Pest Outbreaks

1st Mahmut Osmanovic
dept. Computer Science
Jönköping School of Engineering.
Jönköping, Sweden
osma21nr@student.ju.se

2nd Isac Paulsson
dept. Computer Science
Jönköping School of Engineering.
Jönköping, Sweden
pais20dv@student.ju.se

3rd Sebastian Tuura
dept. Computer Science
Jönköping School of Engineering.
Jönköping, Sweden
tuse20tz@student.ju.se

4th Simon De Reuver
dept. Computer Science
Jönköping School of Engineering.
Jönköping, Sweden
desi20kt@student.ju.se

5th Ivo Österberg Nilsson
dept. Computer Science
Jönköping School of Engineering.
Jönköping, Sweden
osiv20jw@student.ju.se

***Abstract*—This study investigates the application of various machine learning (ML) methods for predicting disease severity in crops for the upcoming week. A comprehensive dataset spanning from 2016 to 2023 was acquired, including crop-related data, pest population metrics, and meteorological data from Sweden. These features were used to train multiple ML models, including HistGradientBoostingRegressor, XGBRegressor, and several deep learning (DL) architectures, such as Feedforward Neural Networks, Transformers, and Long Short-Term Memory (LSTM) networks. A DecisionTreeRegressor served as the baseline model. The models were evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and mean squared error (MSE).**

Feature importance analysis revealed that the first-order lagged target had the most significant impact on predicting a specific disease severity. Other highly or moderately correlated features to the target included the cumulative weather parameters and their lagged values, as well as the geographic location of a particular field.

Experimental results show that HistGradientBoostingRegressor achieved the highest predictive accuracy with a R^2 of 0.88, MAE of 0.89 and MSE of 11.79.

This study shows the potential of machine learning in disease severity forecasting which could lead to more effective and data-driven decisions in agricultural management strategies. Furthermore, it stresses that data driven decisions can enable farmers to take proactive measures against pest outbreaks and prevent losses of a harvest due to pests, benefiting both the farmers and the population as a whole.

***Index Terms*—Pest Forecasting, Machine Learning, Data Science, Artificial Intelligence, Deep Learning, Neural Networks, Predictive Analytics, Time Series Analysis, Crop Protection.**

I. Introduction

Combating pests is crucial for maximizing crop yields during harvest. Pests and crop related diseases are a leading cause of yield reduction. Not only that, inefficient use of pesticides leads to increased economical losses and environmental pollution; it may also contribute to the breakdown of resistance in crops such as wheat [1]. With recent advancements in technology, machine learning (ML), and especially deep learning (DL) new opportunities have emerged to address these challenges. In our study, we employ ML and DL to forecast disease severity up to one week in advance, using a comprehensive dataset that combines historical agricultural records with historical weather data. Notably, the agricultural records were provided by the Swedish Board of Agriculture (Jordbruksverket), ensuring that our data is both reliable and extensive. This combined approach captures the dynamic reality between environmental factors and pest behavior, enabling our models to predict upcoming disease severity and empower farmers to take proactive measures, such as optimized pesticide applications, to significantly mitigate pest damage and enhance crop yields.

As previously mentioned, environment plays a significant role in both crop and pest development. In this study, we have combined agricultural data with weather data to create an enhanced and enriched dataset. Weather conditions, such as temperature, precipitation, humidity, wind speed, and solar radiation, not only influence crop growth but also affect pest behavior and reproduction cycles. However, integrating weather data presents challenges. While historical weather data is well-documented and reliable, forecast data, which is used for prediction, inherently contains uncertainties. Forecast models are subject to errors and natural variability, meaning that predictions based on forecasted weather conditions may deviate from actual outcomes.

In this paper, we present week-by-week pest forecasting models that integrate both historical agriculture data and weather variables. Our primary contributions include: (1) a comprehensive dataset of weekly pest observations and meteorological data for crops in Sweden, (2) evaluating a number of machine learning models, including gradient boosting and neural network architectures, for one-week-ahead pest prediction, and (3) assessing the influence of weather features on model accuracy. The rest of this paper is structured as follows: Section II reviews related work, while Section III provides relevant background to the study. In Section IV, we describe the methodology, including data collection and model selection. Section V presents the results and analysis of our experiments. Finally, Section VI discusses potential directions for future research and concludes the paper.

II. Related Work

ML has become increasingly relevant in agricultural forecasting, particularly in modeling plant diseases and pest outbreaks using environmental and temporal data.

Domingues et al. provided a broad survey of ML applications in agriculture, highlighting CNNs, LSTMs, and the integration of environmental data in disease detection and forecasting [2]. Shahoveisi et al. evaluated ML models for predicting *Sclerotinia sclerotiorum* in canola and dry bean using air temperature and leaf wetness. ANN models outperformed logistic regression and other methods in both classification and regression tasks, suggesting ANN's suitability for disease risk modeling [3]. Abuley et al. predicted Alternaria risk in potatoes using aerobiological and weather data. CART models achieved 86% accuracy when prior conidia levels were included, emphasizing the importance of lagged features—an insight that aligns with our own findings [4]. Skelsey applied anomaly detection models using only disease outbreak data to forecast potato late blight [5] . Their ensemble of unsupervised methods achieved high accuracy with fewer false alerts, showing promise for data-limited settings. Kaundal et al. used SVMs for rice blast forecasting based on weekly weather averages [6]. Their models outperformed traditional regression and neural networks. While these studies demonstrate the power of ML in crop disease prediction, most focus on a single pest or disease, or specific model types. Our work expands on these by evaluating a wide range of ML models, including Transformer and LSTM architectures, across multiple pest-crop combinations in Sweden using weekly pest and weather data. [7] proposed weather-based statistical models for forecasting crop yields and pest outbreaks across different regions in India. Their approach combined composite weather indices with regression techniques and neural networks, demonstrating that forecasts could be made weeks ahead of harvest.

III. Background

The task to predict disease severity on crops was assigned by *The Swedish Board of Agriculture (Jordbruksverket)*. Jordbruksverkets API *Prognos och Varnings API*, contains data regarding crop growth and disease spread over time. This data was subsequently examined and explored in conjunction with SMHI's weather data, gathered from their *Meterological API*. The objective was open ended, we decided on predicting the value of the disease severity per week using the sources named above.

A. Agricultural Pest Challenges

Crop pests are a major challenge to agricultural productivity worldwide and in regions like southern Sweden where temperature varies heavily. Pests significantly reduce crop yields and profits. Globally, an estimated 20–40% of crop production is lost annually due to pests and diseases, representing around US$220 billion in economic losses each year [8]. Crops like winter wheat showed about a 6–7% yield reduction on average when fungal diseases were not controlled [9].

Beyond yield impacts, crops severely affected by pests incur substantial management costs and environmental consequences. Farmers often rely on chemical pesticides to protect crops, but heavy use of these compounds can degrade environmental quality. Furthermore, pests' ability to develop resistance is a growing concern. Over 600 pest species worldwide have evolved resistance to at least one pesticide, undermining the efficacy of standard crop protection measures [10]. Together, these challenges highlight the need for more sustainable pesticide usage strategies to secure yields, and reduce both economic losses and harmful environmental impacts.

B. Role of Weather Data in Pest Prediction

Weather data plays a crucial role in predicting pest behavior and crop growth. Variables such as temperature, precipitation, humidity, wind speed, and solar time significantly influence both crops and pests. Integrating weather data into the agriculture data presents several challenges. One major issue is data granularity, as weather conditions can vary significantly over small distances and short periods. This variability makes it difficult to obtain accurate and representative data for specific locations. Additionally, weather measurements inherently contain uncertainties, particularly in forecast data. Forecast models are subject to errors and natural variability, which can lead to deviations between predicted and actual weather conditions [11]. These uncertainties affect our models reliability and accuracy.

In conclusion, although integrating weather data is challenging due to issues of granularity and uncertainty, it remains essential for enriching datasets and, for the models, capturing the underlying patterns and relations.

IV. Method

Data was sourced from Jordbruksverkets *Prognos och Varnings API* as well as SMHI's *Metrology API*, with the intent of exploring how weather variables affect the growth of diseases on different crops. The following models where tested *HistGradientBoostingRegressor, XGBRegressor, FeedForwardNeuralNetworks, SupportVectorRegressor, Transformer, LongShortTermMemory and DecisionTreeRegressor*. Subsequently their performance was evaluated using the following metrics *MAE, MSE,* R^2.

A. Data

Data has been gathered from Jordbruksverket in the period 2016-2023, for the crops *Höstvete*, *Rågvete* and *Vårkorn*. *Höstvete* has been filtered to include Bladfläcksvampar, Brunrost, Svartpricksjuka, Gulrost and Mjöldagg. *Rågvete* has been filtered to include Brunrost, Gulrost, Sköldfläcksjuka, Mjöldagg and Bladfläcksvampar. *Vårkorn* has been filtered to include Sköldfläcksjuka, Kornets bladfläcksjuka, Mjöldagg, Havrebladlus and Sädesbladlus. The quantity of datapoints for each dataset is specified by table I.

TABLE I
NUMBER OF DATA POINTS BY DATASET

Crop	Höstvete	Rågvete	Vårkorn
# Measurements	21568	2557	11234

TABLE II
CROP AND ASSOCIATED PESTS

Crop	Pest 1	Pest 2	Pest 3	Pest 4	Pest 5
Höstvete	Bladfläcksvampar	Brunrost	Gulrost	Mjöldagg	Svartpricksjuka
Rågvete	Bladfläcksvampar	Brunrost	Gulrost	Mjöldagg	Sköldfläcksjuka
Vårkorn	Korenets bladfläcksjuka	Mjöldagg	Sköldfläcksjuka		

All variants of bladlus have later been excluded due to low predictive performance. In addition, they exhibit negligibly low correlation with all other targets, including each other. Data from Jordbruksverket has a weekly frequency. Series in the data are given an identifier by inserting an id with *graderingstillfalleList*, this has been named *Series_id*. Series have been filtered to include the 90% most common week spans, generally 9 weeks or more. Furthermore, exploration of how sensitive the different *Höstvete* strains are to the variety of pests was done using the recommendations distributed by Jordbruksverket in *Bekämpningsrekommendationer, svampar och insekter 2024*. This resulted in a better performance of the models, however this was later removed due to too many sensitivity values where missing for the different combinations of crops.

Weather data is gathered from SMHI open data API which contains data from weather stations. Data gathered from SMHI includes *Air temperature (min, max)*, *Precipitation(sum, max)*, *Sunshine time(sum)*, *Dewpoint temperature (mean, max)*, *Relative humidity (mean)* and *Long wave irradiance(mean)*. The values from stations are aggregated by day. Values are collected based on geographic distance from field to weather station. A mean of the values from the closest 3 weather stations is used to approximate daily field weather. Daily weather data is further aggregated to weekly frequency using data from one week prior to the listed measurement date.

Yearly cumulative weather features were calculated by field for *Precipitation(sum)*, *Dewpoint temperature (mean)*, *Relative humidity (mean)*, *Air temperature (max)* and *Sunshine time(sum)* using the gathered weather data within the series, generally from week 15 and forwards [12].

B. Exploratory Data Analysis

Whilst measuring Jordburksverket randomly samples a plant and subsequently chooses the three most prominent leaves of that same plant to measure for pests. Observing figure 1 we can see a plunge in the middle of the growth season for the value of the pests. This is due to the three most prominent leaves changing during growth. Furthermore for the specific example of *Svartpricksjuka* which is measured in percentages the amount might stay the same but the leaves have gotten bigger reducing the % of the leaves being covered.

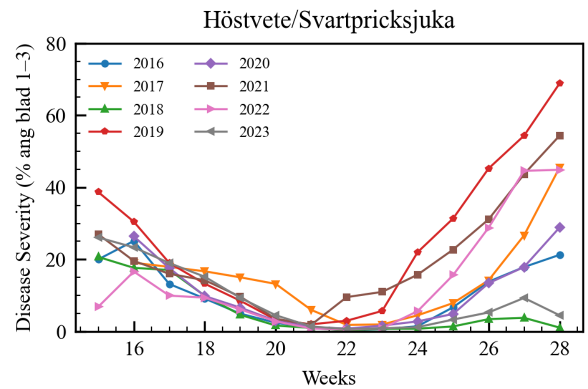

Fig. 1. Disease Severity by Year 2016-2023.

As seen in figure 1 the average disease severity varies significantly by year. 2018 and 2023 stand out as years with relatively low amounts of pest, not exceeding 10% after week 21 where the other years experience an increase in disease severity. We also noticed a reduction in predictive performance for both years.

TABLE III
SCB ANNUAL HARVEST WINTER WHEAT (KG/ACRE)

Year	2015	2016	2017	2018	2019	2020	2021	2022	2023
Harvest (kg/acre)	7570	6680	7360	4790	7730	7450	6600	7220	5750

This correlates well to table III where both 2018 and 2023 are years with a low harvest yield. This is attributed to adverse weather conditions, with droughts occurring in both 2018 and 2023 [13].

TABLE IV
MI COLOR INTERVALS

Color	Red	Blue	Green
MI Intervals	$0.1 <$ nats	$0.1 \leq$ nats ≤ 0.35	nats > 0.35

Observe that the feature color scheme in the mutual information feature plots is highlighted in table IV.

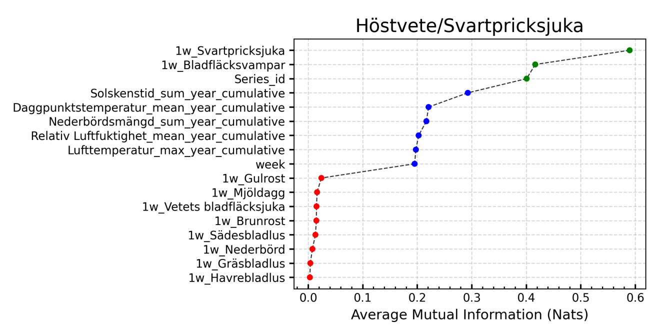

Fig. 2. Mutual Information between Höstvete and Svartpricksjuka features.

A key observation early on was that the lagged value of the target was highly correlated with the target itself. This correlation held true across all crop and pest combinations. Additionally, we observed that the weather parameters did not initially provide any meaningful contribution in predicting our target. Our hypothesis was that the model would benefit from having certain features that could be indicative of weather or not a particular year was about to be exceptionally dry or hot. Cumulative weather parameters were introduced, signified the total amount of, say rain. The value for each of the cumulative parameters was calculated using the gathered weather parameters from the three closest available SMHI weather stations to the farming field. Historical maximum and minimum values were used to scale each weather feature between 0 and 1. Some parameter values were summed over, at each new data point, indicating, say the quantity of rainfall until data date.

We noticed an increase with their inclusion in our training data. Subsequent exclusion of initial weather data demonstrated their superfluousness (did not impact model model performance) and they were excluded.

Another key finding, clearly highlighted in figure 2, and further illuminated for in 3, was that some pests were highly correlated with each other. The rolling spearman correlation coefficient between svartpricksjuka and bladfläcksvampar is chosen as a representative example, where the coefficient generally is found to be positive. Other pests sometimes negatively correlated. Nevertheless, this feature significantly boosted our model performance across all models.

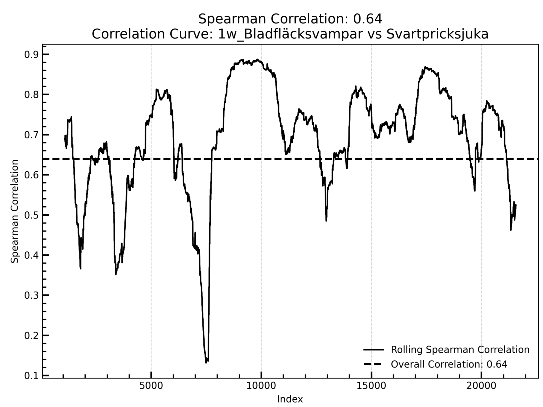

Fig. 3. Rolling Spearman Correlation between svartpricksjuka and the one week lagged value of bladfläcksvampar in the höstvete dataset.

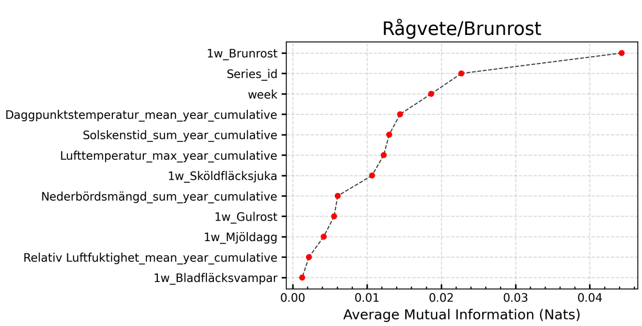

Fig. 4. Mutual Information between Rågvete and Brunrost features.

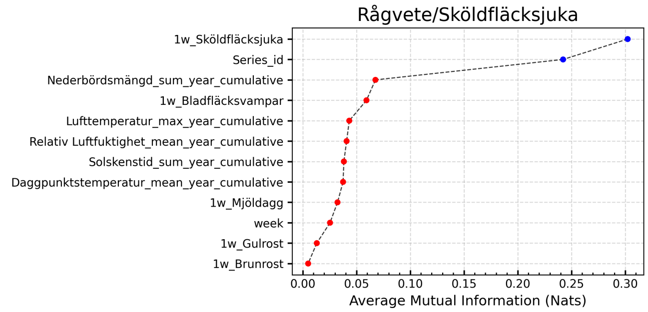

Fig. 5. Mutual Information between Rågvete and Sköldfläcksjuka features.

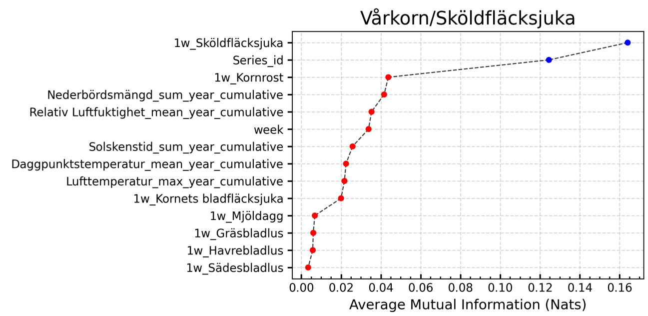

Fig. 6. Mutual Information between Vårkorn and Sköldfläcksjuka.

Average Mutual Information (AMI) is a measure used to quantify the amount of information shared between two variables. By calculating the AMI between features and the target variable we are able to Investigate the amount of information shared between the features and the target. Examining figure 2 shows that for the combination *Höstvete* and *Svartpricksjuka* some features like cumulative weather values, the previous weeks value for both *Svartpricksjuka* and *Bladfläckssvampar* aswell as the week provides a high amount of information on the target. Notably *Series_id* also has a high AMI. This suggests that there is a lot of variability between the fields.

C. Model Selection

The following models where selected and subsequently their performance was evaluated against each other: *HistGradientBoostingRegressor, XGBRegressor, FeedForwardNeuralNetworks, SupportVectorRegressor, Transformer, LongShortTermMemory* and finally the baseline model *DecisionTreeRegressor*. Gradient boosted models have been show to be highly effective for structured tabular data. XGBoost in particular, has proven to be highly effective for regression tasks [14]. Whereas the DecisionTreeRegressor model serves a good baseline model due to it's simplicity and interpretability. Furthermore FeedForwardNeuralNetwork was tested due to its capability of learning complex and non linear relationships in the data. LongShortTermMemory models have been shown to be good at handling time-dependent processes [15]. Finally Transformer and SupportVectorRegressor, as for Transformer models have demonstrated improved performance over recurrent models in certain time-series applications [16]. LSTM and Transformer models are not included in the final results due to lack of adequate tuning.

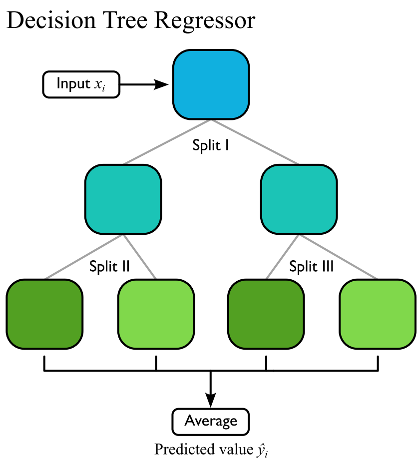

Fig. 7. Structure of Decision tree regressor

Decision tree Regressors, detailed in figure 7 are a tree like model created by conditionally evaluating features to find suitable splits. For regression good splits are those who maximize variance reduction in sub-nodes, leaves are assigned the average of training targets that finish in the leaf. Each splitting node corresponds to a feature and a condition. A prediction is made by following the tree according to feature and condition, upon reaching a leaf the average is assigned as the prediction.

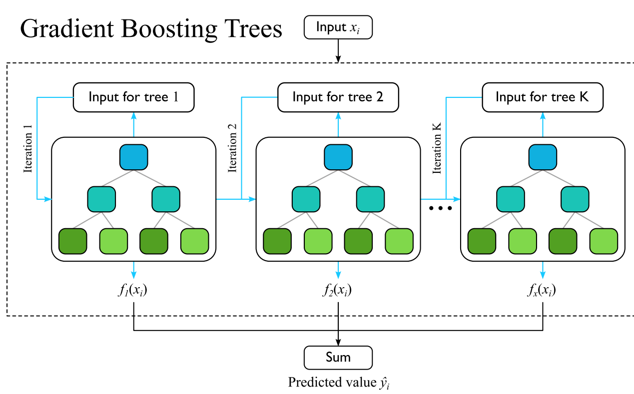

Fig. 8. Structure of Gradient Boosted Trees

Gradient Boosted Trees, detailed in figure 8 is an ensamble model. It is based on training multiple consecutive Decision trees where each tree corrects the residuals of the previous one. For one prediction the outputs of all included decision trees are summed together. Both XGBR and HGBR follow the same basic principle, the main difference is that HGBR bins all data before training. The XGBR module also exposes additional hyper parameters.

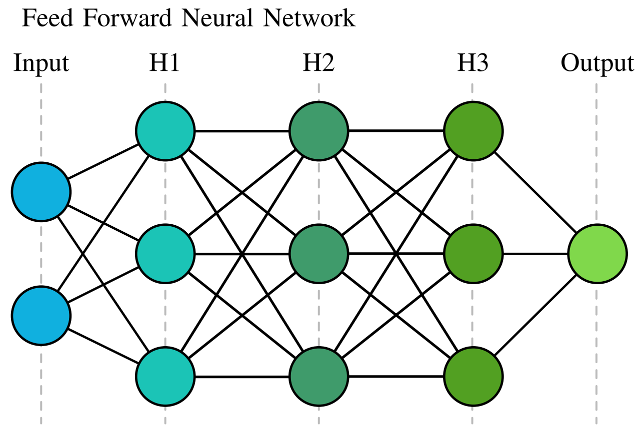

Fig. 9. Structure of Feed Forward Neural Network

FeedForward neural networks is one of the simpler types of neural networks, where in the information only moves forward. Architecturally FeedForward neural networks contains three layers being *input layer, hidden layer & output layer*. Each of the layers are made up by so called *neurons* where they are connected between the layers using *weights*. The *weights* determine how much one *neurons* value affect another *neurons* value. The *neurons* will use *activation functions* such as *ReLU, Leaky ReLU* or *Sigmoid* in order to introduce non-linearity, allowing the network to learn complex patterns. The network is subsequently trained using *back propagation* wherein the network uses *gradient descent* to adjust the *weights*.

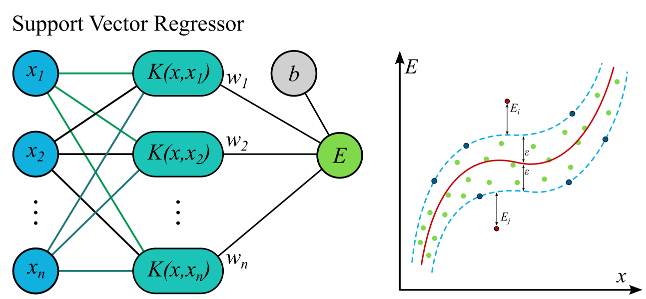

Fig. 10. Structure of Support Vector Regressor

Support Vector Regression is an expansion of Support Vector Machines which introduces a hyperplane around the function to be optimized. The main objective is to determine the optimal hyperplane that closely approximates the continuous function. Support Vector Regressors use *kernels* to transform the input data so that linear relationships can be found, making it easier to fit models. The *Radial Basis Function* is one of the most widely used kernels due to it behaving very similar to gaussian distribution.

$$K(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \tag{1}$$

D. Model Evaluation

In order to provide a representative evaluation of the models, three metrics were selected. MAE (eq 2), MSE (eq 3) and R^2 (eq 4).

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

MAE Provides an intuitive understanding for general predictive performance. Additionally it retains the unit of the target. A weakness in MAE is its focus on general performance, this gives a good representation of the most common cases. A lower value in MAE corresponds to better performance.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{3}$$

MSE Provides a better understanding for predictive performance where the largest errors are given more importance. In our case the distribution of the target has a high positive skew. Therefore MSE provides a good indication of the performance in the less common cases where values in the target are higher resulting in a larger error. This correlates to the most important aspect of our prediction, the peaks. A lower value in MSE corresponds to better performance in the more extreme cases.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

R^2 Measures how well the variance in the target is explained by the model. A higher value in R^2 corresponds to a better fit between the target and the predicted values.

Testing has been performed with 10-fold Cross Validation to ensure robustness in results. To adapt for our case and keep series intact the folds have been created by selecting a percentage of *Series_id*'s without replacement as our test set and training on the remaining data. This ensures that no overlap exists between training and test. This also maintain the integrity of the series and enables testing on the entire dataset. [17]

E. Design of Experiments

To get a detailed and robust result multiple precautions have been taken in the design of the experiment. Table V details the final tuned parameters used for the models.

TABLE V
FINAL PARAMETERS FOR MODELS

Model	Parameter	Value
HistGB	loss	squared_error
	quantile	None
	learning_rate	0.1
	max_iter	100
	max_leaf_nodes	31
	min_samples_leaf	20
	l2_regularization	0.0
	max_features	1.0
	max_bins	255
	early_stopping	auto
DTR	criterion	squared_error
	splitter	best
	max_depth	None
	min_samples_split	2
	min_samples_leaf	1
	min_weight_fraction_leaf	0.0
	ccp_alpha	0.0
FFNN	activation function	elu
	loss	rmse
	hidden_layer_sizes	3*[2*input_shape]
	quantile	0.55
	learning_rate	0.000835
	dropout_rate	0.1
	epochs	12
	batch_size	32
	patience	3
	validation_split	0.15
XGBR	objective	squared_error
	booster	gbtree
	n_estimators	100
	learning_rate	0.3
	max_depth	6
	min_child_weight	1
	gamma	0
	subsample	1
	colsample_bytree	1
	reg_alpha	0
	reg_lambda	1
SVR	C	10
	epsilon	0.01
	gamma	auto
	kernel	rbf
	degree	3
	coef0	0.0

All models have been trained and evaluated for each pest and crop combination. This provides detailed information on the performance every model for every pest and crop. In order to provide similar conditions for all models, the same cross validation splits have been used in all evaluations.

Hyper parameter tuning with cross validation has been performed for all included models before the final evaluation. For HistGB, DTR and XGBR no significant improvement was observed with parameter tuning, hence the default parameters were utilized.

V. RESULT & ANALYSIS

The model performance is detailed through the use of three heatmaps and two tables. Each of the three heatmaps highlights every models performance with respect to one of the three metrics, R^2, MAE and MSE. A blue-yellow color spectrum has been applied where blue signifies a positive outcome and yellow a negative outcome. The heatmaps are constructed in a similar manner, where the y-axis detail a specific crop/pest combination and the x-axis a specific models performance on that combination. Table VI and VII summarizes the results. The rest of the figures are utilized as supplementary results, aiding in the interpretation of the heatmap figures and tables.

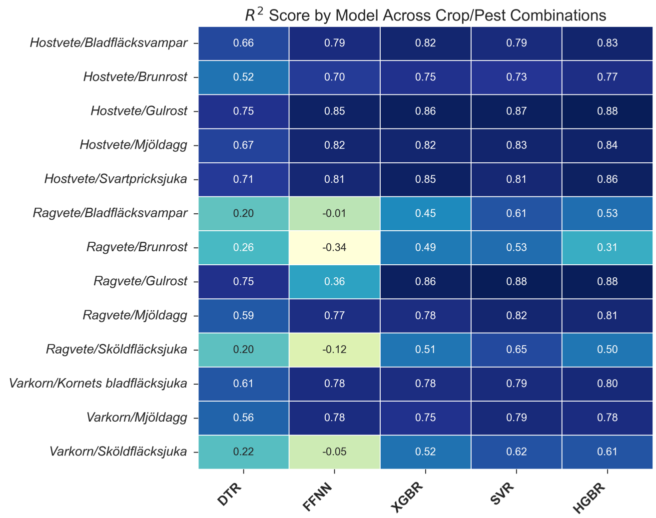

Fig. 11. Vertical dimension details the crop/pest data combination. Horizontal dimension specifies the model in question. The color dimension highlights the R^2 score. Models are ranked from right to left, with the best performing model being placed to the right.

All models demonstrated a satisfactory level of performance for all höstvete and pest combinations. Specifically, more or less all models manage to capture at least $\geq 70\%$ of the variation in the target. The best performing model was HGBR, which managed to achieve an R^2 score for höstvete that approximately varies between 0.80 and 0.90. Its noteworthy that SVR is observed to obtain a better R^2 score for the smallest dataset. Performance is comparable in the medium sized dataset, vårkorn. HGBR performs better than SVR on the largest dataset. Previous research has demonstrated similar outcomes [18, 19]. A reason as to why SVR is able to capture global patterns even with a relatively small dataset is related to its use of support vectors. The support vectors enable the SVR to emphasize on the most important patterns, especially in conjunction with the use of an appropriate kernel function, such as the RBF kernel. With larger datasets, the reliability of most classifiers is increased. That is because the splits are made in such a manner as to reduce both bias and variance. The bias occurs when trying to approximate a target with an overly simplified model, whilst variance refers to errors caused by excessive complexity. The model performance is highly sensitive to the introduction of new data. With the addition of more data, the decision tree splits are further refined, minimizing sensitivity and complexity. Hence, allowing for greater generalization.

Through visual inspection we can extrapolate that our models perform the worst on the rågvete dataset. The the FFNN model was conspicuously under performing. In fact, the FFNN obtained a negative R^2 score two out of three times. Neural networks have the capacity to capture a lot of complexity. Table V specifies the neural network size to being $3\times[2\times\text{input shape}]$, across all experiments. Recall that the rågvete dataset was modestly small, no larger than 2557 data points I. Thus, the model did not obtain an adequate amount of data points to sufficiently capture the variation in the target.

Nonetheless, even the better performing models, SVR and HGBR performed poorly in general. There are several possible explanations for this observation. Firstly, the features in and of themselves may be inadequate predictors of the target. Secondly, there could be too few data points. If the correlative nature between the features and the target is complex, the models will require larger sample sizes of training data. More qualitative training data facilities learning as correlation ostensibly become richer and more nuanced, allowing models to better capture the complexities within the data. Lastly, it could simply be that the models (SVR and HBGR) did not sufficiently exploit the existing feature correlations with the target. Further parameter optimization would likely enhance individual dataset performance. Examination of figure 4 clearly highlights that the lack of mutual information between the features and the target. The entailment being that the features are terrible predictors of the target in this circumstance. However, it may still be that the correlative nature is highly complex and that additional data is required. Given the almost non-existing correlative nature of between the features and the target, we strongly suspect that additional features are required. Two of the other low performing models are related to sköldfläcksjuka as a target. Figures 5 and 6 clearly highlight the insufficiency of meaningful correlations between features and target.

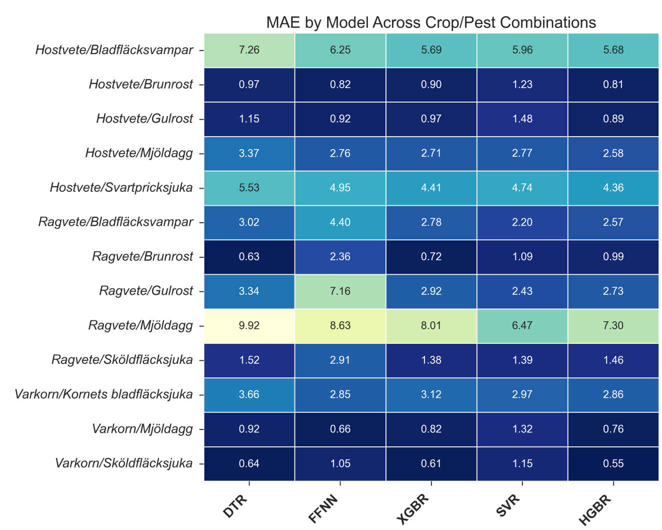

Fig. 12. Vertical dimension details the crop/pest data combination. Horizontal dimension specifies the model in question. The color dimension highlights the mean average error.

The MAE heatmap above (figure 12) details a substantial variance in performance. Some targets, such as gulrost in the höstvete dataset can be predicted with a mean absolute error of 0.55. In other cases, we observe that our models struggle with predictive accuracy. Take for instance mjöldagg in the rågvete dataset, with a high MAE in comparison to the other crop/pest combinations. However, mjöldagg is not as difficult to predict in other crops. In the höstvete data, it is on average, across all models, able to be predicted within an absolute error of 2.83. Figure 13 attempts to contextualize the results.

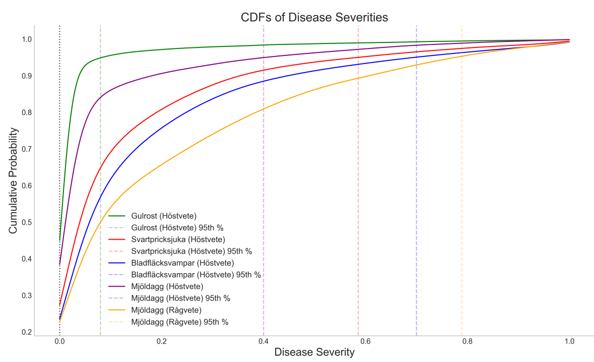

Fig. 13. Cumulative Distribution Functions of Disease Severities

Notice that the pest performing models are trained and tested on data in which exhibit lower amounts of variance (and hence complexity) within the target. The entailment being that it is simpler for the models to predict low variance targets. The prediction of any small pest value number (in the case of the less complex targets) immediately yields a low MAE. In other words, probability of predicting most of the values approximately correctly becomes higher the lower the target variance.

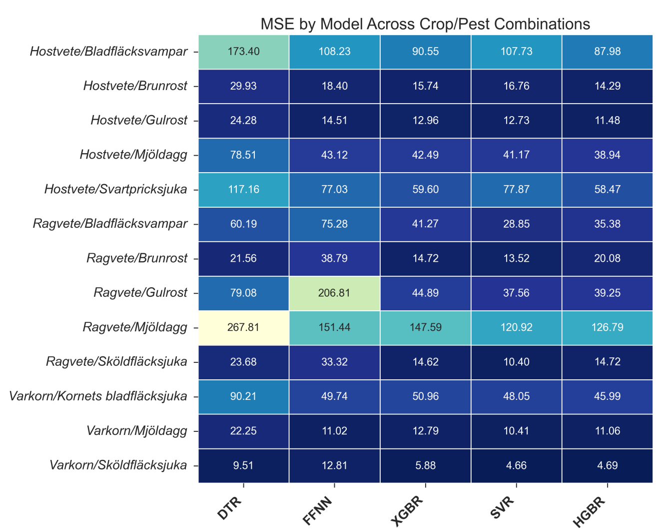

Fig. 14. Vertical dimension details the crop/pest data combination. Horizontal dimension specifies the model in question. The color dimension highlights the mean squared error.

Figure 14 highlights that most of the models perform well (indicated by the vast amounts of blueness) on most of the evaluated combinations.

The MSE error punishes outliers to a greater extent when compared to MAE. Although varying in fraction, there generally exists a lot of values within the targets that are small in magnitude. Their predictions will often be robust in most models with most combinations. The peaks are significantly more difficult but also crucial to correctly predict. Which consequentially renders the MSE metric as a solid proxy for peak over and under estimations.

Observe that again, SVR tends to outperform HGBR (and XGBR) on the smallest dataset, they perform equal in the middle sized dataset and HGBR performs better at the largest (for the reasons previously alluded to).

Its noteworthy that all models comparatively exhibit a deficient performance on the rågvete/mjöldagg combination. However, its not a surprise given the sparseness of the rågvete dataset as well as the complexity in the target. Recall from figure 13 that its target exhibited the greatest amount of variance.

TABLE VI
AVERAGE PERFORMANCE METRICS

Model	R^2	Mean Average Error	Mean Squared Error
HGBR	0.72	2.58	39.16
SVR	0.75	2.71	40.82
XGBR	0.71	2.70	42.62
FFNN	0.47	3.52	64.65
DTR	0.51	3.22	76.74

Table VI details each models average performs across the three chosen metrics across all crop and pest combinations. SVR managed to slightly better capture the average target variance better than HGBR. HGBR performed better on MAE and MSE. It signifies that HGBR on average manage to classify outliers more accurately, hence a better score.

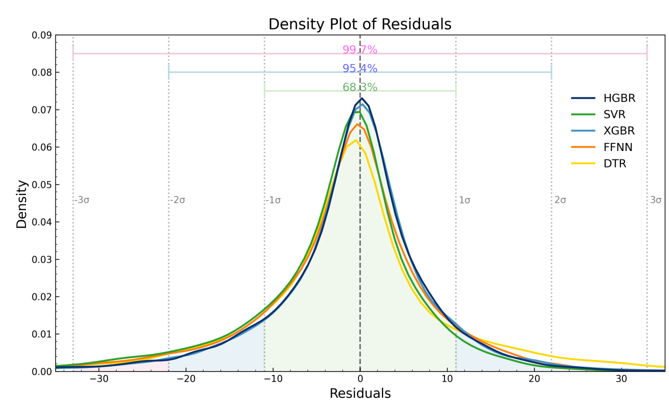

Fig. 15. The density plot highlights each models general performance in terms of over and under predictions. Negative valued residuals specify under-predictions and positive valued residuals detail over-predictions.

The density plot in figure 15 further illuminates the strengths and weaknesses of each model. The residuals are determined by the difference between the predicted and the actual target values. Hence, negative residuals signify under predictions and positive residuals specify over predictions. The ideal density plot for a model is heavily concentrated around 0, having a high peak and thin ends. Note that HGBR has the highest peak. Whilst it is true that SVR over predicts to a slightly lesser extend than HGBR, it also tends to under predict slightly more.

TABLE VII
PERFORMANCE METRICS FOR
HISTGRADIENTBOOSTINGREGRESSOR

Crop	Disease	R^2	MAE	MSE
Höstvete	Bladfläcksvampar	0.83	5.67	87.52
	Brunrost	0.77	0.80	14.40
	Svartpricksjuka	0.85	4.37	58.71
	Gulrost	0.88	0.89	11.79
	Mjöldagg	0.84	2.56	38.45
	Average	**0.83**	**2.86**	**30.43**
Rågvete	Brunrost	0.35	1.01	18.76
	Gulrost	0.88	2.66	37.87
	Sköldfläcksjuka	0.52	1.41	13.97
	Mjöldagg	0.80	7.41	130.43
	Bladfläcksvampar	0.51	2.66	36.86
	Average	**0.61**	**3.03**	**47.58**
Vårkorn	Sköldfläcksjuka	0.60	0.55	4.81
	Kornets bladfläcksjuka	0.80	2.84	45.49
	Mjöldagg	0.78	0.76	11.01
	Average	**0.73**	**1.38**	**20.44**

Table VII summarizes the performance of our best model across all specified combinations. Its performance tends to scale with the amount of data points but is also dependent on the complexity of the target. This is highlighted by the higher MSE values on blackfläckssvampar and svarprickssjuka in höstvete. As previously observed, their targets exhibit significantly more complexity than say, gulrost, that remarkably achieves an R^2 value of of 0.88 and a MSE of 11.79. Combinations that achieve a great R^2 and MAE score generally achieve comparatively high MSE score, this occurs because of a relatively high presence of outliers in the target.

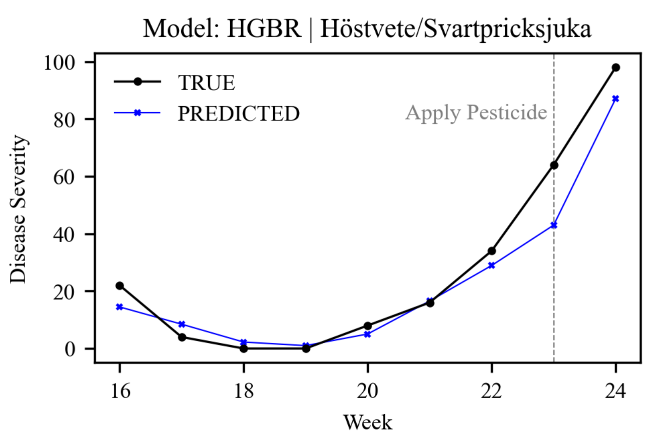

Fig. 16. The y-axis specifies disease severity and the x-axis highlights the current week. The gray vertical line details that pesticide may be applied during week 23.

Figure 16 has been included as a representative example. It highlights the potential usage of our best performing model, HGBR. In particular, it could alarm Jordbruksverket at least one week ahead about sharp raises in the reproduction of a particular pest. Jordbruksverket can with our model already apply pesticide on the highlighted field during week 23, instead of week 24. Note that week 23 corresponds to the recommended range in which pesticide ought to be applied. Our model enables Swedish food products to contain a lesser amount of pesticide (earlier detection of pest outbreak means that a lesser quantity of pesticide is required) whilst simultaneously making possible for farmers to increase their yield of harvest (as fewer crops are damaged).

VI. FUTURE WORK

There are some current issues that could be addressed to further improve results. Distance to weather stations for many of the weather parameters is generally high. A majority is at a distance of over 20 km. The result of this is that weather station data doesn't represent the weather well in a field, causing poor feature performance. To address this we have used a naive method, taking the mean from the three closest weather stations. There are however more advanced methods to estimate weather conditions at a location. Some examples include spatial and spatiotemporal kriging [20]. Another approach is using archived weather forecast data. This is provided by SMHI from their MESAN weather analysis archive. MESAN provides gridded weather forecasts with a spatial resolution of $0.1\,\mathrm{deg}$ or 11 km [21]. This would also allow the use of more weather parameters. Using forecast data would provide the added benefit of training on forecast data which would be required for prediction upon model deployment. This was not implemented due to the size of MESAN data, estimated to be 159 gB.

One feature noted by Jordburksverket to have importance for pest prediction is *leaf wetness duration*. This parameter is not included by SMHI. It could however be estimated from weather data [22].

Another feature that showed promise was sensitivities of the different diseases for the different crop strains. As mentioned the strains of *Höstvete* was tested with regards to the sensitivities. However, this was later removed due to too many of the combinations of disease and strain where missing, specifically for *Vårkorn* and *Rågvete*.

The current models are trained on the entire dataset which has been filtered to include most common week spans. Given the high variance and positive skew in measurement values, outlier detection is challenging. Regular means of outlier detection would remove the majority of data in peak pest growth where prediction is of utmost interest. Future attempts which rigorously try to identify fields with radically different target distributions for the same pests should be analyzed, and either treated separately or dropped. May be attempted using the KL-divergence metric.

References

[1] Shirin Mahmoodi, Meysam Bakhshi Ganje, Kourosh Ahmadi, Yadollah Dalvand, Amir Naghibi, and Nathaniel K. Newlands. Modeling spatiotemporal distribution of yellow rust wheat pathogen using machine learning algorithms: insights from environmental assessment. *Environmental Technology & Innovation*, 2024.

[2] Tiago Domingues, Tomás Brandão, and João C. Ferreira. Machine learning for detection and prediction of crop diseases and pests: A comprehensive survey. *Agriculture*, 12(9), 2022.

[3] Mojtaba Shahoveisi, Bruce D. Gossen, Steven F. Hwang, Tika B. Adhikari, Sabine Banniza, and Luis E. del Río Mendoza. Modeling risk of *Sclerotinia* disease development on canola and dry bean using machine learning. *Frontiers in Plant Science*, 13:846601, 2022.

[4] Isaac K. Abuley, Julie Kikkert, and Christine D. Smart. Importance of meteorological parameters and airborne conidia to predict risk of *Alternaria* in potato using machine learning. *Frontiers in Artificial Intelligence*, 5:923045, 2022.

[5] Peter Skelsey. Forecasting risk of crop disease with anomaly detection algorithms. *Computers and Electronics in Agriculture*, 182:106008, 2021.

[6] Rajeev Kaundal, Amit Kapoor, and G. P. S. Raghava. Machine learning techniques in disease forecasting: A case study on rice blast prediction. *BMC Bioinformatics*, 7(1):485, 2006.

[7] Ranjana Agrawal and S. C. Mehta. Weather based forecasting of crop yields, pests and diseases - iasri models. *Journal of the Indian Society of Agricultural Statistics*, 2007.

[8] Sara Francesconi. High-throughput and point-of-care detection of wheat fungal diseases: Potentialities of molecular and phenomics techniques toward in-field applicability. *Frontiers in Agronomy*, 4, sep 2022.

[9] F. Laidig, T. Feike, B. Klocke, J. Macholdt, T. Miedaner, D. Rentel, and H. P. Piepho. Yield reduction due to diseases and lodging and impact of input intensity on yield in variety trials in five cereal crops. *Euphytica*, 218:150, 2022.

[10] MarketsandMarkets. Pest control market projected to reach $27.5 billion by 2025, oct 2019.

[11] Rabiu Olatinwo and Gerrit Hoogenboom. Chapter 4 - weather-based pest forecasting for efficient crop protection. In Dharam P. Abrol, editor, *Integrated Pest Management*, pages 59–78. Academic Press, San Diego, 2014.

[12] D. E. te Beest, M. W. Shaw, S. Pietravalle, and F. van den Bosch. A predictive model for early-warning of septoria leaf blotch on winter wheat. *European Journal of Plant Pathology*, 124(3):413–425, 2009.

[13] Sveriges Meteorologiska och Hydrologiska Institut. Historiska torrperioder. Accessed: 2023-10-03.

[14] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'16)*, 2016.

[15] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 1997. Fakultät für Informatik, Technische Universität München, Germany and IDSIA, Corso Elvezia 36, 6900 Lugano, Switzerland.

[16] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

[17] Seymour Geisser. The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350):320–328, 1975.

[18] M. Zou, W. G. Jiang, Q. H. Qin, Y. C. Liu, and M. L. Li. Optimized xgboost model with small dataset for predicting relative density of ti-6al-4v parts manufactured by selective laser melting. *Materials*, 15(15):5298, Aug 2022.

[19] Przemyslaw Kupidura, Agnieszka Kepa, and Piotr Krawczyk. Comparative analysis of the performance of selected machine learning algorithms depending on the size of the training sample. *Reports on Geodesy and Geoinformatics*, 1(1):1–10, 2023.

[20] Aamina Batool, Sufian Ahmad, Ayesha Waseem, Veysi Kartal, Zulfiqar Ali, and Muhammad Mohsin. Evaluating variogram models and kriging approaches for analyzing spatial trends in precipitation simulations from global climate models. *Acta Geophysica*, February 2025.

[21] Swedish Meteorological and Hydrological Institute (SMHI). Mesan - an operational mesoscale analysis system. Technical report, Swedish Meteorological and Hydrological Institute, 2006.

[22] Rajkumar Dhakar, Basavaraj R. Amogi, Gajanan S. Kothawade, and Lav R. Khot. Simplified mechanistic model for estimating leaf wetness. *Agricultural and Forest Meteorology*, 363:110399, 2025.