

EXPO-AGRI: PRIMI RISULTATI RELATIVI AD UN MODELLO PREDITTIVO APPLICATO A INFEZIONI DI PERONOSPORA DEL BASILICO IN SERRA

A. CASTELLINI¹, A. FARINELLI¹, A. MINUTO², D. QUAGLIA¹, I. SECCO³, F. TINIVELLA²

¹ Dipartimento di Informatica, Università degli Studi di Verona – Strada le Grazie, 15, 37134 Verona, Italia

² Centro di Sperimentazione e Assistenza Agricola (Ce.R.S.A.A.), Regione Rollo, 98, 17031 Albenga, Italia

³ Agricontrol Snc Di Balbo & Secco, Reg. Str. Torre Pernice, 13/6, Albenga, Italia
federico.tinivella@gmail.com

RIASSUNTO

Nell'ambito dell'agricoltura di precisione la possibilità di prevedere e avere sotto controllo il comportamento delle piante all'interno di ambienti confinati è divenuto un requisito sempre più importante. In questo contesto, gli strumenti basati sulle reti di sensori e sull'intelligenza artificiale rappresentano fattori chiave per ottimizzare i processi di acquisizione dei dati, la modellizzazione matematica e il processo decisionale. In questo contributo è presentata un'architettura generale per il controllo automatico dell'ambiente serra. Nello specifico, è stato elaborato un modello predittivo preliminare in grado di prevedere il rischio di comparsa di nuove infezioni di peronospora del basilico (*Peronospora belbahrii*) su piante di basilico coltivato in serra. L'architettura possiede tre elementi principali di innovazione: nuovi tipi di sensori sono utilizzati per acquisire informazione sullo stato delle piante, predittori matematici sono generati a partire da questa informazione per mezzo di sofisticati metodi di processamento, ed infine i predittori più informativi sono selezionati in modo automatico per mezzo di tecniche di regolarizzazione.

Parole chiave: sensori, predittori, basilico, peronospora, modellizzazione

SUMMARY

EXPO-AGRI: FIRST RESULTS REGARDING A PREDICTIVE MODEL ADOPTED FOR BASIL DOWNY MILDEW INFECTIONS IN GREENHOUSE

Predicting and controlling plant behavior in controlled environments is a growing requirement in precision agriculture. Sensor networks and artificial intelligence methods represent key aspects for optimizing the processes of data acquisition, mathematical modeling and decision making. This paper presents a general architecture for automatic greenhouse control. In particular, a preliminary model for predicting the risk of new infections of downy mildew (*Peronospora belbahrii*) on sweet basil in greenhouse was designed. The architecture has three main elements of innovation: new kinds of sensors are used to acquire information about the state of the plants, model predictors are generated from this information by non-trivial processing methods, and informative predictors are automatically selected using regularization techniques.

Keywords: sensors, predictors, basil, downy mildew, modeling

INTRODUZIONE

In agricoltura le principali coltivazioni subiscono annualmente perdite che ammontano a circa il 20% dell'intera produzione a causa dei danni provocati da patogeni vegetali (ad es. funghi, batteri, virus) (Schumann, 1991). Per decenni i tradizionali prodotti fitosanitari hanno garantito indubbi vantaggi alle colture attraverso l'efficace contenimento di patogeni e parassiti. Più recentemente, l'accresciuta preoccupazione da parte dei consumatori nei confronti dell'impiego di prodotti chimici in agricoltura e dei loro potenziali effetti negativi sull'ambiente e sulla salute umana, ha portato alla messa a punto di una legislazione armonizzata volta alla riduzione dei rischi derivanti dall'uso dei prodotti fitosanitari.

In questo ambito, l'impegno di strumenti propri dell'agricoltura di precisione diventa fondamentale. Storicamente i primi approcci in questo senso sono consistiti nella raccolta di differenti variabili di tipo fisico come umidità, temperatura dell'aria e del suolo attraverso sensori caratterizzati da basso costo, robustezza, autoalimentati e in grado di comunicare in modalità *wireless* (Kassim et al., 2014). L'interpretazione dei grafici derivanti dall'elaborazione delle succitate variabili era demandato al tecnico agronomo. Un più evoluto tipo di approccio ha previsto poi lo sviluppo di modelli automatici di predizione che potessero interpretare l'enorme numero di dati raccolti da parte dei sensori e fornire come output la probabilità che si verificasse un evento critico come la comparsa e lo sviluppo di una malattia. In Aegerter et al. (2003), ad esempio, lo studio dei fattori ambientali che condizionano la comparsa della peronospora della rosa ha portato allo sviluppo di un modello predittivo applicato alla coltivazione in vivaio di questa specie. Kim et al. (2014) hanno sviluppato un modello previsionale relativo al rischio di comparsa di peronospora su ibrido di rovo.

Infine in Behmann et al. (2015) è possibile trovare un'indagine relativa a metodi avanzati di "machine learning" per il rilevamento di stress biotici a carico delle piante. La valutazione dei rischi è utilizzata per guidare le azioni degli agricoltori, ad esempio per quanto riguarda l'impiego degli agrofarmaci. I Sistemi di Supporto alle Decisioni (Decision Support Systems – DSS) rappresentano uno dei principali ambiti di applicazione dell'intelligenza artificiale alle tecniche di produzione in agricoltura (Jensen e Jensen, 1996).

Il progetto Expo-Agri si propone di migliorare la tecnologia DSS quando applicata alle coltivazioni in serra. Mentre in DSS di tipo tradizionale le azioni che vengono suggerite sono poi attuate manualmente, nell'approccio proposto in questa sede viene realizzato un sistema di controllo automatico a ciclo chiuso dove un regolatore dà inizio ad una serie di azioni specifiche sugli attuatori presenti all'interno della serra. In modo particolare, temperatura, umidità, luce e bagnature possono essere regolate automaticamente al fine di creare condizioni sfavorevoli allo sviluppo della malattia. Come caso studio è stata scelta la coltivazione di basilico (*Ocimum basilicum* L.), specie coltivata in Italia su circa 900 ha di superficie in serra e pieno campi (dati Ce.R.S.A.A.). La peronospora del basilico, causata dall'oomicete biotrofo *Peronospora belbahrii*, è diventato un patogeno chiave del basilico in molti Paesi. In Italia ha fatto la sua comparsa in varie aziende agricole in Liguria nel 2003 e ha causato ingenti perdite di produzione in tutte le aree in cui il basilico è coltivato. Lo sviluppo della malattia è correlato alle condizioni ambientali, nello specifico un'elevata umidità relativa (80-100%), temperature miti (20-25 °C, ottimale 20 °C), bagnatura fogliare di almeno 6 ore (Garibaldi et al., 2007).

Il lavoro presenta i risultati intermedi del progetto Expo-Agri, nello specifico l'architettura di controllo proposta, la metodologia utilizzata per creare il modello predittivo e l'introduzione di un nuovo sensore di temperatura fogliare che, in base ai risultati preliminari, è in grado di apportare nuove utili informazioni. La metodologia adottata per creare il modello predittivo è particolarmente interessante dal momento che le serie di dati provenienti dai

sensori non viene utilizzata tout court ma è stato sviluppato un sistema dedicato di estrazione per generare predittori in grado di fornire maggiori informazioni sul comportamento della pianta. Inoltre viene adottata una metodologia automatica per selezionare le caratteristiche che veicolano più informazioni da inserire all'interno del modello predittivo. Questo metodo aiuta anche ad identificare i sensori che possono più utilmente essere impiegati in funzione del fenomeno che si intende studiare. Il contenuto della presente pubblicazione è stato oggetto di presentazione all'interno del convegno internazionale "IEEE Biomedical Circuits and Systems Conference (BioCAS)" tenutosi a Torino in data 19-21 ottobre 2017 (Castellini et al., 2017).

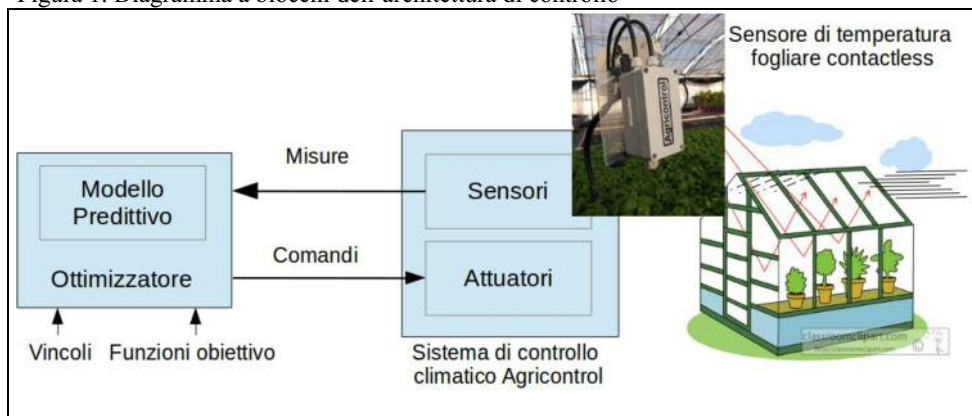
MATERIALI E METODI

Il progetto ha l'obiettivo di definire un'architettura generale per il controllo automatico di serra. La tecnica di controllo è basata su un modello predittivo collegato ad un sistema di acquisizione dati e di attuazione. Le componenti sono qui di seguito descritte.

A. Architettura del sistema

L'architettura di controllo (Figura 1) segue un approccio detto *model predictive* (Maciejowski, 2001). Il modello predittivo è alimentato da misurazioni provenienti dal sistema di acquisizione dati. Queste misurazioni sono utilizzate per stimare serie temporali di evoluzioni future del sistema serra in funzione di diversi comandi di controllo. Serie temporali e possibili comandi sono processati da un modulo di ottimizzazione che individua i comandi che minimizzano l'errore tra l'evoluzione reale del sistema e quella desiderata dall'utente che, in particolare, punta ad avere un basso rischio di infezione del patogeno. Anche se il focus attuale è quello di minimizzare l'infezione, il modulo di ottimizzazione potrebbe consentire in futuro di raggiungere diversi obiettivi, come la minimizzazione dei costi di coltivazione o quella del consumo di energia ed acqua.

Figura 1. Diagramma a blocchi dell'architettura di controllo



B. Sistemi di acquisizione dati ed attuazione

Il sistema di acquisizione dati e controllo di serra è basato sul sistema di controllo climatico distribuito MCX di Agricontrol. Tutti i sensori e attuatori sono collegati ad un PC supervisore attraverso un collegamento seriale multi-punto. Il nuovo sistema di controllo può usare sia misure interne alla serra (ad es. temperatura dell'aria della serra, umidità relativa, intensità della luce, radiazione solare, radiazione PAR e UVA) sia esterne come velocità e direzione del

vento e precipitazioni. Un bus seriale RS-485 collega i seguenti sensori per il monitoraggio dei bancali sperimentali:

- sensore di temperatura e umidità bulbo asciutto e bulbo umido;
- sensore PT100 di temperatura del substrato;
- sensore ad infrarossi per la misura senza contatto della temperatura fogliare;
- sensore di umidità del terreno basato su *frequency domain reflectometry*.

Il sensore ad infrarossi per la temperatura fogliare è stato espressamente sviluppato durante il progetto (dettaglio in alto a destra di Figura 1). Esso è basato sul sensore ad infrarossi Melexis MLX90614-EBAC. L'intervallo di misura è 0-100 °C con risoluzione di 0.1 °C, accuratezza di ± 0.5 °C e angolo di vista di 35°.

Il sistema MCX periodicamente invia un comando di lettura ai sensori attraverso il bus seriale e riceve le risposte con le letture.

C. Dataset

Le prove di coltivazione sono state realizzate in una serra ferro/vetro presso le strutture del Ce.R.S.A.A.. Due bancali in metallo di 5 m² di superficie sono stati riempiti con un substrato di coltivazione di tipo commerciale (torba 70% v/v, pomice 25% v/v, argilla 5% v/v) e seminati con semi di basilico cv "Italiano classico" alla dose di 2 g/m². I bancali sono stati riscaldati attraverso tappetini elettrici disposti al di sotto del substrato al fine di mantenere il substrato alla temperatura di 30 °C fino all'emergenza delle piantine, dopodiché il bancale 1 è stato mantenuto caldo per l'intera durata della prova mentre il bancale 2 è stato spento.

In seguito alla comparsa di sintomi di peronospora (clorosi fogliare e presenza di sporangi di colore grigio-bruno sulla pagina inferiore delle foglie), sono stati condotti rilievi giornalieri relativi alla diffusione del patogeno su entrambi i bancali espressi come numero di foglie mostranti sintomi su un totale di 50 foglie osservate all'interno di un'area scelta arbitrariamente all'interno del bancale replicata 3 volte. Le prove di coltivazione sono state interrotte quando la diffusione di peronospora non mostrava un incremento in due successivi rilievi. E' stata inoltre valutata la produzione espressa come peso fresco della porzione epigea della piante raccolte in un'area 20x20 cm su ciascun bancale in 4 replicazioni.

Sono state condotte due prove, la prima dal 19/05 al 27/07/2016, la seconda dal 19/09 al 28/10/2016. Ogni esperimento ha generato due insiemi di dati (dataset), uno relativo al bancale 1, l'altro relativo al bancale 2, ciascuno contenente una matrice di variabili ambientali generate dai sensori e un vettore (sequenza) di valori di incidenza della peronospora (% di foglie infette). I sensori avevano un intervallo di campionamento di 10 minuti e hanno permesso di misurare le grandezze riportate in tabella 1.

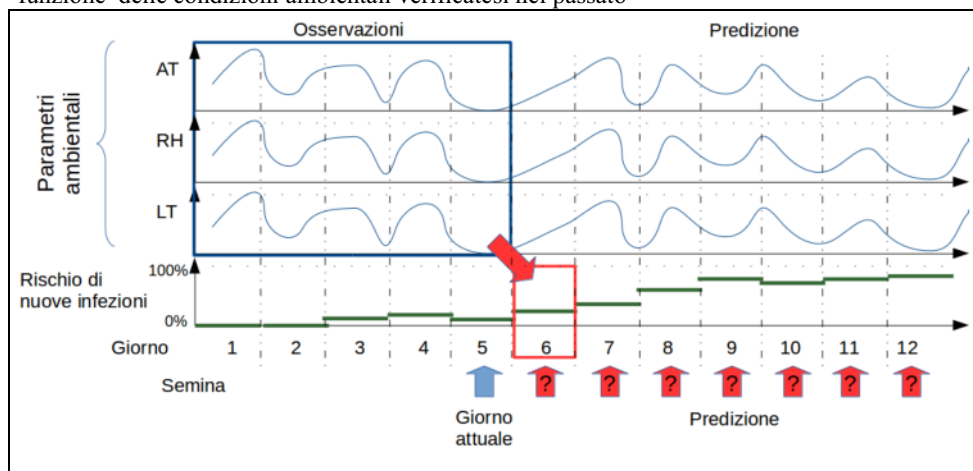
Tabella 1. Variabili del dataset

#	Nome	Descrizione
1	AT	Temperatura dell'aria [°C]
2	RH	Umidità relativa [%]
3	AH	Umidità assoluta [Kg/m ³]
4	DWP	Dew point [°C]
5	BT	Temperatura bancale [°C]
6	LT	Temperatura fogliare [°C]
7	IL	Illuminanza [lx]
8	PPFD	Photosynthetic photon flux density [$\mu\text{mol}/\text{m}^2\text{s}$]
9	RAD	Radiazione [W/m ²]
10	UVA	Radiazione UVA [W/m ²]

D. Modellizzazione: definizione del problema

In considerazione del dataset descritto nella precedente sezione, l'obiettivo è stato quello di definire dei modelli in grado di predire il rischio di nuove infezioni di peronospora in funzione dall'evoluzione nel tempo dei parametri ambientali. La Figura 2 mostra una rappresentazione grafica del problema. Essa rappresenta, in particolare, una situazione in cui l'agricoltore al quinto giorno dopo la semina desidera conoscere il rischio di nuove infezioni data l'evoluzione delle condizioni ambientali degli ultimi cinque giorni.

Figura 2. Definizione del problema: il rischio di nuove infezioni di peronospora è calcolato in funzione delle condizioni ambientali verificatesi nel passato



E. Modellizzazione: generazione dei predittori

Un problema fondamentale nella generazione dei modelli predittivi concerne l'identificazione di predittori informativi. Da un punto di vista agronomico, i predittori rappresentano condizioni legate allo sviluppo e diffusione della malattia. I nostri modelli puntano ad identificare relazioni tra predittori (le variabili indipendenti) e rischio di nuove infezioni (la variabile dipendente).

I predittori sono stati generati aggregando osservazioni appartenenti alle serie temporali dei parametri ambientali seguendo regole proposte da esperti oppure semplicemente suddividendo l'intero intervallo di valori di ciascun parametro ambientale in sottointervalli. La tabella 2 mostra la lista di predittori generata in questo modo. Il predittore $AT_{21-25}RH_{80-100}D_{1-10}$, ad esempio, rappresenta il numero di osservazioni in cui negli ultimi 10 giorni (D) la temperatura dell'aria (AT) assumeva un valore tra 21°C e 25°C ed allo stesso tempo l'umidità relativa (RH) assumeva un valore tra l'80% ed il 100%. Il predittore $AT-DEW_{<2}D_{1-10}$ rappresenta invece il numero di osservazioni in cui negli ultimi 10 giorni (D) la differenza tra la temperatura dell'aria ed il dew point (AT-DEW) è stato inferiore a 2°C. Il significato degli altri predittori può essere compreso dalla loro nomenclatura utilizzando la stessa codifica. I primi cinque predittori (nelle righe 1-3 della Tabella 2) sono stati suggeriti da esperti in accordo con la conoscenza nota in letteratura (Garibaldi et al., 2007; Cohen et al., 2013; Elad et al., 2016), mentre gli altri predittori sono stati realizzati seguendo l'approccio definito in Aegerter et al., 2003. In particolare, secondo quest'ultimo approccio, l'intero intervallo di

temperature osservate (15-35°C) è stato suddiviso in intervalli di 5°C e l'intero intervallo di umidità relativa osservata (65-95%) è stato suddiviso in sotto-intervalli del 10%. Inoltre sono stati aggiunti quattro predittori in cui l'intero intervallo di temperature è suddiviso in sotto-intervalli di 5°C considerando però solo le osservazioni con alta umidità relativa (80%-95%), e quattro predittori che considerano le stesse condizioni ma con la temperatura della foglia al posto di quella dell'aria. Questi ultimi sono un tentativo di sostituire dei predittori basati sulla bagnatura fogliare che non sono disponibili nel nostro dataset ma hanno dimostrato un elevato potere predittivo nella predizione del rischio di peronospora della rosa (Aegerter et al., 2003). Si sta attualmente valutando l'aggiunta di nuovi sensori all'architettura di acquisizione dati per monitorare tale parametro in esperimenti futuri. Solo un piccolo sottoinsieme dei predittori sin qui descritti sono stati inseriti nei nostri modelli. La loro selezione si è basata su metodi statistici descritti nella prossima sezione.

Tabella 2. Predittori a disposizione per la generazione dei modelli

AT ₂₁₋₂₅ RH ₈₀₋₁₀₀ D ₁₋₁₀	AT-DEW _{<2} D ₁₋₁₀
AT ₂₁₋₂₅ D ₁₋₁₀	LT-DEW _{<2} D ₁₋₁₀
LT-DEW _{<4} D ₁₋₁₀	AT ₁₅₋₂₀ D ₁₋₁₀
AT ₂₀₋₂₅ D ₁₋₁₀	AT ₂₅₋₃₀ D ₁₋₁₀
AT ₃₀₋₃₅ D ₁₋₁₀	RH ₆₅₋₇₅ D ₁₋₁₀
RH ₇₅₋₈₅ D ₁₋₁₀	RH ₈₅₋₉₅ D ₁₋₁₀
AT ₁₅₋₂₀ RH ₈₀₋₉₅ D ₁₋₁₀	AT ₂₀₋₂₅ RH ₈₀₋₉₅ D ₁₋₁₀
AT ₂₅₋₃₀ RH ₈₀₋₉₅ D ₁₋₁₀	AT ₃₀₋₃₅ RH ₈₀₋₉₅ D ₁₋₁₀
LT ₁₅₋₂₀ D ₁₋₁₀	LT ₂₀₋₂₅ D ₁₋₁₀
LT ₂₅₋₃₀ D ₁₋₁₀	LT ₃₀₋₃₅ D ₁₋₁₀
LT ₁₅₋₂₀ RH ₈₀₋₉₅ D ₁₋₁₀	LT ₂₀₋₂₅ RH ₈₀₋₉₅ D ₁₋₁₀
LT ₂₅₋₃₀ RH ₈₀₋₉₅ D ₁₋₁₀	LT ₃₀₋₃₅ RH ₈₀₋₉₅ D ₁₋₁₀

F. Modellazione: regressione logistica con regolarizzazione

La regressione logistica regolarizzata è stata utilizzata per predire il rischio di nuove infezioni dati i valori dei predittori, poiché la variabile dipendente in questo problema è binaria, con valore 1 nel caso di “nuove infezioni” e 0 nel caso di “nessuna nuova infezione”, come descritto in Aegerter et al. (2003). Sono stati utilizzati metodi di regolarizzazione (Hastie et al., 2016) per selezionare i predittori più informativi tra quelli nella lista di Tabella 2. Sia Y , la variabile dipendente binaria (valori 0/1), ed X un vettore contenente i valori di p predittori, il modello di regressione logistica esprime la probabilità condizionale della classe (cioè 0/1) per mezzo di una combinazione lineare dei predittori:

$$\begin{aligned}
 P(Y = 0 \mid x) &= \frac{1}{1 + e^{-(\beta_0 + x^T \beta)}}, \\
 P(Y = 1 \mid x) &= 1 - P(Y = 0 \mid x) = \frac{1}{1 + e^{+(\beta_0 + x^T \beta)}}
 \end{aligned}
 \tag{1}$$

le quali implicano che:

$$\log \frac{P(Y=0|x)}{P(Y=1|x)} = \beta_0 + x^T \beta. \quad (2)$$

Il modello è generato calcolando i parametri (β_0, β) tramite massimizzazione detta di *log-likelihood* penalizzata:

$$\max_{(\beta_0, \beta) \in \mathbb{R}^{p+1}} \left[\frac{1}{N} \sum_{i=1}^N \{ I(y_i=0) \log p(x_i) + I(y_i=1) \log (1-p(x_i)) \} - \lambda P_\alpha(\beta) \right] \quad (3)$$

dove I è la funzione che ritorna il valore di verità (valori 0/1) del suo argomento, y_i è l' i -esima osservazione della peronospora (anch'essa binaria), x_j è l' i -esima osservazione dei parametri ambientali, $p(x_i) = P(Y=0 | x_i)$ rappresenta la probabilità nell'equazione (1) per l'osservazione i ad un particolare valore dei parametri (β_0, β) , λ è un parametro che controlla il contributo del termine di regolarizzazione e P_α è il termine di regolarizzazione, un compromesso tra la penalità utilizzata in ridge regression ($\alpha=0$) e la penalità utilizzata in LASSO ($\alpha=1$) (Tibshirani, 1994).

G. Preprocessamento dei dati

I dati grezzi sono stati caricati da file di log dei sensori e fusi in una singola matrice dalla quale tutti i predittori sono stati calcolati. Valori anomali dovuti a guasti del sistema di acquisizione sono stati rimossi e valori mancanti dell'incidenza di peronospora sono stati interpolati per mezzo di interpolazione logistica poiché una crescita di tipo logistico è tipica nelle epidemie su piante (Schumann, 1991). A causa di queste assunzioni, confermate anche dai dati sperimentali, in tutti gli esperimenti qui analizzati la variabile dipendente possiede valore 0 (nessuna nuova infezione) dall'inizio dell'esperimento fino alla prima osservazione della peronospora, e valore 1 (nuova infezione) a partire dalla prima osservazione della peronospora fino al termine dell'esperimento. Il modello presentato nella prossima sezione è stato addestrato con dati a partire dal 15° giorno dalla semina, poiché prima di questo momento la presenza del patogeno è stata assunta come non visibile. Queste assunzioni sembrano appropriate per la nostra configurazione sperimentale poiché i due esperimenti presi in esame possiedono condizioni ambientali simili e di conseguenza i trend di crescita delle piante sono simili. Strategie più complesse, basate sulle reali dimensioni delle piante dovranno essere prese in considerazione in caso di condizioni ambientali non omogenee.

RISULTATI

Il modello predittivo presentato di seguito è stato addestrato considerando i dati di entrambi gli esperimenti a disposizione sul bancale 2 (substrato non riscaldato) in modo da massimizzare la quantità di informazione inclusa nel modello stesso. Tutti i predittori di Tabella 2 sono stati passati all'algoritmo di stima del modello ed una penalità di tipo LASSO ($\alpha=1$ nell'equazione (3)) è stata utilizzata per generare un modello logistico multivariato in grado di classificare “nuova infezione” e “nessuna nuova infezione”. L'effetto di diversi parametri λ (vedi equazione (3)) sulle prestazioni del modello è stato valutato in termini di devianza binomiale in uno schema di cross-validazione con *10-fold*. Il percorso di

regolarizzazione generato è mostrato nella parte superiore di Figura 3. Esso mostra la relazione tra il valore di λ , il numero di variabili del modello e le sue prestazioni. Analizzando tutti i modelli nel percorso di regolarizzazione si è osservato che, anche se il modello con minima devianza binomiale di cross-validazione possiede cinque variabili, l'errore di misclassificazione dei modelli con due e tre variabili (chiamati rispettivamente M_2 ed M_3 di seguito) continua a mantenere valori molto bassi. La Tabella 3 mostra i coefficienti di questi modelli e le loro prestazioni in termini di Matthews Correlation coefficient (MCC).

Tabella 3. Coefficienti e prestazioni de modelli M_2 ed M_3

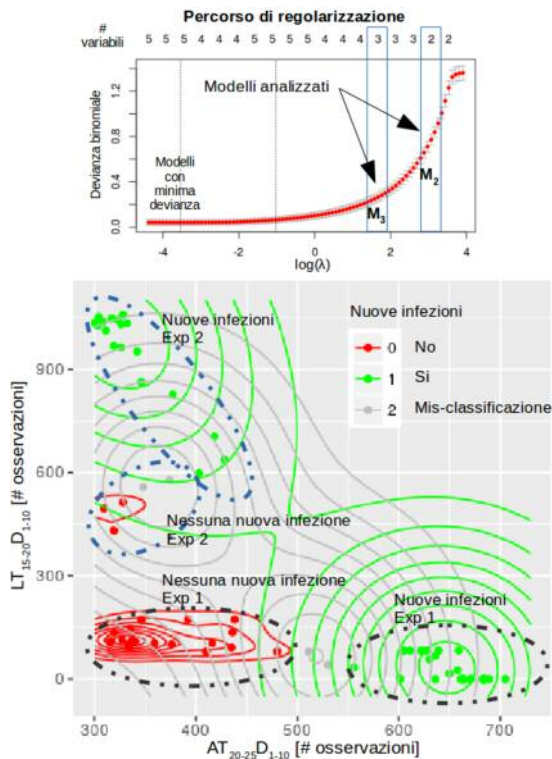
Predittore	Coefficienti modello M2	Coefficienti modello M3
Intercetta (β_0)	-6.5210	-19.4619
$AT_{20-25}D_{1-10}$	0.0123	0.0410
$LT_{15-20}D_{1-10}$	0.0043	0.0129
$LT_{20-25}D_{1-10}$	-	-0.0056
Prestazioni (MCC)	0.88	1.0

Il modello M_2 , il più sparso, non classifica perfettamente il training set ma possiede comunque ottime prestazioni (MCC=0.88). Inoltre, esso utilizza due predittori di interesse: *i*) il numero di osservazioni in cui la temperatura dell’aria ha avuto valori tra i 20°C ed i 25°C negli ultimi dieci giorni ($AT_{20-25}D_{1-10}$), il cui effetto di promozione della peronospora del basilico confermato dalla letteratura (Garibaldi et al., 2007; Elad et al., 2016); *ii*) il numero di osservazioni in cui la temperatura fogliare ha avuto valori tra i 15°C ed i 20°C negli ultimi dieci giorni ($LT_{15-20}D_{1-10}$), che mostra come la temperatura della foglia abbia un’alta rilevanza rispetto al rischio di nuove infezioni. In particolare, il secondo predittore suggerisce sia nuovi esperimenti scientifici che nuove strade per la predizione della peronospora per mezzo di sensori economici. Si osserva che i modelli meno sparsi nel percorso di regolarizzazione (non riportati nella presente pubblicazione) contengono anche altre variabili di interesse, come ad esempio il numero di osservazioni in cui l’umidità relativa ha assunto valori tra il 65% ed il 75% negli ultimi 10 giorni ($RH_{65-75}D_{1-10}$) che è stato dimostrato avere un’influenza negativa sullo sviluppo della peronospora del basilico (Elad et al., 2016).

La parte inferiore di Figura 3 fornisce una visione più approfondita del modello M_2 . Essa mostra la distribuzione de punti del training set proiettati nello spazio bidimensionale dei predittori $AT_{20-25}D_{1-10}$ (asse x) e $LT_{15-20}D_{1-10}$ (asse y). I punti di colore rosso rappresentano giorni correttamente classificati come “nessuna nuova infezione” dal modello M_2 . I punti di colore verde rappresentano giorni correttamente classificati come “nuova infezione” dal modello M_2 . I punti di colore grigio rappresentano mis-classificazioni. Risulta chiaro che il predittore $AT_{20-25}D_{1-10}$ è in grado di classificare i punti dell’esperimento 1 (nella parte bassa del grafico) poiché tutti i punti corrispondenti a “nessuna nuova infezione” possiedono valori inferiori a 500 per quel predittore (si noti che in questo contesto un’osservazione rappresenta un campione acquisito dai sensori ogni 10 minuti, quindi 500 osservazioni corrispondono a circa 83 ore) mentre tutti i punti corrispondenti a “nuova infezione” hanno valori maggiori di 500 per quel predittore. D’altra parte, il predittore $LT_{15-20}D_{1-10}$ è in grado di classificare i punti dell’esperimento 2 (nella parte alta del grafico) poiché tutti i punti corrispondenti a “nuova infezione” possiedono valori inferiori a 600 (circa 100 ore) per quel predittore mentre tutti i punti corrispondenti a “nuova infezione” hanno valori maggiori di 600 per quel predittore. Il predittore $AT_{20-25}D_{1-10}$ però non fornisce buone prestazioni per l’esperimento 2 ed il predittore $LT_{15-20}D_{1-10}$ non fornisce buone prestazioni per l’esperimento 1. Questo

comportamento suggerisce la presenza di condizioni multiple (possibilmente dipendenti) nella promozione di nuove infezioni di peronospora. Nell’esperimento 1, ad esempio, la causa di nuove infezioni sembra essere legata alla lunga permanenza della temperatura dell’aria nell’intervallo tra 20°C e 25°C, mentre nell’esperimento 2 tale causa sembra essere legata alla lunga permanenza della temperatura fogliare nell’intervallo tra i 15°C ed i 20°C (poiché la temperatura dell’aria non sta nell’intervallo tra i 20°C ed i 25°C per lunghi intervalli di tempo in tale esperimento). Questa ipotesi dovrà essere validata da ulteriori esperimenti (attualmente in corso) ma i risultati attuali si considerano molto promettenti poiché mostrano chiare regolarità nelle distribuzioni dei dati di “nuova infezione” e “nessuna nuova infezione”. Tali regolarità solitamente sono visibili solo quando si considerano predittori informativi rispetto al fenomeno in esame.

Figura 3. Modelli addestrati sugli esperimenti 1 e 2, bancale 2. In alto, il percorso di regolarizzazione e le relative posizioni dei modelli M_2 ed M_3 . In basso, distribuzione dei dati proiettati sullo spazio bidimensionale definito dai predittori $AT_{20-25}D_{1-10}$ e $LT_{15-20}D_{1-10}$



CONCLUSIONI

Questo lavoro presenta una architettura basata su sensori e metodi di analisi statistica ed intelligenza artificiale applicati al contenimento della peronospora del basilico in serra. Un modulo fondamentale di questa architettura è rappresentato dal modello matematico che predice il rischio della malattia a partire dall’analisi di serie temporali di parametri ambientali

acquisiti per mezzo di opportuni sensori. Il modello preliminare qui presentato suggerisce possibili relazioni tra le condizioni ambientali o della pianta ed il rischio di nuove infezioni ma richiede ulteriori esperimenti per essere validato ed esteso. Per migliorare questo modello si stanno eseguendo al momento nuovi esperimenti e considerando predittori basati su nuovi parametri, come la bagnatura fogliare e l'intensità luminosa. Metodi automatici per la generazione di tali predittori sono allo studio. Diversi metodi di modellazione, come le reti Bayesiane, saranno considerate nel prossimo futuro e le loro prestazioni confrontate con quelle dell'approccio attuale qui presentato.

LAVORI CITATI

- Aegerter B.J., Nunez J.J., Davis R.M., 2003. Environmental factors affecting rose downy mildew and development of a forecasting model for a nursery production system. *Plant Disease*, 87 (6), 732–738
- Behmann J., Mahlein A., Rumpf T., Römer C., Plümer L., 2015. A review of advanced machine learning methods for the detection of biotic stress in precision crop protection,” *Precision Agriculture*, 16 (3), 239–260
- Castellini A., Farinelli A., Minuto G., Quaglia D., Secco I., Tinivella F., 2017. Expo-Agri: Smart Automatic Greenhouse Control. *Atti 2017 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, 476 – 479
- Cohen Y., Vaknin M., Ben-Naim Y., Rubin A.E., 2013. Light suppresses sporulation and epidemics of *Peronospora belbahrii*. *PLOS-ONE*, 8 (11), e81282, 1-12
- Elad Y. , Omer C., Nisan Z., Harari D., Goren H., Adler U., Silverman D., Biton S., 2016. Passive heat treatment of sweet basil crops suppresses *Peronospora belbahrii* downy mildew. *Annals of Applied Biology*, 168, 373 - 389
- Garibaldi A., Bertetti D., Gullino M.L., 2007. Effect of leaf wetness duration and temperature on infection of downy mildew (*Peronospora* sp.) of basil. *Journal of Plant Diseases and Protection*, 114 (1), 6–8
- Hastie T., Tibshirani R., Friedman J., 2016. The elements of statistical learning. Springer, Stati Uniti, 745 pp.
- Jensen A.L., Jensen F.V., 1996. MIDAS: An influence diagram for management of mildew in winter wheat. *Atti Twelfth Annual Conference on Uncertainty in Artificial Intelligence*, 349–356.
- Kassim M. R. M., Mat I., Harun A.N., 2014. Wireless sensor network in precision agriculture application. *Atti International Conference on Computer, Information and Telecommunication Systems*, 1–5
- Kim K.S., Beresford R.M., Walter M., 2014. Development of a disease risk prediction model for downy mildew (*Peronospora sparsa*) in boysenberry. *Phytopathology*, 104 (1), 50–56
- Maciejowski J., 2001. Predictive control: with constraints. Pearson, Stati Uniti, 352 pp.
- Schumann G.L., 1991. Plant diseases: their biology and social impact. *American Phytopathological Society Press*, Stati Uniti, 397 pp.
- Tibshirani R., 1994. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society, Series B*, 58, 267–288