



HAL
open science

Abnormal Behavior Detection in Farming Stream Data

Juliet Chebet Moso, Stephane Cormier, Hacene Fouchal, Cyril de Runz, John Wandeto

► **To cite this version:**

Juliet Chebet Moso, Stephane Cormier, Hacene Fouchal, Cyril de Runz, John Wandeto. Abnormal Behavior Detection in Farming Stream Data. International Conference on Smart and Sustainable Agriculture (SSA), 2021, Virtual conference, France. pp.44-56, 10.1007/978-3-030-88259-4_4. hal-03445069

HAL Id: hal-03445069

<https://hal.science/hal-03445069>

Submitted on 25 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Abnormal behavior detection in Farming stream data

Juliet Chebet Moso^{1,3}, Stéphane Cormier¹, Hacène Fouchal¹, Cyril de Runz²,
and John M. Wandeto³

¹ CReSTIC EA 3804, Université de Reims Champagne-Ardenne, 51097, Reims, France

² BDTLN, LIFAT, University of Tours, Tours, France

³ Computer Science, Dedan Kimathi University of Technology, Nyeri, Kenya
juliet-chebet.moso@etudiant.univ-reims.fr, {stephane.cormier,
hacene.fouchal}@univ-reims.fr, cyril.derunz@univ-tours.fr,
john.wandeto@dkut.ac.ke

Abstract. Farmers have been able to detect, quantify, and respond to spatial and temporal variation in crops thanks to a variety of technological advancements in recent years. Precision farming aims to provide precise targeting of agricultural inputs while reducing waste and negative consequences. Precision agriculture technologies are effective instruments for increasing farm sustainability and production. These technologies provide ways to create more with less resources. Nowadays, improving agricultural production efficiency and crop yields is impossible without the use of contemporary digital technologies and smart machinery. The introduction of high-accuracy GPS technology into farm machinery, such as combine harvesters, has been a significant component of precision farming. In this paper, we present a streaming-based methodology for detecting anomalies in combine harvester GPS recordings. The key hypothesis is that, “similar points in a feature space have similar anomaly scores”. We examine a data-driven strategy with the goal of applying unsupervised detection algorithms to find anomalies on the fly. Based on the results of the experiments, we can conclude that LSCP beats all other strategies when the number of base detectors is varied. The AUCPR performance of LSCP with two base detectors (HBOS and MCD) is 8.02 percent better than the second best technique MCD.

Keywords: Anomaly detection · Data streams · Precision farming · unsupervised learning.

1 Introduction

Smart Agriculture refers to the application of digital methods to innovate, control, and optimize agricultural production systems. Human intervention in agriculture is boosted by digital transformation, which aids in reducing effort, implementing particular measures, calibrating the use of chemical products on soil and crops, as well as ensuring and boosting yield. It also aids in the management

of all procedures that permit or support agricultural output, such as economic and administrative ones. Smart Agriculture’s goal is to provide solutions that can be used by all farmers, independent of farm size, location, or industry, while leveraging scale effects and keeping costs low. The benefits envisioned from the introduction and integration of technology processes in agriculture are now attributed to increased production and quality efficiency, cost reduction, input optimization, and environmental impact minimization.

The concept behind anomaly detection research is that exploitative behaviour differs quantitatively from normal system behavior. Anomalies are “patterns in data that do not conform to a well-defined notion of normal behaviour” [5]. These patterns reflect fresh, unexpected, or unknown data. Anomaly detection may be performed as a binary classification problem in the presence of labeled data using supervised learning approaches, with data labels being either normal or abnormal. This is seldom the case due to the lack of labeled data and the rarity of abnormal events. Unsupervised learning techniques are used in most anomaly detection systems since huge volumes of unlabeled data are available [5]. Intrusion detection, fraud detection, fault detection, system health monitoring, event detection in sensor networks, identifying ecological changes, and defect identification in images using machine vision are some of the examples of anomaly detection. Anomalies can be classified as erroneous data due to device failure/system failures or as odd data indicating rare/exceptional occurrences that occurred despite being abnormal [19]. Surprising data is unique to each instance, making it challenging to categorize these abnormalities using standard approaches.

This study examines anomaly detection in agricultural fields by analysing GPS logs generated by a combine harvester during wheat harvesting season. We utilize techniques based on hypothesis testing on the occurrence of an anomalous movement pattern during in-field harvesting. The primary assumption of our analysis is that, “Normal instances are far more frequent than anomalies”. The key hypothesis is that, “similar points in a feature space have similar anomaly scores”. When applied to grain harvesting : ”If a combine harvester changes its speed rapidly at a particular point within the field, then it implies an anomaly has occurred“. Our focus is on nonrecurring harvesting disruptions whose occurrence is usually unexpected and random.

We propose the following contributions:

1. We present a detailed state of the art on anomaly detection techniques with a focus on precision agriculture;
2. We define and investigate the issue of completely unsupervised anomaly ensemble construction.
3. We propose an ensemble based methodology for detection of anomalies from a data stream of combine harvester GPS logs;
4. We evaluate the methodology using a real data-set of GPS logs generated by a combine harvester during wheat harvesting and compare its performance with state of the art techniques under the streaming context.

The rest of this paper is structured as follows: Section 2 presents the state of the art investigation on Specialized Micro Transportation Systems, anomaly de-

tection techniques and anomaly detection in agricultural data. Section 3 presents the problem statement. Section 4 presents the methodology. Section 5 presents the experimental evaluation of the proposed technique, and Section 6 presents the conclusion and future work.

2 Related works

This section introduces works on Specialized Micro Transportation Systems, anomaly detection techniques, and application of anomaly detection to agricultural data.

2.1 Specialized Micro Transportation Systems

Intelligent Transportation Systems (ITS) offer substantial advances in transportation system safety, mobility, productivity and environmental conservation [7]. Specialized Micro Transportation Systems (SMTS) are a subset of ITS which comprise of a limited number of vehicles cooperating on a specialized job [23]. SMTS vehicles frequently follow predictable and recognized patterns of activity, which vary depending on the operation at hand. Due to the highly developed and mechanized state of agriculture, it is quite usual for farmers to use numerous vehicles in various essential agricultural operations in order to increase efficiency. Harvesting is perhaps one of the more engaging of these tasks, since it often involves three types of vehicles: combine harvesters, grain carts, and semi trucks. During grain harvesting a single combine harvester is sufficient for the majority of harvesting activities. However, due to its limited storage capacity and transport speed, a semi truck is generally needed to convey harvested grain from fields to a grain elevator for storage or sale. Since trucks are not adapted to the uneven field surfaces, grain carts are engaged to deliver the grain from the harvesters to the trucks at the edges of the fields. With good coordination between these three types of vehicles, a reasonable level of efficiency may be achieved in completing the work. ITS may be used to enhance SMTS management, monitoring, and efficiency.

2.2 Anomaly detection techniques

A lot of work has gone into developing a wide range of anomaly detection algorithms, which can be broadly classified as: Classification based, Nearest Neighbour based, Clustering based, Statistical methods, Information Theoretic based, Spectral and Graph based methods [5], [18]. These categories are summarised in figure 1 together with the workflow followed in anomaly detection.

In the discovery of outliers in high-dimensional data, the isolation-based technique employs the Isolation Forest (IForest) anomaly detection algorithm [11]. IForest is an unsupervised method that does not require labelled data and does not presume data distribution. It is also non-parametric and performs well on unbiased data with minimal noise points [2]. IForest is made up of a forest of

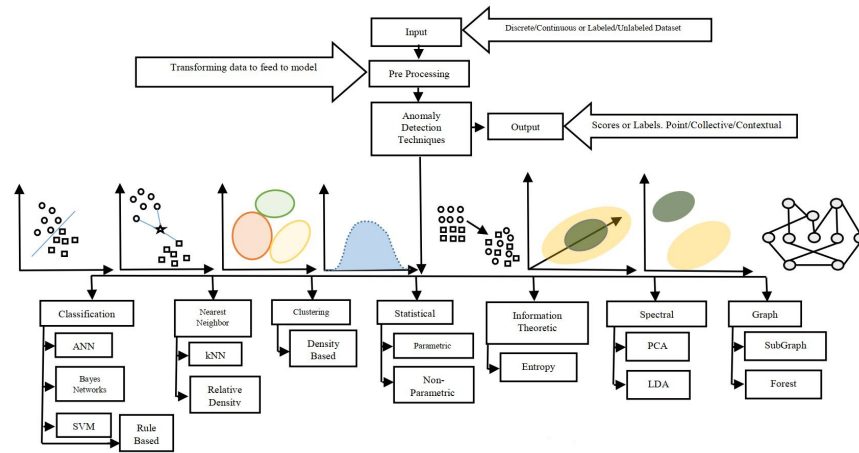


Fig. 1. Anomaly detection workflow and anomaly detection techniques [18]

randomly generated isolation trees (*itrees*). Because anomalies are uncommon and distinct from regular occurrences, a tree can be built to separate each one. At each tree, a determination is made as to whether or not an observation is normal. Anomalies are isolated closer to the tree’s root due to their sensitivity to isolation, whereas normal points are isolated further down.

Locally Selective Combination in Parallel Outlier Ensembles (LSCP) [24] is an unsupervised anomaly detection algorithm that uses the consensus of a test instance’s nearest neighbors in randomly selected feature subspaces to create a local area surrounding it. The Average of Maximum method is used to achieve LSCP, which involves fitting a homogeneous/heterogeneous list of base detectors to the training data, then picking the maximum outlier score to produce a pseudo ground truth for each instance. LSCP investigates both global and local data connections by training base detectors on the whole dataset and prioritizing data locality when combining detectors.

Histogram-based Outlier Score (HBOS) [9] is based on the assumption that features are independent, and hence computes outlier scores by creating histograms for each feature. Modeling the precise features of produced histograms and identifying deviations are used to identify anomalies. HBOS does not require data labeling and does not require any training or learning phase. With scoring-based detection, it also provides a quick computation time. A collection of k one-dimensional histograms is used by the Lightweight on-line detector of anomalies (LODA) [14] to identify anomalies. The probability density of input data projected onto a single projection vector is approximated by each histogram. Because LODA’s output is proportional to the sample’s negative log-likelihood, the greater the anomaly value, the less likely the sample is. This approach may be used to rank features according to their contribution to the sample’s anomalousness since each histogram with sparse projections in LODA provides an

anomaly score on a randomly generated subspace. LODA works effectively with data streams and isn't influenced by missing values.

It is common practice to use diagnostics in conjunction with a traditional fitting approach to try to find outliers. Outliers can, however, impact these traditional approaches, causing the fitted model to fail to discover deviating observations (the masking effect) [16]. By fitting the bulk of the data to the fitted model and then identifying data points that vary from the fitted model, robust statistics is a handy tool for detecting these deviating observations. A highly robust estimator for multivariate anomaly detection is the Minimum Covariance Determinant (MCD) [15]. MCD depends on two multivariate statistics features: the determinant of a matrix and Mahalanobis distances (MD). It looks for observations in the data set with the smallest feasible determinant in the classical covariance matrix.

2.3 Anomaly detection in agricultural data

DeepAnomaly is proposed in [6] where a combination of background subtraction and Deep Learning is used for detection of obstacles and anomalies in an agricultural field. It takes advantage of the fact that agriculture field's visual features are homogeneous, and obstacles occur rarely. The main idea is to detect distant and heavily occluded objects and unknown object types. The detected obstacles included people, barrels, wells and a distant house. DeepAnomaly detects humans more accurately and in real time at larger distances as compared to state-of-the-art techniques. Anomaly detection of time series data by applying two machine learning models, Autoregressive Integrated Moving Average model (ARIMA) and Long Short-Term memory (LSTM) is presented in [1]. A temporal anomaly detection approach is developed that takes into account the temporal connections between digital farm sensor data. LSTM and ARIMA models are tested on real data collected from deployed agricultural sensors with the aim of identifying anomalous data readings. It was discovered that using LSTM results in improved anomaly detection prediction but requires more training time.

The automated identification of crop parcels with abnormal vegetation development is a current problem in precision agriculture. Detecting crop patches with significantly different phenological behaviors from the rest of the crop might aid users like farmers and agricultural cooperatives in improving agricultural practices, disease identification, and fertilizer management. In [13] Isolation forest unsupervised outlier detection algorithm is applied to synthetic aperture radar (SAR) and multispectral images acquired using Sentinel-1 (S1) and Sentinel-2 (S2) satellites to detect the most abnormal wheat and rapeseed crop parcels within a growing season. The four primary types of anomalies investigated were heterogeneity issues, growth anomalies, database errors, and non-agronomic outliers that were deemed irrelevant for crop monitoring. The results of the experiments showed that the use of both S1 and S2 features in anomaly detection resulted in more late growth anomalies and heterogeneous parcels being detected.

Detection of anomalous activity movement of combine harvesters was done in [20] using a Kalman filter and DBSCAN algorithm. The latitude and longitude values of GPS logs are converted to easting and northing coordinates and recursively applied to a Kalman filter that is based on a constant velocity (CV) dynamical model. Kalman filter residual is computed which gives the level of deviation of the filter estimates from the actual measurements. The Kalman filter residual is a measure for the smoothness of the combine harvester motion such that, sudden change in motion results in a higher residual value. The engine load, vehicle speed, and computed Kalman filter residual are then applied to DBSCAN algorithm which generates clusters indicative of the activity being carried out by the combine harvester. Based on the results, it was possible to identify clusters for uniform motion in field, uniform motion on road, stationary points and nonuniform motions (in-field turns, on-road turns, decelerating, accelerating).

3 Problem Statement

In typical combine harvester operation (whether harvesting or travelling between fields), the combine harvester travels in a straight path at a near constant pace. The behaviour of combine harvesters depend on whether they are harvesting in a field or traveling between fields. Changes in combine harvester speed, on the other hand, usually signal in-field or on-road actions such as turning and accelerating. During harvesting a combine harvester should operate at a maximum of 4mph for yield efficiency reasons and a maximum of 20mph on the road or when moving between fields [23]. Moving at high speed can result in grain being missed thus lowering the yield. The focus of this research is the identification of abnormal harvesting behaviour by focusing on local regions (i.e. local contextual anomalies). Detecting anomalies using GPS data logs is a multivariate task where the data is modelled as a data stream and applied to anomaly detection techniques using windowing concept. The primary assumption of the analysis is that, “*Normal instances are far more frequent than anomalies*”. Therefore, the key hypothesis is: *similar points in a feature space have similar anomaly scores*. Harvesting perspective: “*If vehicles change their speed rapidly at a particular point, then it implies an anomaly has occurred*”. A deviation based anomaly detection approach is applied with a focus on in-field harvesting data. To perform multivariate anomaly detection we will consider the following data attributes: latitude, longitude, altitude, speed, bearing and accuracy.

4 Methodology

A sliding window approach is employed in streaming anomaly detection, in which data samples inside a window are sorted by an outlier score, with highly ranked data samples being labeled anomalies. A pipeline framework was adopted where each incoming new instance x_t is passed through a pre-processor (unit norm

scaler) which transforms x_t into a scaled feature vector without changing its dimensions. The scaled feature vector is then processed by the streaming anomaly detection model which predicts the label y_t for the instance. This predicted label is then passed to the running average post-processor which converts the score to the average of all previous scores in the current window. Figure 2 depicts the proposed anomaly detection framework.

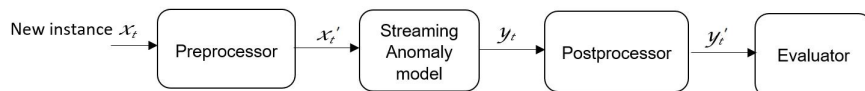


Fig. 2. Anomaly Detection framework

4.1 Data pre-processing and transformation

In this study we used a GPS data-set [21] collected using Nexus 7 tablets on a farm in Colorado U.S.A during 2017 wheat harvesting season. A GPS recorder app was kept running for each involved vehicle (combine harvesters, grain carts and trucks) during harvesting. In a typical harvesting scenario, there are more than one harvester working simultaneously in the field. This results in overlap of the trajectories of the vehicles. A single trajectory is considered as the consolidation of all GPS logs belonging to a single vehicle collected within a day. For the purpose of this study, we use the movement trajectory of a single combine harvester recorded in one day.

Definition: *Trajectory:* A raw trajectory consists of a sequence of n points $T = [p_1, p_2, \dots, p_n]$, in which $p_i = x, y, z, t, A$, where x, y, z represent the position (latitude, longitude and altitude) of the moving vehicle in space, t is the timestamp and A represents other attributes associated with the point (i.e. speed, bearing and accuracy).

Trajectory mining was done using Quantum GIS (QGIS) an open-source cross-platform desktop geographic information system application that supports viewing, editing and visualization of geospatial data. The key interest were the data points generated in the field during harvesting, therefore visualization and map matching was done using QGIS to ensure the GPS points were mapped to a field. The second step was to extract only those points within a specific field using a bounding box. The extracted data exhibited normal harvesting behaviour with all data points below 4mph which is the maximum harvesting speed. To create an evaluation data-set, anomalies were introduced in the original data-set by varying the vehicle speed at specific points along the trajectory such that for specific sections of the trajectory a sequential number of points had their speed increased by a random number between a given range of values above 4mph.

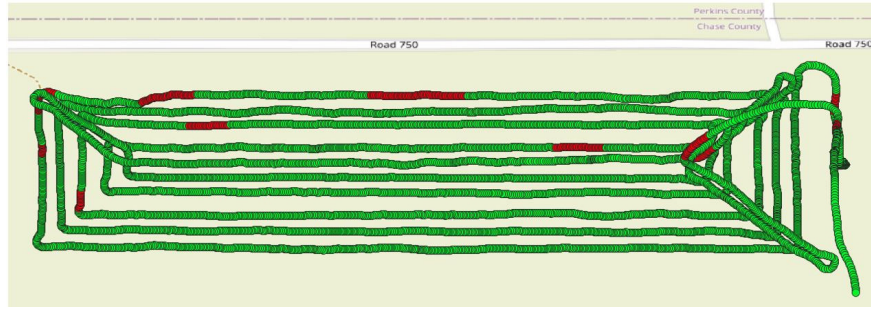


Fig. 3. Area of interest; in-field trajectory of a combine harvester showing normal points in green and anomalies in red

Figure 3 presents the in-field trajectory of a combine harvester after introduction of anomalies with normal points in green, anomalies in red.

Multivariate anomaly detection is performed by considering the latitude, longitude, altitude, speed, bearing and accuracy. The data is modelled as a data stream and applied to ensemble based technique for anomaly detection using windowing concept.

4.2 Performance indicators

In order to evaluate the performance of the different approaches, we use the Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC) [10] and the Area Under the Curve of Precision-Recall (AUCPR) [3]. Both indicators are based on the concepts of:

- True Positive (TP): True Positives are the correctly identified anomalies.
- False Positive (FP): False positive are incorrectly identified normal data.
- True Negative (TN): True negative are correctly identified normal data.
- False Negative (FN): False negative are incorrectly rejected anomalies.

The True Positive Rate (TPR), or Recall, is:

$$TPR = \frac{TP}{TP + FN}.$$

The False Positive Rate (FPR) is:

$$FPR = \frac{FP}{FP + TN}.$$

AUC-ROC receiver operating characteristics are TPR and FPR. The higher the AUC-ROC is, the better the detection is. AUC-ROC is the most popular evaluation measure for unsupervised outlier detection methods [4]. The AUCPR baseline is equivalent to the fraction of positives [17]:

Table 1. Experimental parameters

Parameters	Values
Window sizes	100, 200, 300, 400, 500, 600, 700, 800, 900, 1000
Sliding window size	50
Initial window (training set)	500

$$AUCPR - baseline = \frac{TP}{TP + FP + FN + TN}$$

AUCPR uses Precision and Recall. Precision is the fraction of retrieved instances that are relevant [8].

$$Precision = \frac{TP}{TP + FP}$$

5 Experimental Evaluation and Results

In this section, we present results obtained by experimenting anomaly detection techniques on GPS logs. The algorithms were implemented in Python programming language using PySAD framework [22] which allows us to integrate batch processing algorithms from PyOD [25] and to apply them to streaming data using sliding windows. HBOS, LODA, LSCP, MCD and IForest algorithms were implemented by adapting them to the streaming context using a reference window model. Since LSCP is an ensemble framework, we varied the number of base detectors in its implementation resulting in three models. LSCP_3 implements HBOS, MCD and IForest as the base detectors. LSCP_2 implements MCD and IForest as the base detectors. LSCP_1 implements MCD and HBOS as the base detectors. The reason for variation of base detectors was to attain unbiased overall detection accuracy with little variance by incorporating the capabilities of various base detectors while carefully combining their outputs to form a robust detector.

In the experiments we sort to find answers to the following questions:

- Which model best predicts anomalous cases given a streaming data set?
- Does the size of the window have an impact on how well models identify anomalies?
- What role does data association play in anomaly detection?

Multiple experiments with variation in the window size were carried out to establish the impact of the window size on the performance of the algorithms. Table 1 summarises the parameters used in the experiments.

AUC-ROC and AUCPR were utilized to investigate the algorithms’ performance since they are suited for imbalanced data sets. The applied models output probabilities for each instance which are transformed into predicted labels using a decision threshold. These predicted labels are then used to compute the

Table 2. AUC-ROC performance results for window size variation

Window size	LODA	HBOS	IForest	MCD	LSCP_1	LSCP_2	LSCP_3
100	0.5391	0.8851	0.9590	0.9913	0.9950	0.9960	0.9890
200	0.6529	0.8874	0.9739	0.9948	0.9926	0.9942	0.9891
300	0.7138	0.8950	0.9763	0.9948	0.9940	0.9924	0.9884
400	0.7719	0.9102	0.9788	0.9926	0.9905	0.9938	0.9907
500	0.7894	0.9208	0.9800	0.9953	0.9893	0.9932	0.9904
600	0.8067	0.9257	0.9783	0.9955	0.9911	0.9921	0.9889
700	0.8102	0.9328	0.9786	0.9923	0.9910	0.9927	0.9896
800	0.8168	0.9412	0.9800	0.9917	0.9918	0.9910	0.9899
900	0.8070	0.9502	0.9803	0.9926	0.9912	0.9911	0.9888
1000	0.8104	0.9560	0.9795	0.9900	0.9920	0.9902	0.9888
Average	0.7518	0.9204	0.9765	0.9931	0.9918	0.9927	0.9894

AUC-ROC and AUCPR values. Table 2 and 3 summarises the AUC-ROC and AUCPR results obtained from the experiments.

The experimental results reveal that the performance of LSCP with variation in the number of base detectors is very competitive. The performance is particularly notable in LSCP_1 which achieves a AUCPR value of 0.7661 which is 8.02 percent better than the second best technique MCD. Figure 4 and 5 shows that the performance of all algorithms in terms of AUC-ROC and AUCPR. The performance of LODA and HBOS in terms of AUC-ROC and AUCPR increases progressively with increase in window size. IForest, MCD and variants of LSCP show slight variation in AUC-ROC as the window sizes are varied.

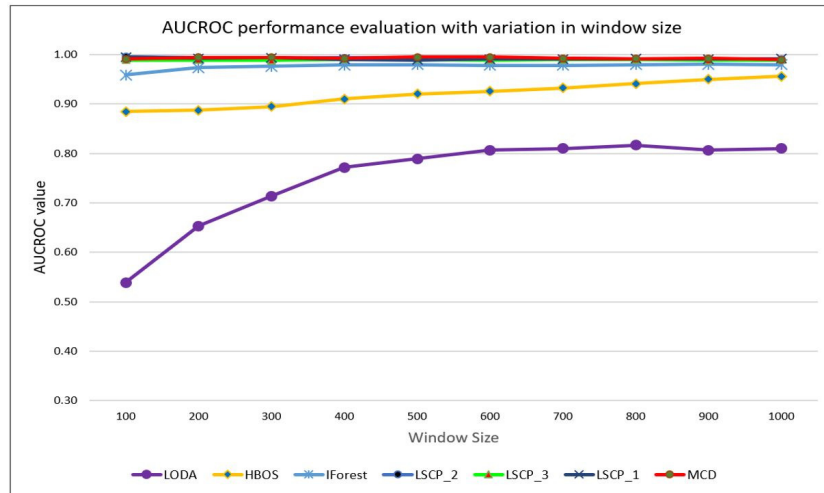
**Fig. 4.** Comparison of models' AUC-ROC performance

Table 3. AUCPR performance results for window size variation

Window size	LODA	HBOS	IForest	MCD	LSCP_1	LSCP_2	LSCP_3
100	0.0464	0.2162	0.4190	0.7072	0.8156	0.8567	0.7426
200	0.0563	0.2778	0.4853	0.7501	0.7730	0.7914	0.6989
300	0.1034	0.2936	0.4882	0.7398	0.8081	0.7538	0.7257
400	0.1283	0.3467	0.4937	0.6450	0.7441	0.7283	0.6844
500	0.2234	0.3626	0.5123	0.7553	0.7521	0.7408	0.7282
600	0.2026	0.3707	0.5079	0.7591	0.7500	0.7178	0.7085
700	0.2017	0.3956	0.5009	0.6404	0.7477	0.7328	0.7055
800	0.1994	0.4128	0.4974	0.6239	0.7637	0.7072	0.6913
900	0.1899	0.4364	0.5055	0.6516	0.7470	0.7229	0.6995
1000	0.1842	0.4662	0.4884	0.5867	0.7602	0.7099	0.7083
Average	0.1536	0.3578	0.4899	0.6859	0.7661	0.7462	0.7093

The baseline AUC-ROC is usually set at a value of 0.5 which suggests no discrimination, 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and more than 0.9 is considered outstanding [12]. Generally, apart from LODA which achieves an average AUC-ROC value of 0.7518 (considered acceptable), all the other approaches achieved AUC-ROC values above 0.9 which is outstanding performance. The baseline AUCPR for the data-set was 0.0333 (based on a sample size of 7416 with 247 true positive anomalies). Based on the results obtained for AUCPR, it is safe to conclude that all the techniques are able to distinguish anomalies from normal instances.

6 Conclusion

Improving agricultural production efficiency and crop yields is impossible without the use of contemporary digital technologies and smart machinery. The introduction of high-accuracy GPS technology into farm machinery, such as combine harvesters, has been a significant component of precision farming. To address the issue of anomaly detection, several techniques have been developed. Anomaly detection has a wide range of applications that necessitate a dependable and precise solution. In this study we have considered the analysis of GPS logs collected from a combine harvester in order to detect anomalous behaviour during wheat harvesting. The implementation of the windowing concept facilitates the detection of anomalies on the fly. It is worth noting that, through the use of machine learning it is possible to detect deviating behaviour of combine harvesters. Also the application of ensemble techniques improves the precision and recall detection performance. Based on the obtained results, LSCP_1 with two base detectors (HBOS and MCD) is able to handle GPS logs and detect anomalies with efficiency and therefore can be used to detect anomalous behaviour during farm operations.

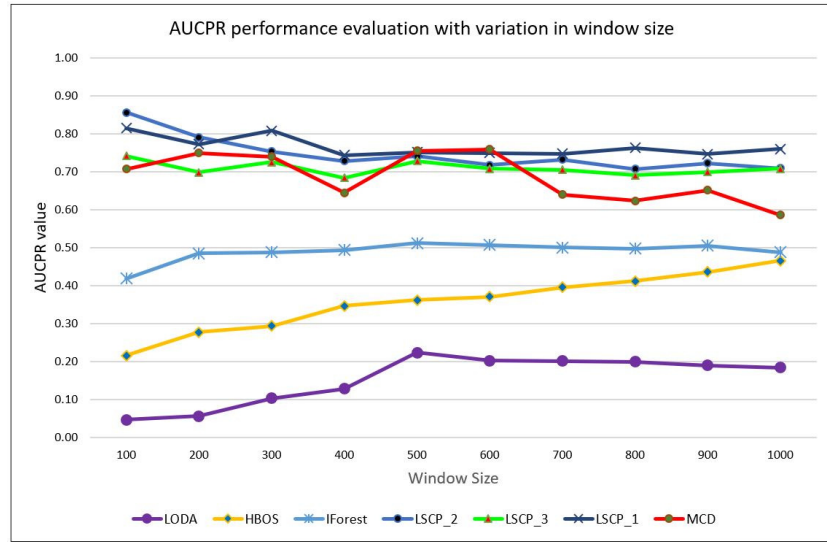


Fig. 5. Comparison of models' AUCPR performance

Our future work will focus on adjustments of the ensemble learning decision rules for efficiency improvement. It will also be interesting to perform anomaly detection in field production using yield values.

References

1. Abdallah, M., Lee, W.J., Raghunathan, N., Mousoulis, C., Sutherland, J.W., Bagchi, S.: Anomaly detection through transfer learning in agriculture and manufacturing iot systems. arXiv preprint arXiv:2102.05814 (2021)
2. Aggarwal, C.C.: Outlier analysis. In: Data mining. pp. 237–263. Springer (2015)
3. Boyd, K., Eng, K.H., Page, C.D.: Area under the precision-recall curve: point estimates and confidence intervals. In: Joint European conference on machine learning and knowledge discovery in databases. pp. 451–466. Springer (2013)
4. Campos, G.O., Zimek, A., Sander, J., Campello, R.J., Micenková, B., Schubert, E., Assent, I., Houle, M.E.: On the evaluation of unsupervised outlier detection: measures, datasets, and an empirical study. *Data mining and knowledge discovery* **30**(4), 891–927 (2016)
5. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey. *ACM computing surveys (CSUR)* **41**(3), 1–58 (2009)
6. Christiansen, P., Nielsen, L.N., Steen, K.A., Jørgensen, R.N., Karstoft, H.: Deep-anomaly: Combining background subtraction and deep learning for detecting obstacles and anomalies in an agricultural field. *Sensors* **16**(11), 1904 (2016)
7. Faezipour, M., Nourani, M., Saeed, A., Addepalli, S.: Progress and challenges in intelligent vehicle area networks. *Communications of the ACM* **55**(2), 90–100 (2012)
8. Fawcett, T.: An introduction to roc analysis. *Pattern recognition letters* **27**(8), 861–874 (2006)

9. Goldstein, M., Dengel, A.: Histogram-based outlier score (hbos): A fast unsupervised anomaly detection algorithm. *KI-2012: Poster and Demo Track* pp. 59–63 (2012)
10. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (roc) curve. *Radiology* **143**(1), 29–36 (1982)
11. Liu, F.T., Ting, K.M., Zhou, Z.H.: Isolation forest. In: 2008 Eighth IEEE International Conference on Data Mining. pp. 413–422. IEEE (2008)
12. Mandrekar, J.N.: Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology* **5**(9), 1315–1316 (2010)
13. Mouret, F., Albughdadi, M., Duthoit, S., Kouamé, D., Rieu, G., Tourneret, J.Y.: Outlier detection at the parcel-level in wheat and rapeseed crops using multispectral and sar time series. *Remote Sensing* **13**(5), 956 (2021)
14. Pevný, T.: Loda: Lightweight on-line detector of anomalies. *Machine Learning* **102**(2), 275–304 (2016)
15. Rousseeuw, P.J., Driessen, K.V.: A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**(3), 212–223 (1999)
16. Rousseeuw, P.J., Hubert, M.: Anomaly detection by robust statistics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* **8**(2), e1236 (2018)
17. Saito, T., Rehmsmeier, M.: The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one* **10**(3), e0118432 (2015)
18. Toshniwal, A., Mahesh, K., Jayashree, R.: Overview of anomaly detection techniques in machine learning. In: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC). pp. 808–815. IEEE (2020)
19. Wang, X., Fagette, A., Sartelet, P., Sun, L.: A probabilistic tensor factorization approach to detect anomalies in spatiotemporal traffic activities. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 1658–1663. IEEE (2019)
20. Wang, Y., Balmos, A., Krogmeier, J., Buckmaster, D.: Data-driven agricultural machinery activity anomaly detection and classification. In: Proceedings of the 14th International Conference on Precision Agriculture (2018)
21. Yaguang, Z., James, K.: Combine kart truck gps data archive (May 2020). <https://doi.org/10.4231/GMH9-8X88>, <https://purr.purdue.edu/publications/3083/2>
22. Yilmaz, S.F., Kozat, S.S.: Pysad: A streaming anomaly detection framework in python. *arXiv preprint arXiv:2009.02572* (2020)
23. Zhang, Y., Balmos, A., Krogmeier, J.V., Buckmaster, D.: Working zone identification for specialized micro transportation systems using gps tracks. In: 2015 IEEE 18th International Conference on Intelligent Transportation Systems. pp. 1779–1784. IEEE (2015)
24. Zhao, Y., Nasrullah, Z., Hryniewicki, M.K., Li, Z.: Lscp: Locally selective combination in parallel outlier ensembles. In: Proceedings of the 2019 SIAM International Conference on Data Mining. pp. 585–593. SIAM (2019)
25. Zhao, Y., Nasrullah, Z., Li, Z.: Pyod: A python toolbox for scalable outlier detection. *Journal of Machine Learning Research* **20**(96), 1–7 (2019), <http://jmlr.org/papers/v20/19-011.html>