\omega_0, \ldots \omega_T$ and corresponding crop state predictions $Y_0', \ldots, Y_T'$ (Figure 2).

3.1 Model Architecture

The proposed model architecture for DMC-MTL is comprised of three parts: the RNN-backbone, the multi-task model, and the parameterization of the biophysical model.

The RNN-backbone (f_θ) contains two linear layers, followed by a Gated Recurrent Unit (GRU; [Chung *et al.*, 2014]), and another linear layer. To support multi-task learning across cultivars, we define $\mathcal{F}_\theta$ which adds a linear embedding layer before f_θ. This embedding layer converts a one-hot encoding of the cultivar into a dense vector, which is concatenated with the daily weather feature vector W_t and passed to f_θ, allowing the model to incorporate cultivar-specific information [Saxena *et al.*, 2023b]. ReLU activations are used, except for the final layer where a $\tanh$ activation is applied. The output of $\mathcal{F}_\theta$, which is in the range $[-1, 1]$, is then rescaled to match the parameter ranges of the biophysical model $\mathcal{M}$.

Figure 2 shows the daily parameterization ω_t of the biophysical model $\mathcal{M}$, the core of our DMC-MTL approach. $\mathcal{F}_\theta$ makes causal parameter predictions by sequentially processing a weather data sequence $W_0, \ldots, W_T$, generating corresponding parameters ω_t at each time step. These parameters are used to parameterize $\mathcal{M}_{\omega_t}$ and along with W_t', to produce phenology prediction Y_{t+1}'.

Biophysical Model Implementation. To learn $\mathcal{F}_\theta$, the biophysical model $\mathcal{M}$ must be differentiable and implemented in a framework that supports gradient backpropagation. In practice, most specialty crop state models are relatively simple and do not require advanced ordinary differential equation solvers. We additionally modify each biophysical model so that the parameters can be updated daily by $\mathcal{F}_\theta$ before each integration step. Parameters for the biophysical models are known to lie in specified ranges. To retain the ability of the DMC-MTL approach to capture complex dependencies among weather features, we choose large ranges for each parameter. Parameters have biophysical meaning; the Base Temp for Emergence parameter is the minimum temperature required before the grapevine can exit dormancy.

4 Extending DMC-MTL: In-Season Tuning

Collecting sufficient historical data for training accurate DMC-MTL models is challenging and often requires many years. In the case of insufficient data, it is beneficial to train the DMC-MTL model with available data and then adjust the model parameters based on *in-season observations* made by growers in the field. To enable in-season adaptation, *error signals* are generated by comparing DMC-MTL model predictions with sparse in-season observations such as occasional measurements of bud stage or cold hardiness. These error signals are then used to adjust future predictions of biophysical model parameters within the same growing season. This continual recalibration reduces bias accumulation seen in static prediction models and improves prediction accuracy.

When in-season observations are available, the daily parameter predictions made by the DMC-MTL approach are adjusted by using an additional neural network conditioned on error signals (Figure 3). While such Error Encoding Networks (EENs) have been used for video frame prediction [Henaff *et al.*, 2017], they have not been explored for the *sparse and low-dimensional* error signals available in crop state tasks. Furthermore, EENs have not previously been applied to hybrid models, so we investigate their effectiveness in reducing in-season predictive error and recalibrating the hybrid DMC-MTL model at small-to-medium horizon lengths.

Model Architecture. We use an EEN with the same model architecture as DMC-MTL ($\mathcal{F}_\theta$) given that observations are sparse and different cultivars may exhibit different error patterns. At each time step t the DMC-MTL model makes a prediction. If an in-season observation is available, the difference between the observation Y_t and model prediction Y_t' is passed to the EEN. If no observation is available, the EEN input is zero. This is consistent with the observation that if the DMC-MTL model makes no error, then the EEN should not change the future model predictions of the biophysical model parameters. We accomplish this by setting the bias terms within the EEN network layers to zero. EEN parameter predictions are combined *additively* with the predicted parameters by the pretrained DMC-MTL and then passed to the biophysical model to predict the next crop state.

Training EENs. Training the In-Season Adaptation model is a two step process. First, a DMC-MTL model is trained to predict parameters of the biophysical model that best predict the observed crop state. Then, the weights of the pretrained RNN $\mathcal{F}_\theta$ are frozen while the EEN is trained. The EEN is trained using the same training data as DMC-MTL, under the assumption that part of the observed crop state cannot be predicted solely by weather and cultivar ID [Henaff *et al.*, 2017]. We use historical crop state observations to mimic the availability of in-season observations in real time. We randomly mask available observations to ensure that the EEN does not learn rely on frequent observations for medium range forecasts, which may not be available at deployment time.

5 Experiment Setup

The performance of our proposed DMC-MTL approach is evaluated on two key criteria: (1) accurate and biologically realistic seasonal predictions and (2) efficient data use across cultivars. We also considered (3) how in-season data can be used to reduce prediction error and (4) robustness to unexpected subseasonal weather patterns, which can be demonstrated by evaluating a model on different weather distribution. Based on these, we design our experiments to answer the following research questions:

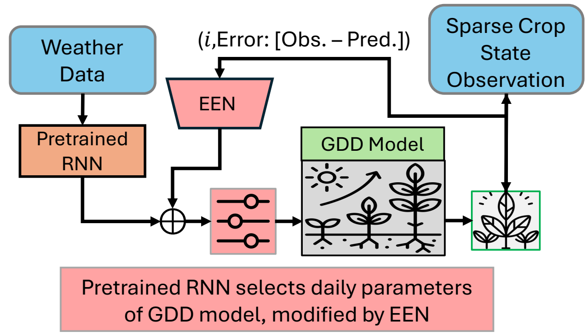

Figure 3: In-Season Adaptation with DMC-MTL. Cultivar i.d. i and error between observed and DMC-MTL predicted crop states are passed to the EEN. The parameters predicted by the EEN are combined additively to the prediction made by DMC-MTL's pretrained RNN before parameterizing the biophysical model.

Q1: (a) How does the average seasonal accuracy of DMC-MTL compare to deployed biophysical, deep learning and hybrid models? (b) Are predictions biologically realistic?

Q2: (a) Does DMC-MTL leverage data efficiently across cultivars? (b) How much per-cultivar data is required?

Q3: Do in-season crop state observations improve DMC-MTL model predictions?

Q4: Does DMC-MTL exhibit robustness to different weather conditions compared to other baselines?

Q5: What percentage of cultivars are accurately predicted by each model type?

5.1 Datasets

Real-World Datasets. We use the grape phenology and cold hardiness of 32 grape cultivars collected between 1988-2025. Phenology (stages of bud break, bloom, and veraison) was observed daily during the non-dormant season, and cold hardiness was measured weekly, biweekly, or monthly during the dormancy season. There are between eight and 21 years of phenological data per cultivar, and between four and 27 years of cold hardiness data per cultivar (43 to 797 samples). Pertinent historical open field weather data is sourced from AgWeatherNet [WSU, 2025].

Synthetic Datasets. To explore the robustness of DMC-MTL to different weather conditions, we generated datasets with two biophysical crop models: (1) the *GDD* model with 31 cultivars [Solow *et al.*, 2025], and (2) the *Ferguson* cold hardiness model with 20 cultivars [Ferguson *et al.*, 2011]. We used historical weather data from the NASAPower database for Washington, USA [NASA, 2025]. We also generated phenology and cold hardiness observations from Vermont, Cal-

	Solution Approaches	Grape Phenology	Grape Cold Hardiness
Q1a:	**DMC-MTL (Ours)**	**7.63 $\pm$ 3.56**	**1.21 $\pm$ 0.39**
	Bio. Model (Deployed)	18.58 $\pm$ 5.03*	2.03 $\pm$ 0.39*
	Bio. Model (Gradient Descent)	12.21 $\pm$ 5.13*	1.88 $\pm$ 0.42*
	Deep-MTL	8.16 $\pm$ 4.20*	1.30 $\pm$ 0.46
	TempHybrid	9.84 $\pm$ 4.35*	3.45 $\pm$ 0.98*
	PINN	8.61 $\pm$ 4.32*	1.30 $\pm$ 0.43
	Residual Hybrid	15.01 $\pm$ 6.00*	1.49 $\pm$ 0.52*
Q2:	DMC-STL	9.57 $\pm$ 3.79*	1.62 $\pm$ 0.34*
	DMC-Agg	9.81 $\pm$ 4.70*	1.51 $\pm$ 0.70*
	DMC-MTL-Mult	11.97 $\pm$ 4.43*	1.57 $\pm$ 0.53*
	DMC-MTL-Add	8.42 $\pm$ 3.56*	1.26 $\pm$ 0.41
	DMC-MTL-MultiH	8.20 $\pm$ 4.15*	1.34 $\pm$ 0.48*

Table 2: The average seasonal error (RMSE in days for phenology and $^\circ C$ for cold hardiness) over all cultivars and five seeds in the testing set for grape phenology and cold hardiness. DMC-MTL is compared against two optimization procedures for the biophysical model, a deep learning approach, three hybrid models, and two DMC variants. Best-in-class results are reported in bold. A * indicates that DMC-MTL yields a *statistically significant improvement* ($p < 0.05$) using the paired t-test relative to the corresponding baseline.

ifornia, and Oregon, USA. For each biophysical model, we generated ten years of data per cultivar using the biophysical models and historical NASA weather data. We randomly masked 88% of the daily cold hardiness samples to resemble the real-world dataset.

5.2 Baselines, Training and Evaluation

Baselines. We consider 11 baselines for our experiments: (1) *Deployed biophysical model*—the *GDD* model for phenology and the *Ferguson* model for cold hardiness that are used by our stakeholders; (2) *Gradient Descent* on the biophysical model parameters. To the best of our knowledge, this baseline has not been used before because the crop models have not been written in a differentiable framework; (3) *TempHybrid*—a hybrid model proposed by Van Bree *et al.* (2025) and adapted to the cold hardiness setting; (4) *Deep-MTL*—a multi-task model that either predicts probabilities for each phenological stage, or a continuous approximation of the cold hardiness. Instead of a $\tanh$ activation, there was a single output feature with no activation function. For cold hardiness we used the regression model proposed by Saxena *et al.* (2023b); (5) *PINN*—a Physics-informed neural network (PINN) with the same architecture and activation as the Deep-MTL model, trained with an additional loss term to weight biologically realistic predictions [Aawar *et al.*, 2025]; (6) *Residual Hybrid*—a hybrid model that uses an RNN to predict the difference between the biophysical model predictions and the observed crop state with the same network architecture as the Deep-MTL model [Vijayshankar *et al.*, 2021]; (7) *DMC-STL*—a DMC model without the embedding layer, trained on a per-cultivar basis; (8) *DMC-Agg*—a DMC model without the embedding layer and trained on all unlabeled cultivar data; (9) *DMC-MTL-Mult*—a DMC-MTL variant that uses a multiplicative embedding; (10) *DMC-MTL-Add*—a DMC-MTL variant that uses an additive embedding; (11) *DMC-MTL-MultiH*—a DMC-MTL variant that uses per-task prediction heads.

Baselines 1-6 to evaluate the efficacy of our approach against the biophysical, hybrid and deep learning baselines. Baselines 7-11 evaluate the efficacy of multi-task learning.

Model Training Protocol. For all experiments, we split the available grape cultivar data into training and testing sets. To build the test set, we withheld two seasons of data per cultivar from the training set. For the cultivars with the least amount of data, this resulted in two years of data in both the training and testing sets. Hyperparameters were selected using a validation set consisting of one season per cultivar.

Every model was trained for 400 epochs using a learning rate of 0.0002. We decreased the learning rate by a factor of 0.9 after a 10 epoch plateau of the training loss. For the deep learning phenology model we used Cross Entropy loss and used PINN loss for the PINN hybrid model with $p = 0.5$. For all other models, we used mean squared error loss function, masking days that did not have a ground truth observation.

Evaluation Protocol. We trained each model five times with different data splits and reported the average root mean squared error (RMSE) across cultivars on the test sets. For phenology, the RMSE was the cumulative error in days over the predictions for bud break, bloom, and veraison. For cold hardiness we reported the RMSE in degrees Celsius over all unmasked samples during the testing year.

6 Results and Discussion

Q1a: Average Performance of DMC-MTL. Table 2 shows the average RMSE values for grape phenology and cold hardiness predictions using different approaches, across 32 cultivars. The results show that DMC-MTL dramatically outperformed the biophysical models that are currently used by growers in PNW region—the GDD model for phenology and Ferguson model for cold hardiness. DMC-MTL also improved over gradient descent optimization for both phenology and cold hardiness. These results indicate the importance of better optimization procedures for crop model calibration, and how DMC-MTL can improve upon standard practices. Further, DMC-MTL improved upon the Deep-MTL, TempHybrid, PINN, and Residual Hybrid models, demonstrating the importance of dynamic parameterization, inclusion of exogenous weather features, and temporal information. We performed the paired t-test aggregated over all cultivars to con-

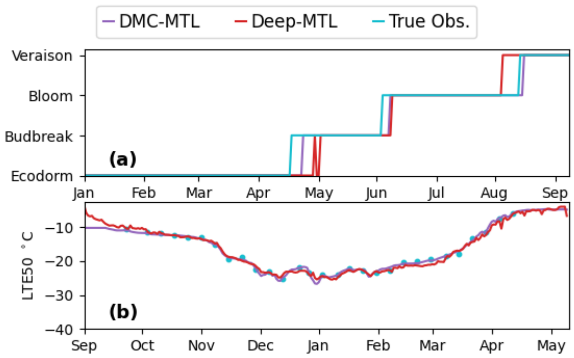

Figure 4: DMC-MTL, Classification, and Regression model predictions for (a) grape phenology and (b) grape cold hardiness. DMC-MTL makes biologically realistic predictions while deep learning model predictions do not always respect biologicaly laws.

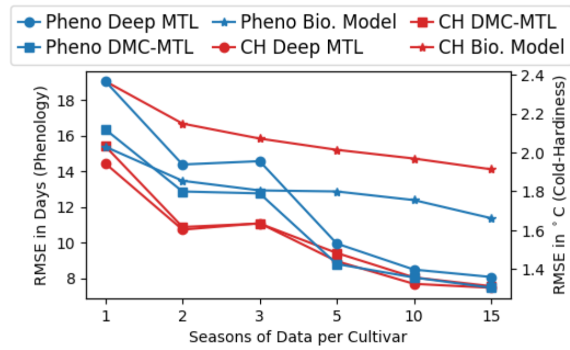

Figure 5: The performance of DMC-MTL models compared to deep learning and biophysical models under limited per-cultivar training data for grape phenology (Pheno) and cold hardiness (CH). Results are averaged over five seeds using the same two-seasons-per-cultivar evaluation sets.

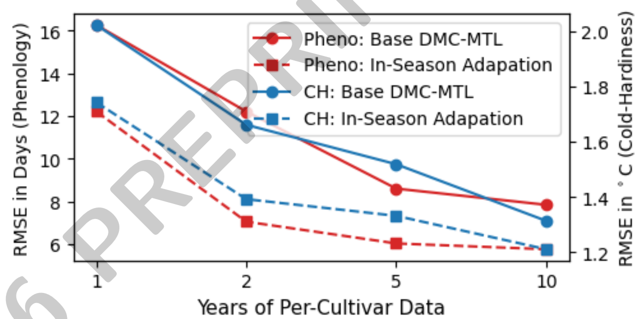

Figure 6: Performance of the base DMC-MTL with increasing per-cultivar data compared to the performance of the DMC-MTL with the additional in-season adaptation training for grape phenology (Pheno) and cold hardiness (CH). Results averaged over five seeds.

firm when our DMC-MTL performance improvements were statistically significant ($p < 0.05$). Overall, our results indicate that our hybrid modeling approach is more accurate and reliable for predicting crop state tasks.

Q1b: Biological Realism. Biologically realistic predictions are critical for interpreting medium-range forecasts that growers rely on to plan their vineyard operations. Figure 4 shows that Deep-MTL model incorrectly predicts bloom, reverts to bud break, then re-enters bloom three days later. Similarly, for cold hardiness, Deep-MTL overestimates the biologically plausible cold hardiness early in the growing season. In contrast, our DMC-MTL consistently makes biologically realistic crop state predictions.

Q2a: Multi-Task Data Sharing. Results in Table 2 show that DMC-MTL outperforms DMC-STL and DMC-Agg, demonstrating that both single-task learning and naive aggregation are insufficient for these crop state prediction tasks. Other implementations of multi-task learning (DMC-MTL-Mult/Add/MultiH) exhibited lower prediction accuracy compared to DMC-MTL. These results demonstrate that our learned one-hot concatenation embedding approach best encoded task-specific information, thereby efficiently leveraging data between cultivars.

Q2b: DMC-MTL Data Requirements. To evaluate the per-cultivar data requirements of DMC-MTL and compare its data efficiency to other models, we vary the number of seasons of available per-cultivar data from one to 15 during training. Our results in Figure 5 show that the average error decreases across all model types, as more data becomes available. While biophysical models seem to perform better in the low data regime, the deep models perform better with more data. In contrast, DMC-MTL outperforms both methods across the entire range of data sizes.

Q3: In-Season Adaptation with In-Field Observations. We evaluate the in-season adaptation approach against DMC-MTL by incorporating real-time in-field data. Using the same training data and protocol as the DMC-MTL models, we trained the in-season adaptation models with increasing amounts of per-cultivar data. Figure 6 shows the performance on the same testing data. By using the in-season observations to refine parameters predicted by the DMC-MTL model, the EEN yields the most benefit where DMC-MTL prediction errors are highest (the most data-limited settings). With only one season of data per cultivar, in-season adaptation reduces error by over 25%, enabling more actionable forecasts. As DMC-MTL accuracy improves when additional data is available during training, the benefit of the EEN diminishes, indicating that in-season adaptation is most valuable when historical per-cultivar data is scarce. Furthermore, this approach is only viable when we cannot overfit the model to the training data as is the case when observations are subjective and noisy in the phenology and cold hardiness domains.

Q4: Robustness to Differing Weather Conditions. Current crop models are calibrated on a site specific basis, limiting their applicability to regions with sufficient historical data. Further, they assume that weather conditions will remain consistent and do not account for extreme weather events which is critical for broader adoption. To evaluate our approach's robustness to varying weather conditions, we trained models on synthetic data from Washington, USA, and evaluated them on data from Vermont, Oregon, and California, which have moderately similar weather patterns. Exper-

Approach	Phenology				cold hardiness			
	WA (Train Loc.)	VT	CA	OR	WA (Train Loc.)	VT	CA	OR
DMC-MTL	5.9 ± 2.7	8.8 ± 5.6	30.4 ± 9.3	17.7 ± 0.5	0.42 ± 0.28	0.76 ± 0.31	1.37 ± 0.16	3.59 ± 0.24
Deep-MTL	6.1 ± 3.0	96.2 ± 10.8	$120. \pm 1.4$	78.0 ± 1.9	0.34 ± 0.22	5.98 ± 1.57	6.01 ± 0.07	3.73 ± 0.46
PINN	5.3 ± 2.9	60.5 ± 14.2	59.8 ± 1.3	58.4 ± 3.9	0.39 ± 0.23	4.86 ± 1.41	8.38 ± 0.25	4.02 ± 0.40
TempHybrid	6.4 ± 5.2	97.3 ± 16.2	$118. \pm 20.$	$83.0 \pm 18.$	4.25 ± 0.98	5.60 ± 0.98	8.57 ± 1.14	6.43 ± 1.33

Table 3: RMSE (days and $^{\circ}C$) for grape phenology and cold hardiness respectively, evaluated on unseen data sampled from the training location (WA) and from locations with a moderately similar weather (Vermont, California, and Oregon). Results averaged over five seeds.

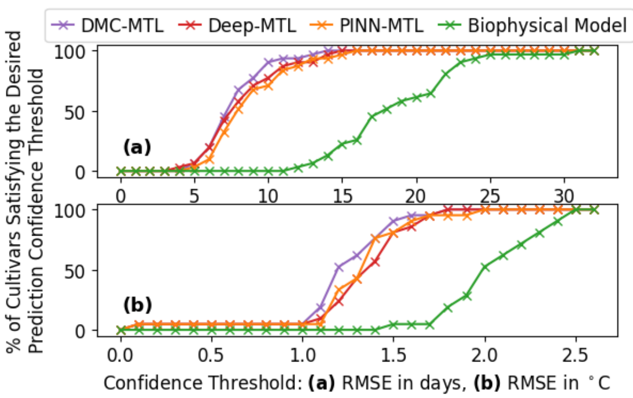

Figure 7: Percentage of all cultivars with cumulative error below a given RMSE threshold modeling a grape grower's tolerance for model prediction error. Results are reported over five seeds for (a) phenology and (b) cold hardiness.

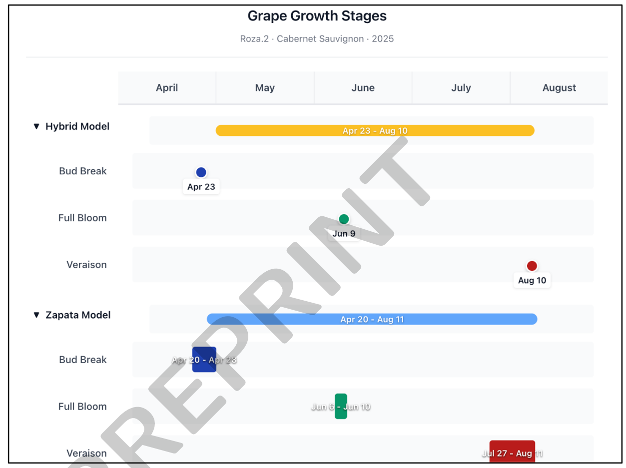

Figure 8: Screenshot of the publicly accessible user interface on AgWeatherNet. The weather station-specific phenology forecasts during the 2025 season for the Cabernet Sauvignon cultivar of the Zapata (GDD-based biophysical model) are compared with the hybrid (DMC-MTL) model.

iments were performed using synthetic phenology and cold hardiness datasets. Table 3 shows the cumulative RMSE in days on the Washington test set and Vermont, Oregon, and California test sets. While all models performed similar on the Washington test set, deep learning and other hybrid models produced large errors on the Vermont, Oregon, and California test sets. In contrast, DMC-MTL had a marginal increase in error on the test set, demonstrating robustness to varying weather conditions and its ability to produce usable predictions in unseen weather conditions.

Q5: Accuracy of Per-Cultivar Predictions. While DMC-MTL substantially reduces prediction error compared to biophysical models, its practical adoption depends on growers' error tolerance. To assess the potential model use, we evaluate the proportion of cultivars with RMSE below predefined thresholds, reflecting varying tolerance levels. We vary the RMSE tolerance threshold in the range $[0, 2.5]$ for cold hardiness and $[0, 30]$ for phenology. Our results in Figure 7 show that DMC-MTL consistently performs better than the baselines for both phenology and cold hardiness. In addition, DMC-MTL's biologically realistic predictions position its medium range forecasts to be predicted unambiguously, positioning it to be widely used in the field.

7 Deployment of DMC-MTL Models

DMC-MTL models for grape phenology were recently deployed on AgWeatherNet [WSU, 2025] for the 2026 growing season. Figure 8 is a screenshot of the user interface visible to growers, showing our model predictions for a specific weather station and grape cultivar (Cabernet Sauvignon) and that of the Zapata model (a GDD-based model that we compared against in Table 2) [Zapata *et al.*, 2017]. Stakeholders in the PNW widely use AgWeatherNet and will have access to both DMC-MTL and biophysical model forecasts to inform their vineyard operations which we will monitor for long-term impact. Continued analysis of deployed DMC-MTL models will provide insights to viticulturists on our interdisciplinary team to better understand the shortcomings of deployed biophysical models and improve understanding of grape phenological development. Outputs of all models are monitored for consistency and advice is posted in the case of rare climactic events that yield unexpected model predictions.

8 Conclusion and Future Work

We present a novel deep learning method that predicts the *parameters* of biophysical models. Our results show that leveraging the benefits of both deep network architecture and biophysical models can outperform both methods individually. In very data-limited settings, our in-season adaptation method provides growers with a more accurate prediction tool that is especially beneficial when waiting for more data is not feasible. Future work will aim to develop uncertainty quantification methods for crop state prediction with our hybrid modeling framework and expand predictions to additional regions.

Acknowledgments

The authors thank Lynn Mills and Zilia Khaliullina at the Irrigated Agriculture Research and Extension Center (IAREC), Washington State University, for their invaluable support in collecting and sharing grape phenology and cold hardiness data. The authors would also like to thank Sanjita Bhavirisetty, Jaitun Patel, and Dheeraj Vurukuti for their assistance in the deployment of DMC-MTL phenology models on AgWeatherNet. This research was supported by USDA NIFA award No. 2021-67021-35344 (AgAID AI Institute).

References

[Aawar *et al.*, 2025] Majd Al Aawar, Srikar Mutnuri, Mansooreh Montazerin, and Ajitesh Srivastava. Dynamics-Based Feature Augmentation of Graph Neural Networks for Variant Emergence Prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 27793–27801, April 2025.

[Badeck *et al.*, 2004] Franz-W. Badeck, Alberte Bondeau, Kristin Böttcher, Daniel Doktor, Wolfgang Lucht, Jörg Schaber, and Stephen Sitch. Responses of spring phenology to climate change. *New Phytologist*, 162(2):295–309, 2004.

[Cai *et al.*, 2021] Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-Informed Neural Networks for Heat Transfer Problems. *Journal of Heat Transfer*, 143(060801), April 2021.

[Caruana, 1997] Rich Caruana. Multitask Learning. *Machine Learning*, 28(1):41–75, July 1997.

[Changpinyo *et al.*, 2018] Soravit Changpinyo, Hexiang Hu, and Fei Sha. Multi-Task Learning for Sequence Tagging: An Empirical Study. In *International Conference on Computational Linguistics*, August 2018.

[Chung *et al.*, 2014] Junyoung Chung, Caglar Gulcehre, KyungHyun Cho, and Yoshua Bengio. Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling. In *NIPS 2014 Workshop on Deep Learning and Representation Learning*, December 2014.

[de Mattos Neto *et al.*, 2022] Paulo S. G. de Mattos Neto, George D. C. Cavalcanti, Domingos S. de O. Santos Júnior, and Eraylson G. Silva. Hybrid systems using residual modeling for sea surface temperature forecasting. *Scientific Reports*, 12(1):487, January 2022.

[Feng *et al.*, 2022] Dapeng Feng, Jiangtao Liu, Kathryn Lawson, and Chaopeng Shen. Differentiable, Learnable, Regionalized Process-Based Models With Multiphysical Outputs can Approach State-Of-The-Art Hydrologic Prediction Accuracy. *Water Resources Research*, 58(10):e2022WR032404, 2022.

[Ferguson *et al.*, 2011] John C. Ferguson, Julie M. Tarara, Lynn J. Mills, Gary G. Grove, and Markus Keller. Dynamic thermal time model of cold hardiness for dormant grapevine buds. *Annals of Botany*, 107(3):389, January 2011.

[Ferguson *et al.*, 2014] John C. Ferguson, Michelle M. Moyer, Lynn J. Mills, Gerrit Hoogenboom, and Markus Keller. Modeling Dormant Bud Cold Hardiness and Budbreak in Twenty-Three Vitis Genotypes Reveals Variation by Region of Origin. *American Journal of Enology and Viticulture*, 65(1):59–71, March 2014.

[Greer *et al.*, 2006] Dennis H. Greer, Jens N. Wünsche, Cara L. Norling, and Harry N. Wiggins. Root-zone temperatures affect phenology of bud break, flower cluster development, shoot extension growth and gas exchange of 'Braeburn' (Malus domestica) apple trees. *Tree Physiology*, 26(1):105–111, January 2006.

[Guralnick *et al.*, 2024] Robert Guralnick, Theresa Crimmins, Erin Grady, and Lindsay Campbell. Phenological response to climatic change depends on spring warming velocity. *Communications Earth & Environment*, 5(1):634, October 2024.

[Henaff *et al.*, 2017] Mikael Henaff, Junbo Zhao, and Yann LeCun. Prediction Under Uncertainty with Error-Encoding Networks. *arXiv:1711.04994*, November 2017.

[Jia *et al.*, 2021] Xiaowei Jia, Jared Willard, Anuj Karpatne, Jordan S. Read, Jacob A. Zwart, Michael Steinbach, and Vipin Kumar. Physics-Guided Machine Learning for Scientific Discovery: An Application in Simulating Lake Temperature Profiles. *ACM/IMS Trans. Data Sci.*, 2(3):20:1–20:26, May 2021.

[Karniadakis *et al.*, 2021] George Em Karniadakis, Ioannis G. Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. Physics-informed machine learning. *Nature Reviews Physics*, 3(6):422–440, June 2021.

[Karpatne *et al.*, 2017] Anuj Karpatne, Gowtham Atluri, James Faghmous, Michael Steinbach, Arindam Banerjee, Auroop Ganguly, Shashi Shekhar, Nagiza Samatova, and Vipin Kumar. Theory-guided Data Science: A New Paradigm for Scientific Discovery from Data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2318–2331, October 2017.

[Keller and Tarara, 2010] Markus Keller and Julie M. Tarara. Warm spring temperatures induce persistent season-long changes in shoot development in grapevines. *Annals of Botany*, 106(1):131–141, July 2010.

[Keller *et al.*, 2016] Markus Keller, Pascual Romero, Hemant Gohil, Russell P. Smithyman, William R. Riley, L. Federico Casassa, and James F. Harbertson. Deficit Irrigation Alters Grapevine Growth, Physiology, and Fruit Microclimate. *American Journal of Enology and Viticulture*, 67(4):426–435, October 2016.

[Knowling *et al.*, 2021] Matthew J. Knowling, Bree Bennett, Bertram Ostendorf, Seth Westra, Rob R. Walker, Anne Pellegrino, Everard J. Edwards, Cassandra Collins, Vinay Pagay, and Dylan Grigg. Bridging the gap between data and decisions: A review of process-based models for viticulture. *Agricultural Systems*, 193:103209, October 2021.

[Milani and Cawley, 2024] Alejandro Milani and Alejandro Mac Cawley. Analyzing the impact of forecast errors in the planning of wine grape harvesting operations using a multi-stage stochastic model approach. *arXiv:2405.19997*, May 2024.

[NASA, 2025] NASA. Nasapower database. https://power.larc.nasa.gov/, 2025.

[Parker *et al.*, 2013] Amber Parker, Inaki Garcia De Cortázar-Atauri, Isabelle Chuine, Gérard Barbeau, Benjamin Bois, Jean-Michel Boursiquot, Jean-Yves Cahurel, Marion Claverie, Thierry Dufourcq, Laurence Gény, Guy Guimberteau, Rainer W. Hofmann, Olivier Jacquet, Thierry Lacombe, Christine Monamy, Hernan Ojeda, Laurent Panigai, Jean-Christophe Payan, Begoña Rodriquez Lovelle, Emmanuel Rouchaud, Christophe Schneider, Jean-Laurent Spring, Paolo Storchi, Diego Tomasi, William Trambouze, Michael Trought, and Cornelis Van Leeuwen. Classification of varieties for their timing of flowering and veraison using a modelling approach: A case study for the grapevine species Vitis vinifera L. *Agricultural and Forest Meteorology*, 180:249–264, October 2013.

[Raissi *et al.*, 2019] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, February 2019.

[Reynolds, 2022] Andrew G. Reynolds. 11 - Viticultural and vineyard management practices and their effects on grape and wine quality. In *Managing Wine Quality (Second Edition)*, pages 443–539. Woodhead Publishing Series in Food Science, Technology and Nutrition, January 2022.

[Rogiers *et al.*, 2022] Suzy Y. Rogiers, Dennis H. Greer, Yin Liu, Tintu Baby, and Zeyu Xiao. Impact of climate change on grape berry ripening: An assessment of adaptation strategies for the Australian vineyard. *Frontiers in Plant Science*, 13:1094633, December 2022.

[Ruder, 2017] Sebastian Ruder. An Overview of Multi-Task Learning in Deep Neural Networks. *arXiv:1706.05098*, June 2017.

[Saxena *et al.*, 2023a] Aseem Saxena, Paola Pesantez-Cabrera, Rohan Ballapragada, Markus Keller, and Alan Fern. Multi-Task Learning for Budbreak Prediction. *arXiv:2301.01815*, January 2023.

[Saxena *et al.*, 2023b] Aseem Saxena, Paola Pesantez-Cabrera, Rohan Ballapragada, Kin-Ho Lam, Markus Keller, and Alan Fern. Grape Cold Hardiness Prediction via Multi-Task Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 15717–15723, 2023.

[Shen *et al.*, 2023] Chaopeng Shen, Alison Appling, Pierre Gentine, Toshiyuki Bandai, Hoshin Gupta, Alexandre Tartakovsky, Marco Baity-Jesi, Fabrizio Fenicia, Daniel Kifer, Xiaofeng Liu, Li Li, Dapeng Feng, Wei Ren, Yi Zheng, Ciaran Harman, Martyn Clark, Matthew Farthing, and Praveen Kumar. Differentiable modeling to unify machine learning and physical models and advance Geosciences. *arXiv:2301.04027*, May 2023.

[Solow *et al.*, 2025] William Solow, Sandhya Saisubramanian, and Alan Fern. Wofostgym: A crop simulator for learning annual and perennial crop management strategies. In *Reinforcement Learning Conference (RLC) 2025*, 2025.

[Unagar *et al.*, 2021] Ajaykumar Unagar, Yuan Tian, Manuel Arias Chao, and Olga Fink. Learning to Calibrate Battery Models in Real-Time with Deep Reinforcement Learning. *Energies*, 14(5):1361, January 2021.

[Van Bree *et al.*, 2025] Ron Van Bree, Diego Marcos, and Ioannis N. Athanasiadis. Hybrid Phenology Modeling for Predicting Temperature Effects on Tree Dormancy. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 28458–28466, April 2025.

[Vijayshankar *et al.*, 2021] Sanjana Vijayshankar, Jennifer King, and Peter Seiler. Wind Power Forecasting using LSTMs. In *2021 60th IEEE Conference on Decision and Control (CDC)*, pages 3658–3663, December 2021.

[Wang *et al.*, 2023] Pengfei Wang, Ze Zhu, Wenlong Liang, Longtao Liao, and Jiashuang Wan. Hybrid mechanistic and neural network modeling of nuclear reactors. *Energy*, 282:128931, November 2023.

[Willard *et al.*, 2023] Jared Willard, Xiaowei Jia, Shaoming Xu, Michael Steinbach, and Vipin Kumar. Integrating Scientific Knowledge with Machine Learning for Engineering and Environmental Systems. *ACM Computing Surveys*, 55(4):1–37, April 2023.

[WSU, 2025] WSU. Agweathernet. https://weather.wsu.edu, 2025.

[Zapata *et al.*, 2017] Diana Zapata, Melba Salazar-Gutierrez, Bernardo Chaves, Markus Keller, and Gerrit Hoogenboom. Predicting Key Phenological Stages for 17 Grapevine Cultivars (Vitis vinifera L.). *American Journal of Enology and Viticulture*, 68(1):60–72, January 2017.

[Zhang and Yang, 2018] Yu Zhang and Qiang Yang. An overview of multi-task learning. *National Science Review*, 5(1):30–43, January 2018.