

Hybrid machine learning approach for anomaly detection

Lai Kai Lok, Vazeerudeen Abdul Hameed, Muhammad Ehsan Rana

School of Computing, Asia Pacific University of Technology and Innovation (APU), Bukit Jalil, Malaysia

Article Info

Article history:

Received Jun 15, 2021

Revised Mar 31, 2022

Accepted Jun 9, 2022

Keywords:

Linear regression

Machine learning

Supervised learning

Support vector machine

Unsupervised learning

ABSTRACT

This research aims to improve anomaly detection performance by developing two variants of hybrid models combining supervised and unsupervised machine learning techniques. Supervised models cannot detect new or unseen types of anomaly. Hence in variant 1, a supervised model that detects normal samples is followed by an unsupervised learning model to screen anomaly. The unsupervised model is weak in differentiating between noise and fraud. Hence in variant 2, the hybrid model incorporates an unsupervised model that detects anomaly is followed by a supervised model to validate an anomaly. Three different datasets are used for model evaluation. The experiment is begun with 5 supervised models and 3 unsupervised models. After performance evaluation, 2 supervised models with the highest F1-Score and one unsupervised model with the best recall value are selected for hybrid model development. The variant 1 hybrid model recorded the best recall value across all the experiments, indicating that it is the best at detecting actual fraud and less likely to miss it compared to other models. The variant 2 hybrid model can improve the precision score significantly compared to the original unsupervised model, indicating that it is better in separating noise from fraud.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Vazeerudeen Abdul Hameed

School of Computing, Asia Pacific University of Technology and Innovation (APU)

Bukit Jalil, Kuala Lumpur 57000, Malaysia

Email: vazeerudeen@gmail.com

1. INTRODUCTION

Anomaly detection is the process of extracting outliers in a dataset whose complexity is amplified by the complex nature of the systems that process the data. Data is often unstructured which is a weakness that causes systems to be vulnerable to intruders. Anomaly detection systems could be manually created by experts. Various checkpoints and thresholds could be set to monitor the possible outliers. However, this would require extensive human interference and monitoring to maintain the thresholds at the right levels to minimize the possibility of false positives. A much viable alternative could be the use of machine learning approaches to monitor and detect anomaly.

2. LITERATURE REVIEW

Anomaly detection plays a significant role in different domains. In manufacturing, unscheduled shutdowns and accidents can be avoided while the efficiency of production can be improved with effective anomaly detection [1]. Anomaly detection in the finance domain can reduce loss due to credit card fraud and improve customers' confidence [2]. To ensure the privacy and security of internet users, effective anomaly detection in the form of internet intrusion detection is needed. This can also avoid crucial systems like military or healthcare infrastructure from cyber-attack [3].

There are two major types of anomalies in the manufacturing domain. The first type is the abnormal activity during the production process, mainly on the production machine's condition or environment. The second type is the defect or the quality of the end product. Long short-term memory based machine learning methods is used by both Verner and Mukherjee [4] and Hsieh *et al.* [5] to detect the anomalies in the sensor data from the production line. In the study of Quatrini *et al.* [1] and Qosim and Zulkarnain [6], random forest (RF), which is a type of ensemble learning performed the best for detecting anomalies in the production process. For checking the quality of the solder paste, Zheng *et al.* [7] proposed a hybrid method consisting of isolation forest, k-means clustering and transfer-learning while [8] is using a generative adversarial network (GAN). Both methods are performing better than conventional machine learning techniques.

Credit card fraud is a major problem in the finance industry. To improve the performance of credit card fraud detection, a resampling technique is applied to solve the class imbalance problem [9]. However, it is concluded that it is not effective enough. In other studies by Baabdullah *et al.* [10] and Rtayli and Enneya [11], it is shown that the resampling technique can improve the model's performance. The three studies are using different datasets. Ensemble learning methods are reported to perform the best in credit card fraud detection [12], [13]. Unsupervised K-means clustering method is compared with isolation forest and displayed a better reading in terms of area under precision recall curve (AUC-PR) [14]. This indicated that the unsupervised method is better in detecting anomalies especially the unseen type during training.

Anomaly detection in the internet security domain is mainly aimed at detecting the cyber-attack type of abnormal activity. In the study of Hasan *et al.* [15], RF is once again showing the best performance in detecting cyber-attack among other supervised machine learning techniques. Unsupervised machine learning methods are shown to perform better for detecting the new or unfamiliar type of cyber-attack [16], [17]. There is also a study where intelligent algorithms are used to improve the performance of machine learning models [18]. Both supervised and unsupervised machine learning algorithms have proven to be viable in solving several real-time problems [19], [20].

There are contradictory conclusions made on the effectiveness of the resampling technique for solving the class imbalance issue. Supervised machine learning techniques failed to detect unseen or new types of anomalies while unsupervised machine learning techniques tend to classify noises as anomalies [21], [22]. Thus, two variations of hybrid models which combine both the supervised and unsupervised techniques are proposed so that it can exceed the performance of conventional machine learning techniques in anomaly detection.

3. RESEARCH METHOD

The flow of the research is divided into three main stages as shown in Figure 1. Stage 1 is mainly on data preparation. In this stage, three different datasets are collected followed by data pre-processing and data splitting. In stage 2, conventional supervised and unsupervised machine learning models are used for detecting fraud in all the three different datasets. In the final stage 3, hybrid models are developed and evaluated together with the resampling technique.

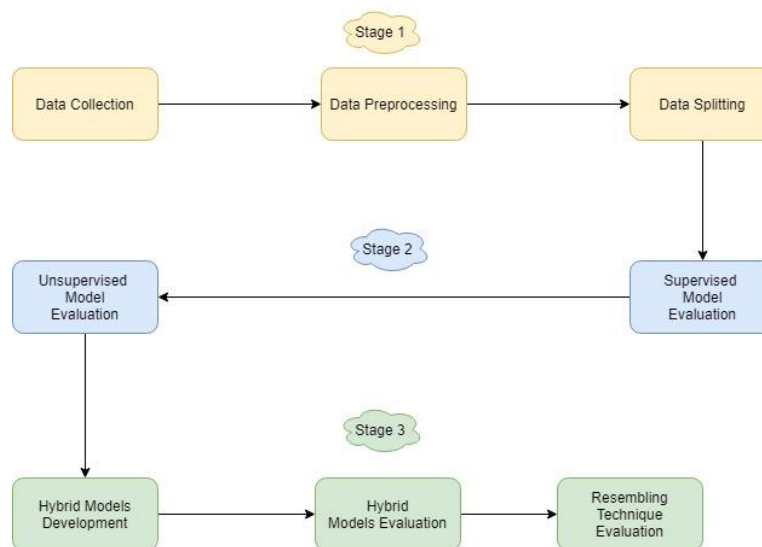


Figure 1. Three main stages of the research

3.1. Dataset description

Two different credit card dataset and one synthetic financial transaction dataset are used in this research. The details of these dataset are described in this section. The first dataset is the ULB credit card transactions dataset, downloaded from the Kaggle website [23]. The transactions in this dataset are made by European cardholders in a two days period of September 2013. The targeted variable of the dataset is to classify whether a particular credit card transaction is a normal or fraudulent transaction. This dataset has a total of 31 features and 284807 samples. Out of the 284807 samples, only 0.172% or 492 samples are fraudulent transactions. Same as most of the anomaly problems, this dataset is highly imbalanced. The summary of the dataset features is presented in Table 1.

Table 1. Summary of the features in dataset 01

Feature	Definition	Type
Time	The different in time period between the first sample and the current sample in seconds	Numeric
V1-V28	Data transformed by using principle component analysis or (PCA) to protect users' privacy and confidentiality	Numeric
Amount	The transaction amount of the sample	Numeric
Class	The target variable or the classification of the transaction, normal (0) or fraud (1)	Categorical

The PaySim dataset is the second dataset used in this research, downloaded from the Kaggle website [24]. The mobile money transactions are synthetically generated by the PaySim simulator using the real world one-month financial logs data from a mobile money service conducted in an African country. The original financial logs data are obtained from a mobile financial service multinational company that is currently running its business in more than 14 countries. The targeted outcome of the dataset is to identify whether a specific mobile money transaction is a fraud or not. This synthetic dataset has a total number of 6362620 instances and 11 features. There are only 8213 instances or 0.129% of the total instances are fraudulent transactions, which is again highly imbalanced. The summary of the dataset features is presented in Table 2.

Table 2. Summary of the features in dataset 02

Feature	Definition	Type
step	A measure of time, where 1 step equal to 1 hour. The whole dataset has 744 steps equivalence to 30 days of simulation	Numeric
type	The type of mobile money transaction. There are five categories in this dataset, which are cash-in, cash-out, debit, payment and transfer	Categorical
amount	The amount of money involved in the transaction, in local currency	Numeric
nameOrig	The ID of the client who made the transaction	Categorical
oldbalanceOrg	The amount of money left in the original account before the transaction	Numeric
newbalanceOrg	The amount of money left in the original account after the transaction	Numeric
nameDest	The ID of the recipient from the transaction	Categorical
oldbalanceDest	The amount of money left in the recipient's account before the transaction. No information on this if the recipient is merchants	Numeric
newbalanceDest	The amount of money left in the recipient's account after the transaction. No information on this if the recipient is merchants	Numeric
isFraud	The targeted outcome of the classification, whether it is a fraudulent transaction (1) or a normal transaction (0)	Categorical
isFlaggedFraud	This is to regulate the transactions which involve a massive amount of money. A transaction that transfer more than 200,000 is flagged as (1) while less than that is (0)	Categorical

The third dataset is also a credit card dataset, downloaded from the Index of dataset website [25]. This is also the dataset used by Makki *et al.* [9] and Baabdullah *et al.* [11] for fraud detection experiments. The transactions in the dataset are made by the credit card holder who lives in the United State. All the values of the data are already transformed into numerical values. The targeted variable is to identify whether the particular transaction is a fraudulent or legitimate transaction. There is a total of 10,000,000 instances with 9 features each. Out of the 10 million samples, only 596014 or 5.96% are fraudulent transactions. This is the least imbalanced dataset in terms of percentage among the three datasets used in this research. The summary of the 9 features is shown in Table 3.

3.2. Performance evaluation metrics

The precision is derived from true positive (TP) and false positive (FP) as shown in (1). In the context of fraud detection, the precision measures the proportion of correctly predicted fraud out from all the samples that is predicted as fraud by the model. Higher precision means when a model is predicting an instance as fraud, it is more likely that the prediction is correct. This provides a clearer picture on the model performance in fraud detection.

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

The recall is derived from TP and FN as shown in (2). In the context of fraud detection, the recall measures the proportion of correctly predicted fraud out of all the actual fraud in the dataset. Higher recall translates to better performance in detecting fraud. As the recall and precision do not use the true negative (TN) in the calculation, both are not affected by the highly imbalanced characteristic of anomaly detection and show a clearer picture of how well the model performs in detecting fraud.

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

The F1-Score is derived from the precision and recall as shown in (3). It calculates the harmonic mean of both the precision and recall. Compared to the normal mean where it considers each value equally, the harmonic means are heavily affected by low values. F1-Score will only show a high reading if both the precision and recall are high, which give an overall picture of how well the precision and recall values.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} = \frac{2 \times TP}{2 \times TP + FN + FP} \tag{3}$$

Table 3. Summary of the features in dataset 03

Feature	Definition	Type
custID	The customer ID of the credit card holder	Categorical
gender	The Gender of the customer, male or female	Categorical
state	The State of the United State where the customer resides	Categorical
cardholder	The number of credit card owned by the customer, maximum 2, minimum 1	Categorical
balance	The credit card balance in USD	Numeric
numTrans	The total number of transactions made by the customer using the credit card he or she owns	Numeric
numIntTrans	The total number of international transactions made by the customer using the credit card he or she owns	Numeric
creditLine	The credit limit of the customer	Numeric
fraudRisk	The targeted outcome of the classification, whether the transactions associated with a particular customer contain any fraudulent transaction (1) or only normal transaction (0)	Categorical

3.3. Hybrid models development

There are five supervised machine learning models and three unsupervised machine learning models used in this research to evaluate its performance in fraud detection. All the conventional machine learning methods used in this research are shown in Table 4. After evaluating the performance of all the models, two best performing supervised models and one best performing unsupervised model are selected for hybrid model development.

Table 4. List of machine learning models used in this research

Supervised machine learning	Unsupervised machine learning
<ul style="list-style-type: none"> • Decision Tree • Logistic Regression • Support Vector Machine • K-Nearest Neighbour • Random Forest 	<ul style="list-style-type: none"> • K means • One-Class SVM • Isolation Forest

There are two variants of hybrid models being developed in this research. The first variant is to improve the performance of supervised machine learning models in fraud detection, by improving the number of actual fraud being detected or the TP number. As supervised machine learning models are good at detecting known types of fraud while weak in detecting unseen or new types of fraudulent transactions, those samples that are predicted as normal by the supervised models are being sent for second stage screening by

using an unsupervised machine learning model. With this, all the new or unseen types of fraud can be detected as well.

The second variant is mainly focused on improving the performance of unsupervised machine learning models in fraud detection, by reducing the number of falsely identified fraud or the FP number. As unsupervised machine learning models are good at detecting new or unseen types of anomalies but weak at differentiating between noise and actual fraud, those instances that are identified as fraud are being sent for the second stage of filtering by using supervised machine learning model. With this, the noise and the actual fraud can be better separated.

After evaluating the performance of all the eight models in the previous Section, 2 out of five from the supervised machine learning models with the best F1-Score and one out of 3 unsupervised machine learning models with the highest actual fraud identification or best TP number are selected for the hybrid models development. In each variant of the hybrid model, either supervised model followed by unsupervised model or unsupervised model followed by supervised model, two hybrid models will be developed, making up a total of 4 hybrid models with 2 models for each variant.

4. RESULTS AND DISCUSSION

In dataset 01 and dataset 02, where all independent features are transformed into numerical value by using principle component analysis (PCA) or one-hot encoding, RF model showed the best balance between precision and recall, resulting in highest F1-Score. In dataset 03, where the categorical data is not one-hot encoded, represented by using a range of numbers instead, the RF model is not performing well. This is because RF treats these features as a range of numbers with different significance rather than as categorical variables. Table 5 shows the performance of the models for dataset 01 without resampling while the Table 6 demonstrates the performance with resampling. As highlighted, there is a notable difference in the precision and the F1-scores.

Table 5. Performance of each model for Dataset 01 without resample technique

D1 without Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised Machine Learning	DT	124	34	42	90938	0.74699	0.78481	0.76543
	LR	90	68	18	90962	0.83333	0.56962	0.67669
	SVM	126	32	27	90953	0.82353	0.79747	0.81029
	KNN	114	44	6	90974	0.95000	0.72152	0.82014
	RF	123	35	9	90971	0.93182	0.77848	0.84828
Unsupervised Machine Learning	Kmeans	134	24	2601	88379	0.04899	0.84810	0.09264
	OCSVM	118	40	8960	82020	0.01300	0.74684	0.02555
	Isolation Forest	128	30	3673	87307	0.03368	0.81013	0.06466
Hybrid Model	RF -Kmeans	138	20	2725	88255	0.04820	0.87342	0.09136
	KNN -Kmeans	134	24	2717	88263	0.04700	0.84810	0.08907
	Kmeans -RF	119	39	9	90971	0.92969	0.75316	0.83217
	Kmeans -KNN	114	44	5	90975	0.95798	0.72152	0.82310

Table 6. Performance of each model for Dataset 01 with resample technique

D1 with Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised Machine Learning	DT	113	45	60	90920	0.65318	0.71519	0.68278
	LR	128	30	316	90664	0.28829	0.81013	0.42525
	SVM	128	30	115	90865	0.52675	0.81013	0.63840
	KNN	116	42	38	90942	0.75325	0.73418	0.74359
	RF	118	40	13	90967	0.90076	0.74684	0.81661
Unsupervised Machine Learning	K means	77	81	2658	88322	0.02815	0.48734	0.05323
	OCSVM	112	46	9232	81748	0.01199	0.70886	0.02357
	Isolation Forest	124	34	3178	87802	0.03755	0.78481	0.07168
Hybrid Model	RF - IsoF	131	27	3178	87802	0.03959	0.82911	0.07557
	KNN - IsoF	130	28	3197	87783	0.03907	0.82278	0.07461
	IsoF - RF	111	47	13	90967	0.89516	0.70253	0.78723
	IsoF- KNN	110	48	19	90961	0.85271	0.6962	0.76655

Table 7 shows the performance of the models for dataset 02 without resampling while the Table 8 presents the performance of the models with resampling technique. The precision of the logistic regression

(LR) and support vector machine (SVM) models remained unaffected. However, other models such as the k-nearest neighbors (K-NN), RF-IsoF had noticeable differences in the performance.

Among the unsupervised machine learning models, K Means is able to detect the most number of actual frauds only in the experiment of dataset 01 without resampling technique. In all other cases, it is the worst as most actual frauds remained undetected. K Means is a clustering method and it uses the distance between the centroid of the cluster and the sample to decide whether a sample is a fraud or not. When the fraud samples are mixed with the normal samples without clear separation, K Means will not be able to perform. The IsoF model recorded the highest recall in all other cases among the unsupervised models. Compared to supervised models, unsupervised models have a relatively low value of precision as it is unable to differentiate between noise and actual fraud. As unsupervised models do not use the class labelled of the instances or fraud samples for model training, the resampling technique does not improve the performance of these models.

Table 7. Performance of each model for Dataset 02 without resample technique

D2 without Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised	DT	316	12	14	110474	0.95758	0.96341	0.96049
Machine Learning	LR	2	326	0	110488	1.00000	0.00610	0.01212
	SVM	12	316	0	110488	1.00000	0.03659	0.07059
	KNN	161	167	7	110481	0.95833	0.49085	0.64919
	RF	322	6	5	110483	0.98471	0.98171	0.98321
Unsupervised Machine Learning	K means	93	235	3232	107256	0.02797	0.28354	0.05092
	OCSVM	174	154	11142	99346	0.01538	0.53049	0.02989
Hybrid Model	Isolation Forest	226	102	11621	98867	0.01908	0.68902	0.03713
	RF - IsoF	323	5	11626	98862	0.02703	0.98476	0.05262
	DT - IsoF	320	8	11629	98859	0.02678	0.97561	0.05213
	IsoF - RF	225	103	0	110488	1.00000	0.68598	0.81374
	IsoF - DT	222	106	6	110482	0.97368	0.67683	0.79856

Table 8. Performance of each model for Dataset 02 with resample technique

D2 with Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised	DT	322	6	11	110477	0.96697	0.98171	0.97428
Machine Learning	LR	4	324	0	110488	1.00000	0.01220	0.02410
	SVM	20	308	0	110488	1.00000	0.06098	0.11494
	KNN	151	177	12	110476	0.92638	0.46037	0.61507
	RF	325	3	4	110484	0.98784	0.99085	0.98935
Unsupervised Machine Learning	K means	84	244	3241	107247	0.02526	0.25610	0.04599
	OCSVM	168	160	11024	99464	0.01501	0.51220	0.02917
Hybrid Model	Isolation Forest	230	98	10714	99774	0.02102	0.70122	0.04081
	RF - IsoF	326	2	10717	99771	0.02952	0.9939	0.05734
	DT - IsoF	324	4	10720	99768	0.02934	0.98780	0.05698
	IsoF - RF	229	99	1	110487	0.99565	0.69817	0.82079
	IsoF - DT	228	100	5	110483	0.97854	0.69512	0.81283

The application of the resampling technique only shows a trend in improving the recall value for the LR and SVM models. This is because only these two models are able to capitalize on the increase in fraud samples for forming a better decision boundary. When the performance of the supervised models is improved with the application of the resampling technique, those hybrid models that used the improved supervised model are showing better performance as well as shown in Table 9 and Table 10.

Across the three datasets, the variant 1 hybrid models, supervised followed by unsupervised machine learning techniques, displayed improved recall score compared to both the original supervised model and unsupervised model. In fact, across all the six experiments, hybrid model variant 1 is the model that showed the best recall value. This indicates that the variant 1 hybrid model is best in detecting fraud and less likely to miss actual fraud. The precision of the hybrid model is slightly better than the original unsupervised model but lower than that of the supervised model, the same trend can be seen on the F1-Score. The variant 1 hybrid model is definitely a better model compared to the unsupervised model and a better model in terms of detecting actual fraud but when it comes to precision score or the number of FP, supervised models are the better choice.

Across the three datasets, the variant 2 hybrid models, unsupervised followed by supervised machine learning techniques, showed a significant improvement in the precision value and F1-Score compared to the original unsupervised model. In some cases, it is displaying better precision compared to the

original supervised model. This indicated that the variant 2 hybrid model is able to resolve the issue of the weak ability of unsupervised models in differentiating noise from fraud. However, these improvements are associated with a decrease in recall value compared to the original unsupervised models.

For those applications where detecting actual fraud is crucial and missing the actual fraud can bring a significant bad effect, variant 1 hybrid model is a suitable candidate. When there is not much-labelled fraud data and an unsupervised machine learning model is more practical, variant 2 hybrid model can be used to improve the unsupervised model in differentiating noise from fraud.

Table 9. Performance of each model for Dataset 03 without resample Technique

D3 without Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised	DT	2343	2425	2749	72483	0.46013	0.49140	0.47525
Machine Learning	LR	2260	2508	777	74455	0.74416	0.47399	0.57912
	SVM	1727	3041	378	74854	0.82043	0.36221	0.50255
	KNN	2141	2627	1347	73885	0.61382	0.44904	0.51865
	RF	2202	2566	1065	74167	0.67401	0.46183	0.54810
Unsupervised	K means	718	4050	1682	73550	0.29917	0.15059	0.20033
Machine Learning	OCSVM	2063	2705	7442	67790	0.21704	0.43268	0.28908
	Isolation Forest	4171	597	15372	59860	0.21343	0.87479	0.34314
Hybrid Model	RF - IsoF	4180	588	15434	59798	0.21311	0.87668	0.34288
	LR - IsoF	4171	597	15372	59860	0.21343	0.87479	0.34314
	IsoF - RF	2193	2575	1003	74229	0.68617	0.45994	0.55073
	IsOE-LR	2260	2508	777	74455	0.74415	0.47399	0.57912

Table 10. Performance of each model for Dataset 03 with resample technique

D3 with Resample	Model	TP	FN	FP	TN	Precision	Recall	F1-Score
Supervised	DT	2278	2490	2635	72597	0.46367	0.47777	0.47061
Machine Learning	LR	3303	1465	2565	72667	0.56288	0.69274	0.62110
	SVM	3127	1641	2135	73097	0.59426	0.65583	0.62353
	KNN	2866	1902	3701	71531	0.43642	0.60109	0.50569
	RF	2550	2218	1634	73598	0.60946	0.53482	0.56971
Unsupervised	K means	661	4107	1739	73493	0.27542	0.13863	0.18443
Machine Learning	OCSVM	2138	2630	7447	67785	0.22306	0.44841	0.29792
	Isolation Forest	4159	609	15437	59795	0.21224	0.87227	0.34141
Hybrid Model	LR - IsoF	4175	593	15484	59748	0.21237	0.87563	0.34183
	SVM - IsoF	4160	608	15440	59792	0.21224	0.87248	0.34143
	IsoF - LR	3287	1481	2518	72714	0.56624	0.68939	0.62177
	IsoF - SVM	3126	1642	2132	73100	0.59452	0.65562	0.62358

5. CONCLUSION

All the objectives of the research have been achieved. Resampling technique is applied across all three dataset experiments to verify its effectiveness in solving the class imbalance problem. Results showed that it only has a consistent effect on LR and SVM models. Three different types of the dataset are used to investigate its effect on the performance of machine learning models in anomaly detection. Results showed that the transformation of the independent features plays a crucial role in determining the performance of each model. If the features are well transformed, the RF model is able to yield the best F1-Score.

Five supervised models and three unsupervised models are used in all the experiments to study those models' performance in anomaly detection. Two variants of hybrid models are developed for anomaly detection, where variant 1 hybrid model is focusing on improving the TP number of the supervised model while variant 2 is focusing on improving the FP number in unsupervised model. The performance of the two variants of hybrid models is compared and evaluated across all the experiments.




In future, a formula which contains the weightage of the probability from each model from the hybrid model can be developed for predicting the final outcome of the classification to further enhance the performance in anomaly detection. The weightage can be adjusted accordingly depending on the needs of the application. Besides, more types of data from different domains can be used to verify the effectiveness of the proposed hybrid models.

A hybrid model that can detect fraud in real-time over the time series dataset can also be developed. To make sure the machine learning model only selects useful features for model training, intelligence algorithms can be incorporated for feature selection. Lastly, resampling techniques from different python libraries can also be applied to test its effectiveness in improving model performance.




REFERENCES

- [1] E. Quatrini, F. Costantino, G. Di Gravio, and R. Patriarca, "Machine learning for anomaly detection and process phase classification to improve safety and maintenance activities," *Journal of Manufacturing Systems*, vol. 56, pp. 117–132, Jul. 2020, doi: 10.1016/j.jmsy.2020.05.013.
- [2] I. Sadgali, N. Sael, and F. Benabbou, "Performance of machine learning techniques in the detection of financial frauds," *Procedia Computer Science*, vol. 148, pp. 45–54, 2019, doi: 10.1016/j.procs.2019.01.007.
- [3] O. Faraj, D. Megías, A. M. Ahmad, and J. Garcia-Alfaro, "Taxonomy and challenges in machine learning-based approaches to detect attacks in the internet of things," in *ACM International Conference Proceeding Series*, Aug. 2020, pp. 1–10, doi: 10.1145/3407023.3407048.
- [4] A. Verner and S. Mukherjee, "An LSTM-based method for detection and classification of sensor anomalies," in *ACM International Conference Proceeding Series*, Jun. 2020, pp. 39–45, doi: 10.1145/3409073.3409089.
- [5] R. J. Hsieh, J. Chou, and C. H. Ho, "Unsupervised online anomaly detection on multivariate sensing time series data for smart manufacturing," in *Proceedings - 2019 IEEE 12th Conference on Service-Oriented Computing and Applications, SOCA 2019*, Nov. 2019, pp. 90–97, doi: 10.1109/SOCA.2019.00021.
- [6] H. Qosim and Zulkarnain, "Fault detection system using machine learning on synthesis loop ammonia plant," *ACM International Conference Proceeding Series*, pp. 74–80, 2020, doi: 10.1145/3400934.3400950.
- [7] Z. Zheng *et al.*, "Contextual anomaly detection in solder paste inspection with multi-task learning," *ACM Transactions on Intelligent Systems and Technology*, vol. 11, no. 6, pp. 1–17, Dec. 2020, doi: 10.1145/3383261.
- [8] H. Wang, M. Li, F. Ma, S. L. Huang, and L. Zhang, "Poster abstract: Unsupervised anomaly detection via generative adversarial networks," in *IPSN 2019 - Proceedings of the 2019 Information Processing in Sensor Networks*, Apr. 2019, pp. 313–314, doi: 10.1145/3302506.3312605.
- [9] S. Makki, Z. Assaghir, Y. Taher, R. Haque, M. S. Hacid, and H. Zeineddine, "An experimental study with imbalanced classification approaches for credit card fraud detection," *IEEE Access*, vol. 7, pp. 93010–93022, 2019, doi: 10.1109/ACCESS.2019.2927266.
- [10] T. Baabdullah, A. Alzahrani, and D. B. Rawat, "On the comparative study of prediction accuracy for credit card fraud detection with imbalanced classifications," in *Proceedings of the 2020 Spring Simulation Conference, SpringSim 2020*, 2020, doi: 10.22360/SpringSim.2020.CSE.004.
- [11] N. Rtayli and N. Enneya, "Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization," *Journal of Information Security and Applications*, vol. 55, p. 102596, Dec. 2020, doi: 10.1016/j.jisa.2020.102596.
- [12] S. Dhankhad, E. A. Mohammed, and B. Far, "Supervised machine learning algorithms for credit card fraudulent transaction detection: A comparative study," in *Proceedings - 2018 IEEE 19th International Conference on Information Reuse and Integration for Data Science, IRI 2018*, Jul. 2018, pp. 122–125, doi: 10.1109/IRI.2018.00025.
- [13] A. H. Nadim, I. M. Sayem, A. Mutsuddy, and M. S. Chowdhury, "Analysis of machine learning techniques for credit card fraud detection," in *Proceedings - International Conference on Machine Learning and Data Engineering, iCMLDE 2019*, Dec. 2019, pp. 42–47, doi: 10.1109/iCMLDE49015.2019.00019.
- [14] U. Porwal and S. Mukund, "Credit card fraud detection in E-commerce," in *Proceedings - 2019 18th IEEE International Conference on Trust, Security and Privacy in Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering, TrustCom/BigDataSE 2019*, Aug. 2019, pp. 280–287, doi: 10.1109/TrustCom/BigDataSE.2019.00045.
- [15] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," *Internet of Things (Netherlands)*, vol. 7, p. 100059, Sep. 2019, doi: 10.1016/j.iot.2019.100059.
- [16] J. Inoue, Y. Yamagata, Y. Chen, C. M. Poskitt, and J. Sun, "Anomaly detection for a water treatment system using unsupervised machine learning," in *IEEE International Conference on Data Mining Workshops, ICDMW*, Nov. 2017, vol. 2017–November, pp. 1058–1065, doi: 10.1109/ICDMW.2017.149.
- [17] R. Bhatia, S. Benno, J. Esteban, T. V. Lakshman, and J. Grogan, "Unsupervised machine learning for network-centric anomaly detection in IoT," in *Big-DAMA 2019 - Proceedings of the 3rd ACM CoNEXT Workshop on Big Data, Machine Learning and Artificial Intelligence for Data Communication Networks, Part of CoNEXT 2019*, Dec. 2019, pp. 42–48, doi: 10.1145/3359992.3366641.
- [18] H. Zhu, G. Liu, M. Zhou, Y. Xie, A. Abusorrah, and Q. Kang, "Optimizing Weighted Extreme Learning Machines for imbalanced classification and application to credit card fraud detection," *Neurocomputing*, vol. 407, pp. 50–62, Sep. 2020, doi: 10.1016/j.neucom.2020.04.078.
- [19] M. Ehsan Rana and W. Wei, "A machine learning based software project schedule management solution," in *Test Engineering and Management*, 2020, pp. 307–321.
- [20] K. K. Keat, V. Hameed, and M. E. Rana, "Time Prediction algorithm based on distance and real-world conditions," in *CompuSoft an International Journal of Advanced Computer Society*, 2020, vol. 9, no. 9, pp. 3817–3823.
- [21] A. B. Nassif, M. A. Talib, Q. Nasir, and F. M. Dakalbab, "Machine learning for anomaly detection: a systematic review," *IEEE Access*, vol. 9, pp. 78658–78700, 2021, doi: 10.1109/ACCESS.2021.3083060.
- [22] A. Toshniwal, K. Mahesh, and R. Jayashree, "Overview of anomaly detection techniques in machine learning," in *Proceedings of the 4th International Conference on IoT in Social, Mobile, Analytics and Cloud, ISMAC 2020*, Oct. 2020, pp. 808–815, doi: 10.1109/I-SMAC49090.2020.9243329.
- [23] "Credit Card Fraud Detection," Kaggle, 2017. [Online]. Available: <https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud> (accessed May 16, 2021).
- [24] PaySim, "Synthetic financial datasets for fraud detection." www.kaggle.com/datasets/ealaxi/paysim1. <https://www.kaggle.com/datasets/ealaxi/paysim1> (accessed May 16, 2021).
- [25] "Index of/datasets." [Online]. packages.revolutionanalytics.com. <http://packages.revolutionanalytics.com/datasets/> (accessed May 16, 2021).




BIOGRAPHIES OF AUTHORS

Lai Kai Lok    holds a degree in Physics from University Tunku Abdul Rahman (UTAR) and currently is a Master student from Asia Pacific University of Technology & Innovation (APU) Malaysia, majoring in Artificial Intelligence. His research focuses on improving the performance of anomaly detection by developing hybrid models of supervised and unsupervised machine learning techniques. He published one PERFIK 2014 conference paper and co-authored few other papers when he was a research assistant in University Malaya Plasma Lab. He has more than four years experience working as a mechanical design engineer, focusing on 3D modelling of automotive lamp and car parts using CATIA v5, supporting KOITO Japan. He can be contacted at email: tp061241@mail.apu.edu.my.



Vazeerudeen Abdul Hameed    has more than ten years of experience in academia, predominantly in teaching and research. He obtained his PhD in Computer Science from Universiti Teknologi Malaysia. He is associated with Asia Pacific University of Technology & Innovation (APU) Malaysia since 2009. His areas of research include Machine Learning, Deep Learning and Computer Vision in which he has authored and co-authored many journal and conference publications. He can be contacted at email: vazeerudeen@gmail.com.



Muhammad Ehsan Rana    possesses more than 20 years of experience in teaching, research, and academic management. He did his PhD in Software Engineering from Universiti Putra Malaysia. He is associated with Asia Pacific University of Technology & Innovation (APU) Malaysia since 2008. He is an active researcher in the areas of Software Architecture, Cloud Computing, Blockchain and IoT. He has authored and co-authored many journal papers and conference publications in the areas specified. He is also a reviewer of several indexed journals and has participated in the organization of IEEE and other international conferences. He can be contacted at email: muhd_ehsanrana@apu.edu.my.