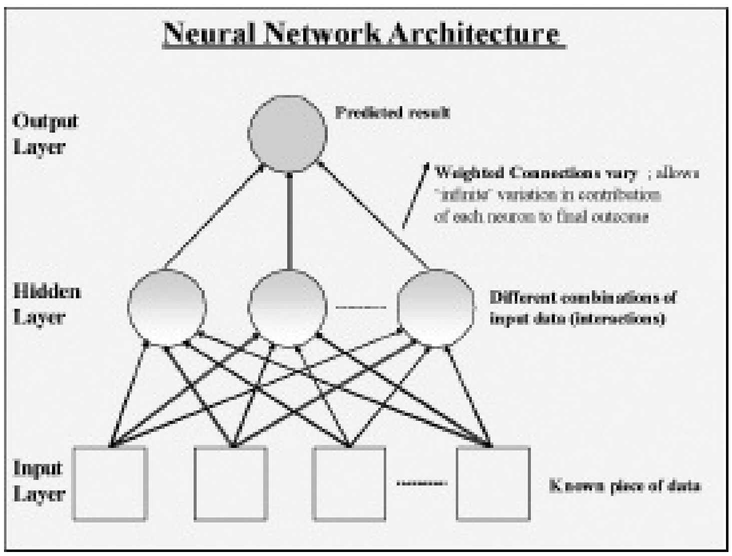
LET'S PUT SOME ARTIFICIAL INTELLIGENCE (A.I.) INTO AGRICULTURE

Dr. Bernie Hill, Neural Network Modeller

Agriculture and Agri-Food Canada, Research Centre, Lethbridge, AB T1J 4B1
hillb@agr.gc.ca; (403)-317-2267

Introduction

Neural network (NN) modelling is an 'artificial intelligence' technique. A NN is a computer simulation of biological (mammalian) neuron layers to form a mathematical model. Like mammalian brains, NN learn patterns from example data (data mining). The result is a non-linear model than can make future predictions from similar data. NN are adept at solving complex problems with many co-related variables. Agricultural systems are usually complex and involve co-related variables such as weather. NN have been around for 50 years, overshadowed until recently by expert systems. Expert systems use a set of rules supplied by an external expert, whereas NN develop their own 'rules' from the data without any pre-conceived biases. NN do not output a convenient formula or set of rules but they do provide a working model for predicting complex problems. NN are commercially available as software programs that run on Pentium PCs.



NN Architecture

A NN begins with an input layer where each neuron represents a piece of known data. Each neuron in the input layer is then connected to every neuron in a 'hidden' layer. The hidden layer neurons represent different combinations of data from the input layer (i.e. like interaction effects between the inputs). The hidden layer allows the NN to solve non-linear (real world) problems. Finally, each neuron in the hidden layer is connected to the output layer where each neuron (only one shown in diagram) represents one possible result. Information about a problem is supplied to the input layer and is then passed through the hidden layer to the output layer via the network of weighted connections. Adjusting the connection weights varies the contribution of each input neuron, and each hidden layer neuron, to the predicted result.

NN methodology

A NN is first 'trained' by supplying input information where the corresponding outputs are known. The NN adjusts itself (via the connection weights) so the predicted output agrees with the known output. The user has no control over the final connection weights, thus NN have a 'black box' aspect to them. Back propagation is commonly used to minimize the errors in fitting the weighted connections to the data. Many iterations are required to solve for the best-fit. One can influence the solution by carefully choosing the inputs and the program parameters. Once a NN is trained, it can be used to predict outputs for input data where the outputs are unknown.

NN modelling requires dividing the dataset into a training set used for model development (often 70% of data) and a validation set (30% of data) hidden from the NN software during training. The validation set is used to test how well the developed model will predict on data it hasn't seen before.

An important step in NN modeling is data pre-processing. This involves deciding how to present the data to the NN software. Often an input can be expressed in several different ways, for example temperatures can be expressed directly (such as 1 to 30 °C) or divided into categories (such as freezing, cold, mild, warm, hot) to represent the temperature on each day. Input data are also often transformed (computing the log, inverse, square root, etc.) so they correlate better with the output(s).

The NN modeling process is trial and error, and requires readjusting input presentations and parameter settings to obtain the 'best' network (the NN that makes the best predictions on validation sets). The best NN can be exported from the software as a matrix of weights (in Visual Basic code). An I/O routine with user-interface is written around this matrix, the code compiled, and distributed as a stand-alone executable file.

Summary of Steps in NN Modelling

1. conceptualize the model, decide on inputs and output(s)
2. create databases for training and validating the model
3. format and pre-process the data, apply transformations
4. train many different NN, vary NN parameters, test with validation sets, choose 'best' model
5. export model as Visual Basic flash code, create a user-interface, compile as an executable file
6. use model to predict unknowns

We began NN modelling of agricultural systems at the Lethbridge Research Centre in 1992. Some of our 'early' NN applications were:

1. predicting pesticide (cyhalothrin) dissipation rates under different weather scenarios
2. classifying suitability of sites for release of beetles for biological control of leafy spurge
3. predicting seeding dates from spring weather patterns (sci. publication)
4. estimating beef carcass quality from ultrasound measurements on live animals
5. predicting meat tenderness from beef carcass measurements 24-h post-mortem (sci. publication)
6. predicting maturity dates for hard red spring wheat in western Canada (sci. publication)
7. predicting seed quality from image analysis (several patents; commercialized by Dupont as the Acurum® Seed Analysis System)

More recently (since 2005), we have applied NN modelling to the following agricultural problems:

1. predicting healthy and sick feedlot animals from their daily feeding patterns
2. land cover classification from satellite image data
3. a real-time camera system to identify weeds in crops for selective spraying (sci. publication)
4. predicting days to harvest, and yields, of greenhouse-grown sweet peppers (sci. publication)
5. predicting cuticle cracking for greenhouse-grown peppers and tomatoes (sci. publication)
6. modelling greenhouse tomato growth
7. predicting the severity of bacterial ring rot on potatoes
8. predicting canola emergence, and yields

Applications no. 1, 2, 7, and 8 (immediately above) will be discussed in more detail. How developed NN models can be used by producers will be demonstrated for applications no. 7 and 8.

We used Predict® v3.13 software (NeuralWare®, Carnegie, PA) on a Pentium desktop PC (3.0 GHz) in the NN modelling applications described below.

Predicting Healthy and Sick Feedlot Animals from their Daily Feeding Patterns, in cooperation with Dr. Karen Schwartzkopf-Genswein (Lethbridge Research Centre)

Background

A significant number of animals get sick (usually with BRD) after being placed in a feedlot. Once identified, sick animals can be pulled, placed in a hospital pen, and treated with antibiotics. Early detection of sick animals reduces antibiotic treatments, and animal and production losses. Medication of healthy animals is undesirable. Although their feedings are irregular, it might be possible to differentiate between healthy and sick animals based on their feeding patterns, in advance of overt symptoms.

Objective

To classify animals, as healthy or sick, on different days prior to the sick animals being pulled by the pen checker.

Methods

Trials were conducted in 1998 (Feb. 17 to Jul. 7) and 2002 (Jan. 21 to Sep. 03) at a commercial feedlot near Amarillo, TX. Feeding behaviour of newly-received calves was monitored for up to 109 d using the GrowSafe® system which uses radio frequencies to record feed bunk attendance. The feeding durations and inter-meal intervals of individual animals were measured. The data (on a per animal basis) were arranged by day before pull (DBP), and datasets compiled such that healthy and sick animals were in a 1:1 ratio, and matched by pen and days-on-trial (DOT). The 1 DBP dataset was comprised of 108 animals for the 1998 trial, and 232 animals for the 2002 trial. In 1998, 39% of the sick animals were pulled by 5 DOT; in 2002, 27% were pulled by 5 DOT.

The following 13 inputs were used for NN modelling: initial body weight, DOT, no. feed bunk visits/d, feeding duration (min, max, avg, SD, total in sec/d), inter-meal intervals (min, max, avg, SD), and max. daily temperature. The output was animal condition, healthy or sick. The data were modelled one DBP at a time. We used a 10x cross-validation procedure (a rotating 10% validation set) to evaluate the NN.

Results

Typically, the NN used 6-12 inputs to classify the animals. Some typical NN architectures were: 9-4-2, 13-5-3, 14-6-2. Average (sick and healthy animals) classification accuracies on different DBP are presented in the Tables below. There were no differences in the accuracies for healthy and sick animals.

In both trials, the classification accuracies were the same (no statistical differences) over the entire range of 1-5 DBP and 1-10 DBP. We had expected the NN accuracies to ‘fall off’ with longer DBP.

How can a NN looking at feeding patterns pick-out sick animals 10 days before the pen checker? We think that rather than feeding behaviour indicating that an animal is in early stages of sickness, the NN are identifying the ‘poor eaters’. These ‘poor eaters’ then get sick more often. Apparently, the ‘poor eaters’ are consistent in their feeding patterns over the range of 1-10 DPB. This implies, once a ‘poor eater’ always a ‘poor eater’. Feedlot managers may be able to identify and manage the ‘poor eaters’?

The min time at the bunk, the min inter-meal interval, and DOT were consistently the most important indicators of each animal’s condition. Sick animals had longer times away from the bunk, but once at the bunk, they spent longer times eating. A lot of the sickness occurred at shorter DOT, i.e. soon after the animals were placed in the feedlot.

Classification Accuracies for the 1998 trial

DBP	% Accuracy* (SD)
1	81 (9)
2	74 (14)
3	74 (13)
4	74 (9)
5	72 (14)

* No significant differences (P>0.05) among accuracies on different DBP

Classification Accuracies for the 2002 trial

DBP	% Accuracy* (SD)
1	75 (5)
2	70 (10)
3	75 (4)
4	75 (8)
5	78 (10)
6	76 (5)
7	74 (11)
8	73 (10)
9	76 (9)
10	75 (11)

* No significant differences (P>0.05) among accuracies on different DBP

Classifying Agricultural Land Cover from SPOT Satellite Image Data, in cooperation with Dr. Anne Smith (Lethbridge Research Centre)

Background

The ability to identify land cover is useful for: censuses, crop insurance, providing inputs for production models, modelling environmental change, and evaluating carbon sequestration. Remote sensing provides a means of acquiring such data over large areas. Individual NN have been used to analyze image data and discriminate among land cover types. We have recently developed a NN tree structure, consisting of a series of linked NN, to try to improve land cover discrimination.

Objective

To use a NN Tree to classify agricultural land cover in southern Alberta.

Methods

The study area (approx. 40x40 mi) was located just east of Lethbridge, AB. It contained a variety of crops, a lot of native rangeland, and some fallow. Ground data (crop identifications, ground 'truthing') were provided by Alberta Financial Services Corp. and Alberta Agriculture and Food.

The SPOT satellite image was acquired on Aug. 19, 2004. Regions of interest (ROIs) were identified within this image as examples of each type of land cover. The ROIs were selected in pairs, one for NN training, and one for NN validation. Reflectance values for 4 bands, plus OSAVI (Optimized soil adjusted vegetation index) data, were extracted from the images on a pixel (20x20 m) by pixel basis.

The NN Tree consisted of 11 individual NN (each delivering a y/n decision) linked together. Land cover was classified pixel by pixel using the NN Tree. The NN Tree had following hierarchical scheme:

classify range&fallow vs. crops, then separate range vs. fallow; classify pulse&root crops vs. other crops, then separate pulses vs. root crops; classify oilseeds vs. other crops; classify cereals vs. forages, then separate cereals vs. broad leaf cereals.

The following 5 inputs were used for NN modelling: 4 reflectance bands B1-B4, plus OSAVI (calculated). The output was the classification (y/n) of each land cover category. NN were validated with the independent data from the designated ROI's.

Results

The 16 crops, rangeland, and fallow were classified into 8 subgroups (see the Table below).

Using 85% accuracy as a desirable threshold:

- Accuracies were good for rangeland, fallow, and broad leaf cereals.
- Accuracies need some improvement for root crops, forage, and cereals.
- Accuracies were poor for pulses and oilseeds.

Our NN Tree will require some fine-tuning to improve accuracies for certain vegetation types. For example, data from July when canola/mustard are flowering, may improve the oilseeds classification.

Accuracies of NN Tree method for classifying different land cover

	Vegetation type	Land cover group	% Accuracy
1	native rangeland	native rangeland	92
2	fallow	fallow	89
3	beans, chickpeas/lentils, peas	pulse crops	65
4	potatoes, sugar beets	root crops	80
5	canola/mustard	oilseeds	64
6	alfalfa, hay, tame pasture	forage	78
7	barley, 4 classes of wheat	cereals	77
8	corn, corn silage	broad leaf cereals	89

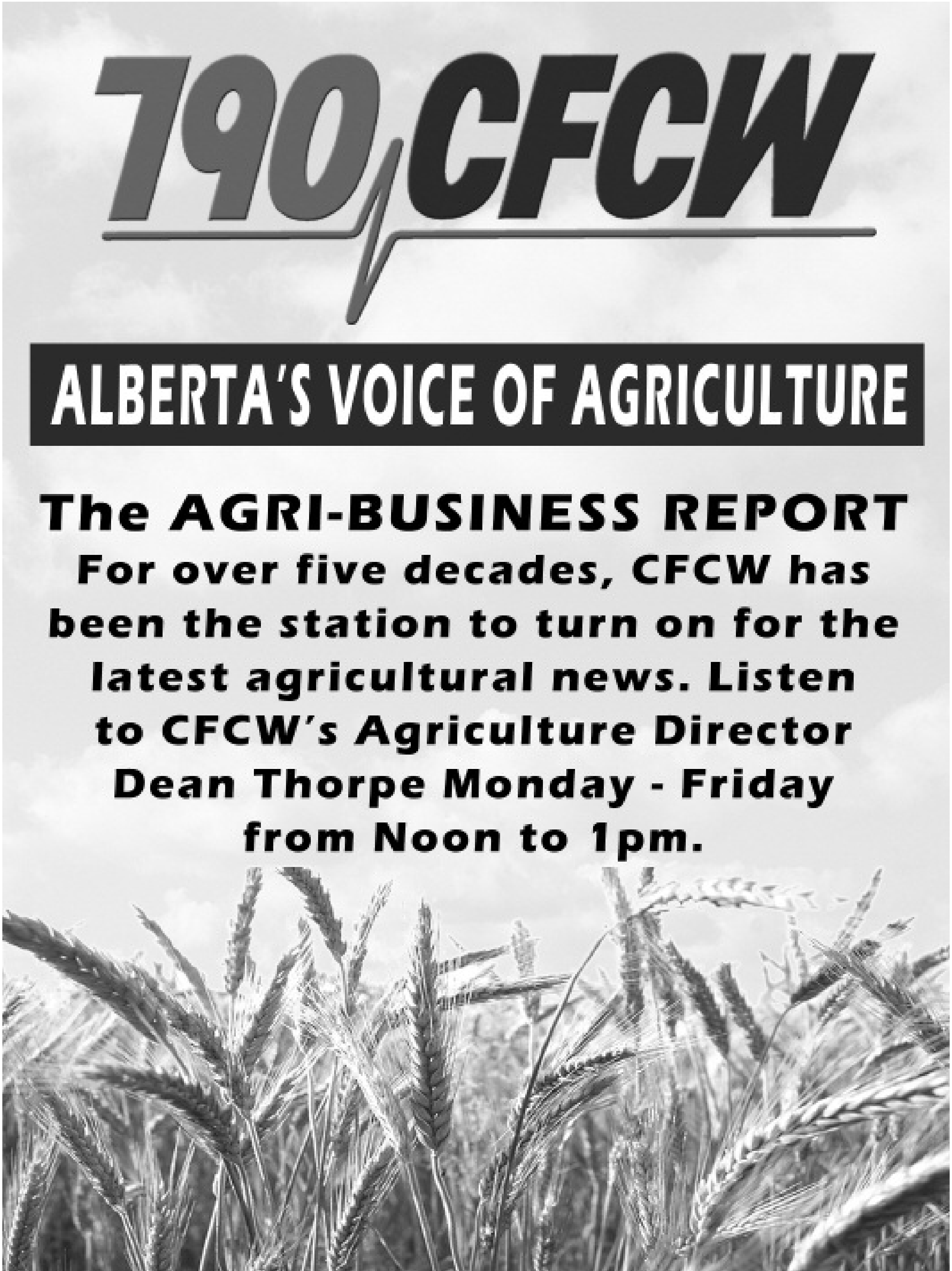
Predicting the Severity of Bacterial Ring Rot on Potatoes, in cooperation with Dr. Larry Kawchuk (Lethbridge Research Centre)

Background

Bacterial ring rot (BRR), for which there is zero tolerance, is caused by a tuber-borne bacterium which overwinters in potato debris, may reside in other hosts (sugar beets), can be spread by insects, and survives on equipment for up to 5 years. If BRR severity can be predicted during the current season, the industry can be alerted when there is high risk for BRR.

Objective

To predict the severity rating of BRR on potato foliage from seasonal weather variables.


790 CFCW
ALBERTA'S VOICE OF AGRICULTURE
The AGRI-BUSINESS REPORT
For over five decades, CFCW has been the station to turn on for the latest agricultural news. Listen to CFCW's Agriculture Director Dean Thorpe Monday - Friday from Noon to 1pm.

Methods

BRR trials were conducted at Stavely, AB (isolated field plots) over 11 different years (1986-1996) with many different varieties. Teton (resistant) and Norland (susceptible) were not included in the data set. The no. data (examples) were n=107.

The following 44 inputs were used for NN modelling: year, variety, seeding date, harvest date, plus 40 weather inputs. The weather was tabulated in 2 wk periods, from May 1 to Sept 30 (10 periods) with 4 weather variables per period, namely: AvgMaxT, AvgMinT, AvgT, TotPcp (to give 40 weather inputs). The output was: visual severity ratings (on a 0-5 scale) for BRR on potato foliage. We used a 10x cross-validation procedure (a rotating 10% validation set) to evaluate the NN.

Results

The original 44 inputs could be reduced to just 7 weather inputs and a good NN model obtained. The best NN model had an architecture of: 8-5-1 (two transformations used for one input). The avg R2 was 0.78 (SD=0.03), with an absolute avg error of 0.5 (0-5 scale).

Sensitivity analysis was conducted to rank the 7 inputs for their importance to the NN model (see Table below). The pcp during the last two weeks of May, and the maxT during the first two weeks of August, were the most important inputs.

The chronology of the weather inputs used in the NN model (see Table below) indicates that BRR severity depends mainly on pcp in late May/June and maxT in July/early August. This agrees with our understanding of conditions that influence BRR severity.

Ranking of importance of the 7 inputs

	input	2-wk period	rank
1	Pcp	May 16-31	**6.1**
2	MaxT	Aug 1-15	**1.3**
3	MaxT	July 1-15	0.76
4	Pcp	Jun 1-15	0.71
5	MaxT	July 16-31	0.61
6	MinT	May 16-31	0.30
7	Pcp	Jun 16-30	0.26

Chronology of inputs used in the NN model

2-wk period	pcp	minT	maxT
May 1-15			
May 15-31	**pcp**	minT	
Jun 1-15	pcp		
Jun 16-30	pcp		
Jul 1-15			maxT
Jul 16-31			maxT
Aug 1-15			**maxT**
Aug 16-31			

Demo: There will be a demo of the deployed NN model to predict BRR severity.

Predicting the Emergence of Seeded Canola Across the Prairies, in cooperation with Murray Hartman (AAFRD, Lacombe)

Background

Canola emergence can be quite variable. Seeding rates are often increased to compensate for poor emergence. This has become an issue with the expense of the new hybrid seed types. If producers could estimate canola emergence more accurately, seeding rates could be adjusted to achieve the desired stands.

Objective

To categorize canola emergence (high, medium, low) from seasonal weather variables.

Methods

Data were obtained from plot trials over the three Prairie provinces (270 from AB, 241 from SK, 72 from MB) over 13 different years between 1985-2006 (mostly from 1999-2006). Varieties were designated as: either hybrid or other; locations as: south, mid, north of each province; soil zones as: black, thin black, dark brown, grey; and tillage as: conventional or direct seeding. Seeding dates were represented as Julian dates. The no. data (examples) were n=583.

The following 22 inputs were used for NN modelling: year, location, soil zone, tillage, variety, seeding date (SDate), plus 16 weather inputs. The weather was tabulated in 1 wk periods, from 6 wk before seeding date until 2 wk after seeding (8 inputs). There were 2 weather variables per period: AvgT and TotPcp (to give 16 weather inputs). The output was: canola emergence category (low 11-50%, medium 51-69%, high 70-100%). We used a 10x cross-validation procedure (a rotating 10% validation set) to evaluate the NN.

Results

The original 22 inputs could be reduced to just 6 inputs (SDate + 5 pcp inputs) and a good NN model obtained. The NN architecture for the best model was: 15-10-3 (more than one transformation used for several inputs). The avg classification accuracy was 0.70 (low=0.73, med=0.74, high=0.64).

It should be noted that, despite all the 22 inputs including location and soil zone, predicting canola emergence came down to the 'moisture before seeding'. Sensitivity analysis was conducted to rank the 6 inputs for their importance to the NN model (see Table below). It is the pcp, 1 and 2 weeks before seeding, that mainly determines canola emergence.

Ranking of importance of the 6 inputs

	input	relative to seeding	rank
1.	Pcp	2 wk before	**421**
2.	Pcp	1 wk before	**228**
3.	Pcp	3 wk before	34
4.	Pcp	4 wk before	20
5.	Pcp	5 wk before	15
6.	SDate	0 wk before	1

Demo: There will be a demo of the deployed NN model to predict canola emergence.